# DOCUMENT INTERCHANGE STANDARDS

## STANDARDS FOR DOCUMENT INTERCHANGE

Exchanging electronic texts between different formats has been a problem for many years. For example, it is difficult to convert a TeX (1) document into an MS Word (2) document. Proprietary solutions exist for specific purposes. For example, Microsoft invented the Rich Text Format (RTF) to facilitate the exchange between different versions of Word and other Microsoft Office software. But the existence of special conversion software (3) shows that even conversions between different versions of the same word processor can be problematic. Adobe's PDF (4) can be used to display complex texts on the Internet by Acrobat and a freeware application, but this solution is less suitable for printing, and reprocessing the PDF format requires yet other software. PostScript (5) is in heavy use as a printing standard but is unsuitable as a format for editing. Basically, ASCII text (6) is the only format that is more or less universally interchangeable.

As with other data types, standards exist that facilitate the interchange of documents. A standard is a documented agreement that contains technical specifications to ensure that objects can be used as described. For example, the format of credit cards, phone cards, and "smart" cards is derived from a standard. Adhering to the standard, which defines an optimal thickness (0.76 mm) ensures that the cards can be used worldwide.

The advent of multimedia and the World Wide Web have extended the concept of a document far beyond the classical media (stone, wood, paper, computer screens). ASCII and its successor UNICODE (7) are standards, but they are too limited to be useful for document interchange or archival. ASCII text does not contain any semantic information, nor does it contain any layout or other accessory information. Proprietary formats such as PDF may provide a solution in a specific environment. However, for archival purposes, open standards are required that are system and vendor independent.

A truly universal solution for document archival consists of the use of a number of standards for document models. This article will describe the purpose of and the interaction between the standards that have been recently adopted in that area: SGML, CALS, HTML, and XML, HyTime, DSSSL, CSS, MHEG and PREMO. Together, they attempt to solve the problem of interchanging and reusing composite documents.

### The Standardization Process

The standardization process is led by special organizations. The best known is the International Organization for Standardization (ISO) (8). ISO is a federation of national standards bodies from some 100 countries, one from each country. ANSI (9), the American National Standards Institute, is the member body from the United States. The mission of ISO is to promote the development of standardization. Standards facilitate the international exchange of goods and services and help to develop intellectual and economic cooperation.

ISO publishes international standards. The creation of a standard takes place according to precise rules and regulations. Before an International Standard (IS) is published, it is circulated among voting participants of a committee as a Draft International Standard (DIS). The DIS is modified according to the comments received during the voting procedure and is finally published as IS.

The ISO staff is excluded from participating in the working groups and the voting committees and is not responsible for the technical content or quality of a standard.

Many ISO standards are joint publications with the International Electrotechnical Commission (IEC) (10). The close links between ISO and IEC are emphasized by the fact that the central secretariats of both organizations can be found in the same building, in Geneva, Switzerland. The IEC concentrates on standards in the fields of electricity, electronics, and related technologies.

Other organizations create open standards, such as the Object Management Group (11), the W3 Consortium (12), the Internet Society [through the IETF (13)], and OASIS, formerly known as the SGML Open Consortium (14). The membership fee for the participation in some of these organizations is directly proportional to a company's turnover, a fact that might bias the creation of a standard.

## THE STANDARD GENERALIZED MARKUP LANGUAGE

The Standard Generalized Markup Language (SGML) (15,16) is a neutral (vendor and system independent) format that allows easy reuse of data. With SGML, data can be published simultaneously on paper or in electronic form from a single source.

### Document Structures

A document has two structures:

- a physical structure which defines the visual attributes of the text (and images) as it is laid out on a page (such as font type and size, white space, positioning of text); physical structure is also called specific structure, i.e., it is fit only for one purpose
- a logical structure which defines the semantics of a text, identifying the meaning of the data (such as headings, titles, hypertext links, cross references) irrespective of a physical page; logical structure is also called generic structure, i.e., it may be reused for multiple purposes

We use SGML to model a documents' logical structure.

