GEOGRAPHIC INFORMATION SYSTEMS

A Geographic Information System (GIS) is a set of computer-based tools that collects, stores, retrieves, manipulates, displays, and analyzes geographic information. Some definitions of GIS include institutions and people besides the computer-based tools and the geographic data. These definitions refer more to a total GIS implementation than to the technology.

Here, *computer-based tools* are hardware (equipment) and software (computer programs). *Geographic information* describes facts about the earth's features, for example, the location and characteristics of rivers, lakes, buildings, and roads. *Collection* of geographic information refers to the process of gathering, in computer-compatible form, facts about features of interest. Facts usually collected are the location of features given by sets of coordinate values (such as latitude, longitude, and sometimes elevation), and attributes such as feature type (highway), name (Interstate 71), and unique characteristics (the northbound lane is closed). *Storing* of geographic information is the process of electronically saving the collected information in permanent computer memory (such as a computer hard disk). Information is saved in structured computer files. These files are sequences of only two characters, 0 and 1, called bits, organized into bytes (eight bits) and words (16–64 bits). These bits represent information stored in the binary system. *Retrieving* geographic information is the process of accessing the computer-compatible files, extracting sets of bits and translating them into information we can understand (for example, information given in our national language). *Manipulation* of geographic data is the process of modifying, copying, and removing from computer permanent memory selected sets of information bits or complete files. *Display* of geographic information is the process of generating and making visible a graphic (and sometimes textual) representation of the information. *Analysis* of geographic information is the process of studying, computing facts from the geographic information, and asking questions (and obtaining answers from the GIS) about features and their relationships. For example, what is the shortest route from my house to my place of work?

HARDWARE AND ITS USE

The main component is the computer (or computers) on which the GIS run. Currently, GIS systems run on desktop computers to mainframes (used as a stand-alone or as a network configuration). In general, GIS operations require handling large amounts of information (fifty megabytes or larger file sizes are common), and in many cases, GIS queries and graphic displays must be generated very quickly. Therefore, important characteristics of computers used for GIS are processing speed, quantity of random access memory (RAM), size of permanent storage devices, resolution of display devices, and speed of communication protocols.

Several peripheral hardware components may be part of the system: printers, plotters, scanners, digitizing tables, and other data collection devices. *Printers* and *plotters* are used to generate text reports and graphics (including maps). High-speed printers with graphics and color capabilities are commonplace today. The number and sophistication of the printers in a GIS organization depend on the amount of text reports to be generated. Plotters allow the generation of oversized graphics. The most common graphic products of a GIS system are maps. As defined by Thompson (1), “Maps are graphic representations of the physical features (natural, artificial, or both) of a part or the whole of the earth’s surface. This representation is made by means of signs and symbols or photographic imagery, at an established scale, on a specified projection, and with the means of orientation indicated.” As this definition indicates, there are two different types of maps: (1) line maps, composed of lines, the type of map we are most familiar with, usually in paper form, for example a road map; and (2) image maps, which are similar to a photograph. Plotters able to plot only line maps are usually less sophisticated (and less expensive) than those able to plot high-quality line and image maps. Plotting size and resolution are other important characteristics of plotters. With some plotters it is possible to plot maps with a size larger than one meter. Higher plotting resolution allows plotting a greater amount of details. Plotting resolution is very important for images. Usually, the larger the map size needed, and the higher the plotting resolution, the more expensive the plotter.

Scanners are devices that sense and decompose a hardcopy image or scene into equal-sized units called pixels and store each pixel in computer-compatible form with corresponding attributes (usually a color value per pixel). The most common use of scanning technology is in fax machines. They take a hardcopy document, sense the document, and generate a set of electric pulses. Sometimes, the fax machine stores the pulses to be transferred later; other times they are transferred right away. In the case of scanners used in GIS, these pulses are stored as bits in a computer file. The image generated is called a raster image. A raster image is composed of pixels. Generally, pixels are square units. Pixel size (the scanner resolution) ranges from a few micrometers (for example, five) to hundreds of micrometers (for example, 100 micrometers). The smaller the pixel size the better the quality of the scanned images, but the larger the size of the computer file and higher the scanner cost. Scanners are used in GIS to convert hardcopy documents to computer-compatible form, especially paper maps.