### Markup Languages

The machine and system independence, and hence the reusability of SGML data, is achieved by adding text strings called *markup* to a document. Markup originates in the publishing industry. In traditional publishing, the manuscript is annotated by a copy-editor with layout instructions for the typesetter. These handwritten annotations are called markup.

### Procedural Markup

*Procedural markup* refers to commands that fit alongside the text and directly influence its processing. For example, "set this text in a 12-point Helvetica bold typeface."

### Descriptive Markup

*Descriptive* commands also provide descriptive information about their purpose, such as "this piece of text is a chapter title."

### Generic Markup

In addition to delimiting parts of a document, *generic markup* indicates the semantics of these parts. For example, "chapter" rather than "new page, 18 point boldface Helvetica, centered." When text is marked as a chapter, any system can render it in the way it is best able to do. Style sheets of modern word processors are an example of generic markup.

Generic markup has two benefits over procedural markup: generic markup achieves higher portability and is more flexible than procedural markup.

### Generalized Markup

*Generalized markup* defines the rules for creating a generic coding language. Some generic markup languages are meta languages. New applications of data thus described become possible, such as automatic database loading. Or we can select a document or a part of it without necessarily having to scan its entire content.

SGML regulates the definition of generalized markup languages.

### The Abstract Nature of SGML

SGML permits its application to an infinite variety of document structures. A concrete markup language, i.e., *grammar,* describes the structure and semantics of a particular class of documents. It is defined in a document type definition (DTD). DTDs exist for books, articles, computer manuals, aircraft manuals, jurisprudence articles, mathematical formulas, tables, patent applications, drug applications, submarine maintenance manuals, and other, usually complex technical documents. The best known SGML application is HyperText Markup Language (HTML), used by the World Wide Web. The logical parts of a document that are defined in the DTD are called elements. The elements in the HTML DTD describe a fairly broad class of general-purpose documents.

**The SGML Syntax.** SGML is a delimiter based language. Like the names of the semantic parts of a document, the characters of the delimiters may be changed. Once a DTD is fixed and installed for use, the syntax is also fixed. Here are the most commonly used delimiters:

| | |
|---|---|
| ⟨ | start-tag open delimiter |
| ⟩ | tag close delimiter |

| | |
|---|---|
| ⟨/ | end-tag open delimiter |
| ⟨html⟩ | start-tag of the html element |
| ⟨/html⟩ | end-tag of the html element |
| = | value indicator |
| ` ` | literal string delimiter |
| & | entity reference open delimiter |
| ; | entity reference close delimiter |
| < | entity reference to the entity "lt" |

An SGML *parser* scans a document for these delimiters. When one is found, it triggers a change in the way it recognizes the data following it.

**Elements.** *Elements* are the logical units that an author (or the software used by the author) should recognize and mark up. Examples of HTML elements are: html, header, title, body, img, a, p. h1, h2, etc. They are marked up by adding start-tags at the start of the element: ⟨header⟩, ⟨title⟩, ⟨img⟩, ⟨a⟩, ⟨p⟩, ⟨h1⟩, ⟨h2⟩; and by adding end-tags at the end of the element: ⟨/header⟩, ⟨/title⟩, ⟨/a⟩, ⟨/h1⟩, ⟨/h2⟩. There are various different elements, depending on the data they contain. Elements can be nested inside each other, and can be defined as required, optional, or required and repeatable.

Element names may be in upper, lower, or mixed case (i.e., ⟨IMG⟩=⟨img⟩=⟨Img⟩).

**Attributes.** It is possible to qualify additional information about an element that goes beyond the structure. This information is given as the value of an *attribute*. For example, the ⟨IMG⟩ and ⟨A⟩ tags both have an attribute that specifics the URL of the image or the target of the link:

```
⟨img src =
"http://lhcb.cern.ch/images/gif/lhcb.gif"⟩
```

The ⟨img⟩ tag is an empty tag, and has no content. In this example of an ⟨a⟩ tag, the URL is contained in the value of the HREF attribute:

```
⟨a name = "z0" href =
"http://lhcb.cern.cb/default.htm"⟩
 LHCb home page⟨/A⟩
```

Notice that the value is always surrounded by double quotes (literal string delimiters), and that the attribute only appears on the element's start-tag. Attribute names may be specified in upper, lower, or mixed case letters (i.e., SRC=src=Src). Attribute values are literals, and are left in the case they are specified.

SGML has a number of predefined attribute types, such an unique identifiers and references thereto. For example, if the attribute "ID" in the tag ⟨P ID=``first''⟩ is declared as a "unique identifier" in the DTD, it can be used as the target of a cross reference link. This could be done by ⟨Pref IDREF=``first''⟩ if IDREF is declared as an attribute of type "reference to a unique identifier" in the DTD. SGML only permits references within the same document. To make a reference outside an SGML document, HyTime is needed.

**Entities.** An entity is a collection of characters that can be referenced as a unit. Among other things, *entities* allow the identification of characters that cannot be entered from the keyboard. For example, in some countries, there is no key corresponding to an e with an acute accent on it, é. This symbol is represented by the entity eacute. To refer to it, the entity is enclosed within an entity reference open delimiter (&) and an entity reference close delimiter (;). Entity references should also be used when a text character is required that is the same as an SGML delimiter. For example, instead of typing "⟨", the reference "<" should be used and instead of "⟩", ">". Parameter entities are a special type of entity that can be used as a variable inside a markup definition.

### Document Type Definitions

A *document type definition* (DTD) is also called an SGML application. The DTD defines the grammar of a concrete generic markup language for a class of similar documents. It defines three types of markup commands: elements, attributes and entities, and their usage constraints. It defines

- the *names* of elements that are permissible
- how *often* an element may appear
- the *order* in which elements must appear
- whether markup, such as the start- or end-tag, may be *omitted*
- the *contents* of elements; i.e., the names of other elements that are allowed to appear inside them, down to the character data level
- tag *attributes* and their default values
- the names of all *entities* that may be used
- any *typewriter conventions* that can be exploited to ease adding markup

A DTD should not contain any information on how to process a document, or what it should look like. Some examples of well known DTDs are

- The World Wide Web uses HTML, the HyperText Markup Language. HTML is the world's largest SGML application.
- Three DTDs designed by the AAP (Association of American Publishers) for books, scientific articles, and serials. This was the first major application of SGML.
- ISO 12083:1994 (17) "Electronic Manuscript Preparation and Markup." These are modernized and improved versions of the AAP document types, plus a DTD for mathematical formulas. This work was mainly done by a workgroup of the publications committee of the European Physical Society.
- Docbook (18). A DTD for technical documentation.
- The Text Encoding Initiative (19). A DTD for the humanities.
- The Continuous Acquisition and Life-Cycle Support (CALS) DTDs which are discussed in the next section.

## CONTINUOUS ACQUISITION AND LIFE-CYCLE SUPPORT

Continuous acquisition and life-cycle support (CALS) began as a Department of Defense (DoD) initiative in the 1980s to

**Table 1. The CALS Standards**

| Standard | Description | Use |
|---|---|---|
| CALS MILS-STD-1840A | Automated Interchange of Technical Information | Overall standard for exchanging and archiving technical information. |
| IGES MIL-D-28000 | Initial Graphic Exchange Specification | A graphics standard for representing 3-D CAD drawings. |
| SGML MIL-M-28001 | Standard Generalized Markup Language | Technical documents should be marked up with SGML. |
| CCITT G4 MIL-R-28002 | Group 4 Facsimile Standard | A standard for describing raster (bit-mapped) data. |
| CGM MIL-D-28003 | Computer Graphics Metafile | A standard format for describing 2-D illustrations with geometric graphics objects. |

exchange technical data with the government in electronic form rather than on paper. CALS is a management philosophy to improve the acquisition and life cycle support process. Moving to electronic creation, storage and transfer of information is not an idea specific to the DoD, and industry is very involved in policy and implementation related to CALS. Industry has informally renamed the CALS acronym to mean "Commerce at Light Speed."

Many government agencies like NASA, the Department of Commerce, and the Department of Energy are involved with encouraging CALS practices and the CALS strategy is practiced widely in the United Kingdom, Australia, Taiwan, Korea, Japan, and many other countries.