Some GIS cannot use raster images to answer geographic questions (queries). Those GIS that can are usually limited in the types of queries they can perform (they can perform queries about individual locations but not geographic features). Most queries need information in vector form. Vector information represents individual geographic features (or parts of features) and is an ordered list of vertex coordinates. Figure 1 shows the differences between raster and vector. *Digitizing tables* are devices that collect vector information from hardcopy documents (especially maps). They consist of a flat surface on which documents can be attached and a cursor or puck with several buttons, used to locate and input coordinate values (and sometimes attributes) into the computer. The result of digitizing is a computer file with a list of coordinate values

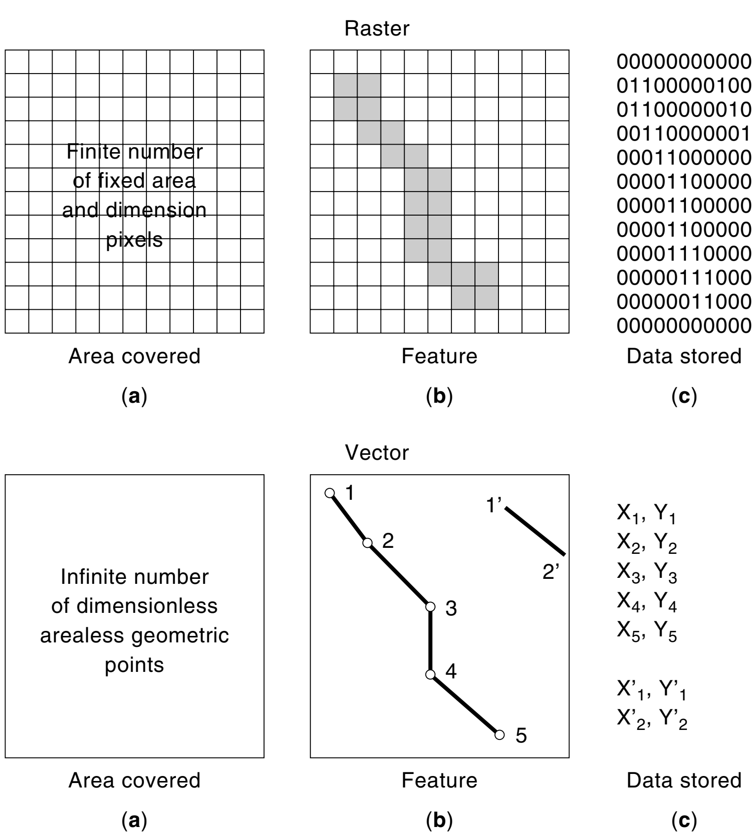

Figure 1. The different structures of raster and vector information, feature representation, and data storage.

and attributes per feature. This method of digitizing is called "heads-down digitizing."

Currently, there is a different technique to generate vector information. This method uses a raster image as a backdrop on the computer terminal. Usually, the image has been georeferenced (transformed into a coordinate system related in some way to the earth). The operator uses the computer mouse to collect the vertices of a geographic feature and to attach attributes. As in the previous case, the output is a computer file with a list of coordinate values and attributes for each feature. This method is called "heads-up digitizing."

SOFTWARE AND ITS USE

Software, as defined by the AGI dictionary (2), is the collection of computer programs, procedures, and rules for the execution of specific tasks on a computer system. A computer program is a logical set of instructions, which tells a computer to perform a sequence of tasks. GIS software provides the functions to collect, store, retrieve, manipulate, query and analyze, and display geographic information. An important component of software today is a graphical user interface (GUI). A GUI is set of graphic tools (icons, buttons, and dialogue boxes) that can be used to communicate with a computer program to input, store, retrieve, manipulate, display, and analyze information and generate different types of output. Pointing with a device such as a mouse to select a particular software application operates most GUI graphic tools. Figure 2 shows a GUI.

GIS software can be divided into five major components (besides the GUI): input, manipulation, database management system, query and analysis, and visualization. *Input software* allows the import of geographic information (location and attributes) into the appropriate computer-compatible format. Two different issues need to be considered: how to transform (convert) analog (paper-based) information into digital form, and how to store information in the appropriate format. Scanning, and heads-down and heads-up digitizing software with different levels of automation, transforms paper-based information (especially graphic) into computer-compatible form. Text information (attributes) can be imported by a combination of scanning and character recognition software, and/or by manual input using a keyboard and/or voice recognition software. In general, each commercial GIS software package has a proprietary format, used to store locations and attributes. Only information in that particular format can be used in that particular GIS. When information is converted from paper into digital form using the tools from that GIS, the result is in the appropriate format. When information is collected using other alternatives, then a file format translation needs to be made. Translators are computer programs that take information stored in a given format and generate a new file (with the same information) in a different format. In some cases, translation results in information loss.

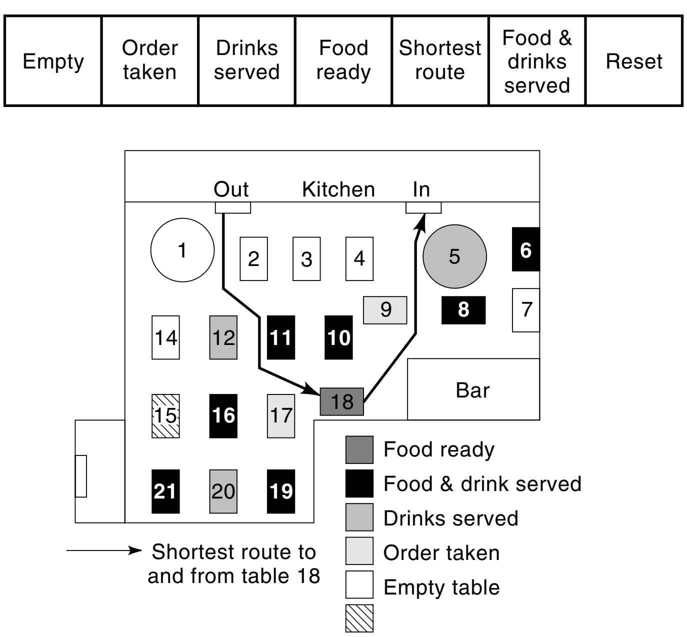

Figure 2. A graphic user interface (GUI) for a GIS in a restaurant setting and the graphic answers to questions about table occupancy, service, and shortest route to Table 18.

Manipulation software allows changing the geographic information by adding, removing, modifying, or duplicating pieces or complete sets of information. Many tools in manipulation software are similar to those in word-processors: create, open, and save a file; cut, copy, paste, undo graphic and attribute information. Many other manipulation tools allow drafting operations of the information, such as: draw a parallel line, square, rectangle, circle, and ellipse; move a graphic element, change color, line width, line style. Other tools allow the logical connection of different geographic features. For example, geographic features that are physically different and unconnected, can be grouped as part of the same layer, level, or overlay (usually, these words have the same meaning). By doing this, they are considered part of a common theme (for example, all rivers in a GIS can be considered part of the same layer: *hydrography*). Then, one can manipulate all features in this layer by a single command. For example, one could change the color of all rivers of the hydrography layer from light to dark blue by a single command.

Database management system (DBMS) is a collection of software for organizing information in a database. This software performs three fundamental operations: storage, manipulation, and retrieval of information from the database. A database is a collection of information organized according to a conceptual structure describing the characteristic of the information and the relationship among their corresponding entities (2). Usually, in a database there are at least two computer files or tables and a set of known relationships, which allows efficient access to specific entities. Entities in this concept are geographic objects (such as a road, house, and tree). Multipurpose DBMS are classified into four categories: inverted list, hierarchical, network, and relational. Healy (3) indicates that for GIS, there are two common approaches to DBMS: the hybrid and the integrated. The hybrid approach is a combination of a commercial DBMS (usually, relational) and direct access operating system files. Positional information (coordinate values) is stored in direct access files and attributes, in the commercial DBMS. This approach increases access speed to positional information and takes advantage of DBMS functions, minimizing development costs. Guptill (4) indicates that in the integrated approach the Standard Query Language (SQL) used to ask questions about the database is replaced by an expanded SQL with spatial operators able to handle points, lines, polygons, and even more complex structures and graphic queries. This expanded SQL sits on top of the relational database. This simplifies geographic information queries.