The major CALS standard families are shown in Table 1.

## THE HYPERTEXT MARKUP LANGUAGE AND THE EXTENSIBLE MARKUP LANGUAGE

### Why HTML?

During the mid 1980s research at CERN in the use of SGML for client-server documentation systems showed that there was a need for a format that could be easily translated and displayed on many different client platforms. Such a system was developed, and it offered a solution for IBM systems (searching, printing, previewing of printable documents). Fewer functions were available on other platforms (searching and viewing of ASCII text only). The system used the IBM RSCS protocol and could be used across Bitnet, a popular, wide area network linking together IBM mainframe computers. It used SGML for text, but for graphics, it was limited to proprietary formats. It soon became clear that a more general system was required based on Transmission Control Protocol/Internet Protocol (TCP/IP) with proper client-server format negotiation. This became the protocol called the Hyper Text Transfer Protocol (HTTP).

To build hypertext into a documentary base, Tim Berners-Lee adopted a flexible document type which he called the Hyper Text Markup Language (HTML) (20). In the interest of simplicity, HTML contained the most common elements that were used by CERN's SGML system at that time, without imposing any particular structure. These were augmented with the tags that were required to support the hypertext paradigm of the web.

### What's Next?

Despite the huge success of the Web, HTML has often been criticized because of its lack of structure. It is also not convenient in a distributed environment where an application would like to take advantage of the client's computing power. One solution to this problem is the Java (21) language, and the HTML extensions that allow the execution of Java programs from pages on the Web. To ensure more efficient data transfer and to avoid the proliferation of HTML dialects the Extensible Markup Language (XML) (22) is being proposed. XML is called extensible because it extends HTML towards a more complete support of SGML. With XML, it will be possible to send richer data over the Internet.

### Extensible Markup Language

The Extensible Markup Language is a subset of SGML. The goal is to enable SGML to be served, received, and processed on the Web in the way that is now possible with HTML. For this reason, XML has been designed for ease of implementation and for interoperability with both SGML and HTML.

The adoption of XML will make several types of applications much easier. For example,

- accessing specialized semantics (transferring personal data from one database to another)
- multiple presentation of data (depending on the reader's background)
- off-loading computation from the server to the client [for CORBA (23) type applications]
- obtaining personalized data (from newspapers)

These applications require data to be encoded using tags that describe a rich set of semantics. An alternative to XML for these applications is proprietary code embedded as "script elements" in HTML documents and delivered in conjunction with proprietary browser plug-ins or Java applets. A very interesting application of XML is the Channel Definition Format (CDF) (24) developed by Microsoft. CDF is an open specification that permits a web publisher to offer frequently updated collections of information, or channels, from any web server for automatic delivery to compatible receiver programs on computers ("server push").

## HYPERMEDIA/TIME BASED STRUCTURING LANGUAGE

The Hypermedia/Time based structuring language (HyTime) is an extension of SGML (25,26) that describes connected or temporal information. It covers different aspects of linking, as well as features for multimedia purposes, including virtual time, scheduling, and synchronization.

### Architectural Forms

HyTime facilities can be integrated into any SGML DTD via the technique of architectural forms. An architectural form can be compared to an object oriented "framework." The design of a particular hypertext construct, such as a link, is defined by the architectural form, ready to be re-used by the DTD writer. By giving an element in a DTD a specific HyTime attribute with values that are specified in the HyTime standard, a HyTime engine knows what action it should take.

### Links and Locations

Hypertext systems like the World Wide Web do not differentiate between a link and its endpoints (the locations or anchors). The link information is treated as a whole; any editing change affects the entire linking construct. For example, the HTML anchor ⟨`A HREF=''http://www.cern.ch/''`⟩ contains the target of the link. The drawbacks of this approach are obvious: as soon as the object behind the target (`''http://www.cern.ch/''`) no longer exists or is moved elsewhere, the link needs to be updated. As the owner of the target is unaware of who links to it, this is an impossible task.