Query and analysis software provides new explicit information about the geographic environment. The distinction between query and analysis is somewhat unclear. Maguire and Dangermond (5) indicate that the difference is a matter of emphasis: "Query functions are concerned with inventory questions such as 'Where is . . .?' Analysis functions deal with questions such as 'What if . . .?'." In general, query and analysis use the location of geographic features, distances, directions, and attributes to generate results. Two characteristic operations of query and analysis are buffering and overlay. Buffering is the operation that finds and highlights an area of user-defined dimension (a buffer) around a geographic feature (or a portion of a geographic feature), and retrieves information inside the buffer, or generates a new feature. Overlay is the operation that compares layers. Layers are compared two at a time by location and/or attributes.

Query and analysis are the capabilities that differentiate GIS from other geographic data applications such as computer-aided mapping, computer-aided drafting (CAD), photogrammetry, and mobile mapping.

Visualization in this context refers to the software for visual representation of geographic data and related facts, facilitating the understanding of geographic phenomena, their analysis, and interrelations. The term visualization in GIS encompasses a larger meaning. As defined by Buttenfield and Mackaness (6), "visualization is the process of representing information synoptically for the purpose of recognizing, communicating, and interpreting pattern and structure. Its domain encompasses the computational, cognitive, and mechanical aspects of generating, organizing, manipulating, and comprehending such representation. Representation may be rendered symbolically, graphically, or iconically and is most often differentiated from other forms of expression (textual, verbal, or formulaic) by virtue of its synoptic format and with qualities traditionally described by the term 'Gestalt.'" It is the confluence of computation, cognition, and graphic design.

Visualization is accomplished through maps, diagrams, and perspective views. A large amount of information is abstracted into graphic symbols. These symbols are endowed with visual variables (size, value, pattern, color, orientation, and shape) that emphasize differences and similarities among those facts represented. The joint representation of the facts shows explicit and implicit information. Explicit information can be accessed by other means such as tables and text. Implicit information requires, in some cases, performing operations with information such as computing the distance between two points on a road. In other cases, by looking at the graphic representation, we can access implicit information. For example, we can find an unexpected relationship between the relief and erosion that is not obvious from the explicit information. This is the power of visualization!

USING GIS

GIS is widely used. Users include national, state, and local agencies; private business (from delivery companies to restaurants, from engineering to law firms); educational institutions (from universities to school districts, from administrators to researchers); and private citizens. As indicated earlier, the use of GIS requires software (that can be acquired from a commercial vendor), hardware (which allows running the GIS software), and data (with the information of interest). As indicated by Worboys (7), "data are only useful when they are part of a structure of interrelationships that form the context of the data. Such a context is provided by the *data model*." Depending on the problem of interest, the data model may be simple or complex. In a restaurant, information about seating arrangement, seating time, drinks, and food are well defined and easily expressed by a simple data model. Fundamentally, you have information for each table about its location, the number of people it seats, and the status of the table (empty or occupied). Once a table is occupied, additional information is recorded: how many people occupy the table; when it was occupied; what drinks were ordered; what food was ordered; the status of the order (drinks are being served, food is being prepared, etc.). Questions such as, What table is empty? How many people can be seated at a table? What table seats seven people? Has the food ordered by table 11 been served? How long before table 11 is free again? are easily answered from the above information with a simple data model (see Figure 2). Of course, a more sophisticated data model will be required if more complex questions are asked of the system. For example, What is the most efficient route to reach a table based on the current table occupancy? If alcoholic drinks are ordered at a table, how much longer will it be occupied than if nonalcoholic drinks are ordered? How long will it be before food is served to Table 11 if the same dish has been ordered nine times in the last few minutes?

Many problems require a complex data model. A nonexhaustive list of GIS applications that require complex models is presented next. This list gives an overview of many fields and applications of GIS:

Siting of a new business. Find the best location in this region for a new factory, based on natural and human resources.

Network analysis. Find the shortest bus routes to pick up students, for a given school.

Utility services. Find the most cost-efficient way to extend the electric service to a new neighborhood.

Land Information System. Generate an inventory of the natural resources of a region and the property-tax revenue, using land parcels as the basic unit.