HyTime makes a distinct difference between links and locations. A link is a reference between two or more locations. A location is the address of a potential anchor point, which is the actual, physical point where the link ends.

This step of indirection guarantees easy maintenance of hyperlinks. HyTime brings about a consistent way of describing hyperlinks between any media.

It should be noted that the World Wide Web is attempting to solve the moving target problem via Universal Resource Numbers, Uniform Resource Names, and similar efforts.

### HyTime Links

Linking in SGML is achieved by assigning ID attributes to elements and referencing these through an IDREF attribute, which is limited to references inside the same document.

HyTime defines two link architectural forms: contextual links (which are part of the document where the link markup resides) and independent links (stored externally to the document that the link markup connects). The richness of HyTime linking lies in the many ways one can describe locations in a document, often as a sequence of stepwise refined addresses known as location ladders.

The separation of links from anchors makes links more robust. It allows for links between any kind of media, such as musical notes.

## DOCUMENT LAYOUT

Document layout is the skill of positioning text and images on a page, or for electronic media, on a screen. Document layout on electronic media is classically achieved via procedural markup. For example in TeX, the effect of the command `\rm{text}` will be that the word "text" will be typeset using a Roman font. Procedural markup can give powerful but nonportable results.

Procedural markup commands depend on the system that will process the document (TeX in the case of the command `\rm{text}`). Another example that illustrates different results when procedural markup is used can be seen when an HTML document is viewed with different Web browsers. The ⟨`TABLE`⟩ tag was introduced by NetScape in version 2.0 of its Navigator. Older browsers like NCSA Mosaic will not display the table layout as intended, although all browers are supposed to ignore unrecognized tags with no loss of content. The issue of rendering of different media on the Web is important but is not the main issue here.

By using generic markups, such as SGML, HTML, or XML, a document can be freely interchanged. However, the generic makeup contains no information on the document layout. There are several ways in which this information can be captured. The Cascading Style Sheet (27) mechanism was invented by the W3 Consortium as a standard way to describe the layout of HTML documents. DSSSL (28) defines the document layout of SGML documents. When (and if) XML replaces HTML, a special version of DSSSL for online document delivery, DSSSL-O (29) will be used instead of the CSS mechanism to convey layout information over the Internet.

### Cascading Style Sheets

Cascading Style Sheets (CSS) are interoperable style sheets that allows authors and readers to attach style (e.g., fonts, colors, and spacing) to HTML documents. They

- allow designers to express typographic effects
- allow externally linked as well as internal and inline style sheets
- be interoperable across Web applications
- support visual, as well as non-visual, output media
- are applied hierarchically (hence "cascading")

Unfortunately, the CSS mechanism is completely different from DSSSL.

### Document Style Semantics and Specification Language

DSSSL (pronounced *dis-sul*), the Document Style Semantics and Specification Language, provides a standardized syntax and layout for SGML documents. DSSSL is declarative, in the sense that style specifications are made by describing final results, not by describing the procedures that are used to create the formatted results.

### The Parts of the DSSSL Standard

DSSSL is divided into different parts. The most important of these are

- the Style Language
- DSSSL-Online
- the Transformation Language
- the Standard Document Query Language
- the Expression Language

The transformation process changes one SGML document (conforming to a certain DTD) into another SGML document (conforming to another DTD). The commands to do this are specified in the expression language. The style language describes the formatting of SGML documents; the style language also uses the expression language. Both the transformation language and the style language can address any object using the standard document query language. HyTime shares the standard document query language with DSSSL.

## MULTIMEDIA AND HYPERMEDIA INFORMATION CODING EXPERT GROUP

The Multimedia and Hypermedia Information Coding Expert Group (MHEG) (30), is a hypermedia architecture for multimedia distribution. MHEG can run in environments with very small resources, such as set-top boxes where Java-enabled browsers are an overload. Although it was originally developed for broadcast applications, MHEG has some substantial advantages for information and point-of-sales terminals as well as interactive TV.

In the MHEG model, a video sequence is a "scene" with moving objects. The "scene" remains constant over some time and is only transmitted once by the server to the client. The moving objects, which are smaller in size, are transmitted continuously. Consequently, MHEG will relieve the load on the server. It employs an object-oriented model, generic enough to format different kinds of multimedia documents, and provides on any network the quality of service people expect from TV. In addition, it offers powerful models of spatial and temporal synchronization between different media, which are not provided for in other standards. The latest MHEG standard is MHEG-5.

### Objectives

MHEG-5 was designed for interactive multimedia applications such as video on demand, home shopping, games, education, and information. The standard allows large applications to be distributed between server and client in a way that the client only requires a small amount of memory.

HTML and MHEG-5 have many concepts in common, such as the focus on declarative code. However, HTML is a document description language, not a format for describing multimedia applications. MHEG-5 is built from the ground up with the needs of multimedia applications in mind, such as synchronization and speed control of streams, handling of stream events, etc. The Java language makes it possible to write applets for multimedia applications that can be embedded in HTML documents. However, the performance and size of current Java systems are inadequate for the limited resources of interactive TV, whereas current MHEG-5 implementations will fit in a few kB.

## PRESENTATION ENVIRONMENT FOR MULTIMEDIA OBJECTS

The Presentation Environment for Multimedia Objects (PREMO) (31) addresses the creation of, presentation of, and interaction with all forms of information using single or multiple media. PREMO is a project which has not yet resulted in a published standard.

### Objectives

The aim of PREMO is the standardization of programming environments for the presentation of multimedia data. PREMO will support still computer graphics, moving computer graphics (animation), synthetic graphics of all types, audio, text, still images, moving images (including video), images coming from imaging operations, and other media types or combinations of media types that can be presented.

PREMO complements the work of other emerging ISO standards on Multimedia, such as MHEG and HyTime. These standards do not aim at the presentation of media objects, but deal with aspects of the interchange of multimedia information.

### Description

The Graphical Kernel System (GKS) was the first ISO standard for computer graphics. It was followed by a series of complimentary standards, addressing different areas of computer graphics such as PHIGS, PHIGS PLUS, and CGM. One of the main differences between PREMO and previous graphics standards is the inclusion of multimedia aspects.

Technology has made it possible to create systems which use, within the same application, different presentation techniques that are not necessarily related to synthetic graphics, for example, video, still images, and sound. Examples of applications where video output, sound, etc., and synthetic graphics (e.g., animation) coexist are numerous. PREMO proposes development environments that are enriched with techniques supporting the display of different media in a consistent way and which allow for the various media-specific presentation techniques to coexist within the same system.

PREMO needs to solve the problem of synchronization of video and sound presentation. This problem is well known in the multimedia community; its integration with the more general demands of a presentation system will obviously be a challenge.

## CONCLUSION

The area of standards for document models is a very rapidly moving field. SGML and its existing applications will be further adopted and exploited. DSSSL and HyTime will inspire new, more sophisticated applications and software. The Internet and Java will push XML, and possibly DSSSL-O. The evolution of MHEG in the interactive TV world seems clear—perhaps MHEG and PREMO will pave the way to a true integration of mass market multimedia applications with the more conventional picture of a document that we have today.

## BIBLIOGRAPHY


1. D. E. Knuth, *The TeX Book,* Reading, MA: Addison-Wesley, 1986. Available www: http://www.tug.org/
2. Microsoft Corporation, *Using Microsoft Word, Word Processing Program 97,* Redmond, WA: Microsoft Corporation, 1997. Available www: http://www.microsoft.com/word/
3. Microsoft Corporation, *Word 6.0 / 95 Binary Convertor for Word 97 now available,* Redmond, WA: Microsoft Corporation, 1997. Available www: http://www.microsoft.com/organizations/corpeval/500.htm
4. Adobe Systems Inc., *Acrobat 3.0 User's Guide,* San Jose, CA: Adobe Systems Inc., 1997. Available www: http://www.adobe.com/studio/tipstechniques/acrcreatepdf/main.html
5. Adobe Systems Inc., *Postscript Language Reference Manual,* Reading, MA: Addison-Wesley, 1990, 2nd ed.
6. *The International Standard Corresponding to ASCII is ISO / IEC 646:1991* (ISO 7-Bit Coded Character Set for Information Interchange), Geneva: ISO, 1991.
7. *The International Standard Corresponding to UNICODE is ISO / IEC 10646-1:1993* (Universal Multiple-Octet Coded Character Set (UCS)—Part 1: Architecture and Basic Multilingual Plane), Geneva: ISO, 1993.
8. The International Organization for Standardization (ISO), 1 rue de Varembé, PO Box 56, CH-1211 Geneva 20, Switzerland. Available www: http://www.iso.ch
9. The American National Standards Institute (ANSI), 11 West 42nd Street, New York 10036. Available www: http://www.ansi.org
10. The International Electrical Commission (IEC), 3 rue de Varembé, PO Box 131, 1211 Geneva 20, Switzerland. Available www: http://www.iec.ch
11. The Object Management Group (OMG), 492 Old Connecticut Path, Framingham, MA 01701. Available www: http://www.omg.org
12. The World-Wide Web Consortium, Massachusetts Institute of Technology, Laboratory for Computer Science, 545 Technology Square, Cambridge, MA 02139. Available www: http://www.w3.org/pub/WWW/
13. The Internet Society creates its standards through the Internet Engineering Task Force (IETF). Available www: http://info.isoc.org/
14. The SGML Open Consortium, 7950 Hwy. 72 W. Suite G201, Madison, AL 35758. Available www: http://www.sgmlopen.org
15. *The International Standard Corresponding to SGML is ISO / IEC 8879:1986* (Standard Generalized Markup Language), Geneva: ISO, 1986.
16. E. van Herwijnen, *Practical SGML,* Boston: Kluwer Academic, 1994, 2nd ed.
17. *The International Standard Corresponding to ISO 12083 is ISO / IEC 12083:1994,* Geneva: ISO, Electronic Manuscript Preparation and Markup, 1994. This standard is also available from ISO in electronic form with examples and explications as "The Annotated ISO 12083."
18. The Davenport group maintain the Docbook DTD. Available www: http://www.ora.com/davenport/
19. The Text Encoding Initiative. Available www: http://www.uic.edu/orgs/tei/
20. B. White, *HTML and the Art of Authoring for the World-Wide Web,* Boston: Kluwer Academic, 1996.
21. P. van der Linden, *Just Java,* Palo Alto, CA: Sun Microsystems Press, 1996.
22. Extensible Markup Language (XML). W3C Working Draft WD-xml-961114. Available www: http://www.w3.org/pub/WWW/TR/WD-xml-961114.html
23. *The Common Object Request Broker: Architecture and Specification (2.),* Object Management Group, 1995.
24. Microsoft Corporation, *Channel Definition Format,* Redmond, WA: Microsoft Corporation, 1997. Available www: http://www.microsoft.com/standards/cdf.htm
25. *The International Standard Corresponding to HyTime is ISO 10744:1992* (Hypermedia/Time-based Structuring Language (HyTime), Geneva: ISO, 1992.
26. S. J. DeRose and D. G. Durand, *Making HyperMedia Work: A User's Guide to HyTime,* Boston: Kluwer Academic, 1994.
27. H. Lie and B. Bos, Cascading Style Sheets, level 1, 1996. Available www: http://www.w3.org/pub/WWW/TR/REC-CSS1-961217.html
28. *The International Standard Corresponding to DSSSL is ISO / IEC 10179-1996* [Document Style Semantics and Specification Language (DSSSL)], Geneva, IOS, 1996.
29. DSSSL Online Application Profile, 1996. Available www: http://sunsite.unc.edu/pub/sun-info/standards/dsssl/dssslo/dssslo.htm
30. ISO/IEC, *Coding of Multimedia and Hypermedia Information (MHEG),* Geneva: ISO, 1900, ISO/IEC 1DIS 3522-1/4.
31. ISO/IEC, Presentation Environment for Multimedia Objects (PREMO), Geneva: IOS, 1900. ISO/IEC 14478. Available www: http://www.cwi.nl/Premo/premo.html


Eric van Herwijnen
CERN