Intelligent car navigation. What are the recommended speeds, geographic coordinates of the path to be followed, street classification, and route restrictions to go from location *A* to location *B*?

Tourist Information System. What is the difference in driving time to go from location *A* to location *B*, following the scenic route instead of the business route? And where, along the scenic route, are the major places of interest located?

Political campaigns. Set the most time-efficient schedule to visit the largest possible number of cities where undecided voters could make the difference during the last week of a political campaign.

Marketing branch location analysis. Find the location and major services to be offered by a new bank branch, based on population density and consumer preferences.

Terrain analysis. Find the most promising site in a region for oil exploration, based on topographic, geological, seismic, and geomorphological information.

QUALITY AND ITS IMPACT IN GIS

The unique advantage of GIS is the capability to analyze and answer geographic questions. If no geographic data is available for a region, of course, it is not possible to use GIS. On the other hand, the validity of the analysis and quality of the answers in GIS are closely related to the quality of the geographic data used. If poor quality or incomplete data were used, the query and analysis would provide poor or incomplete results. Therefore, it is fundamental to know the quality of the information in a GIS. Of course, the quality of the analysis and query capabilities of a GIS is also very important. Perfect geographic data used with poor-quality analysis and query tools generates poor results.

Quality is defined by the *U.S. National Committee Digital Cartographic Data Standard—NCDCDS* (8) as "fitness for use." This definition states that quality is a relative term: Data may be fit to use in a particular application but unfit for another. Therefore, we need to have a very good understanding of the scope of our application to judge the quality of the data to be used. The same committee identifies five quality components in the context of GIS in the Spatial Data Transfer Standard (SDTS): lineage, positional accuracy, attribute accuracy, logical consistency, and completeness.

SDTS is the *U.S. Federal Information Processing Standard—173* and states that "Lineage is information about the sources and processing history of the data." Positional accuracy is "the correctness of the spatial (geographic) location of features." Attribute accuracy is "the correctness of semantic (nonpositional) information ascribed to spatial (geographic) features." Logical consistency is "the validity of relationships (especially topological ones) encoded in the data," and completeness is "the mapping and selection rules and exhaustiveness of feature representation in the data." The International Cartographic Association (ICA) has added two more quality components: semantic accuracy and temporal information. As stated by Guptill and Morrison (9), "semantic accuracy describes the number of features, relationships, or attributes that have been correctly encoded in accordance with a set of feature representation rules." Guptill and Morrison (10) also state that "temporal information describes the date of observation, type of update (creation, modification, deletion, unchanged), and validity periods for spatial (geographic) data records." Most of our understanding about the quality of geographic information is limited to positional accuracy, specifically, point positional accuracy. Schmidley (11) has conducted research in line positional accuracy. Research in attribute accuracy has been done mostly in the remote sensing area, and some in GIS [see Chapter 4 of (9)]. Very little research has been done in the other quality components [see (9)].

To make the problem worse, because of limited digital geographic coverage worldwide, GIS users combine, many times, different sets of geographic information, each set of a different quality level. Most GIS commercial products have no tools to judge the quality of the data used: Therefore, it is up to the GIS user to judge and keep track of information quality.

Another limitation of GIS technology today is the fact that GIS systems, including analysis and query tools, are sold as "black boxes." The user provides the geographic data, and the GIS system provides results. In many cases the methods, algorithms, and implementation techniques are considered proprietary and there is no way for the user to judge their quality.

More and more users are starting to recognize the importance of quality GIS data. As a result, many experts are conducting research in the different aspects of GIS quality.

THE FUTURE OF GIS

GIS is in its formative years. All types of users have accepted the technology and it is a worldwide multibillion-dollar industry. This acceptance will create a great demand for digital geographic information in the near future. Commercial satellites and multisensor platforms generating high-resolution images, mobile mapping technology, and efficient analog-to-digital data conversion systems, are some of the promising approaches to the generation of geographic data.

GIS capabilities are improving. This is a result of the large amount of ongoing research. This research includes the areas of visualization, user interfaces, spatial relation languages, spatial analysis methods, geographic data quality, three-dimensional and spatio-temporal information systems, and open software design. These efforts will result in better, reliable, faster, and more powerful GIS.

BIBLIOGRAPHY

1. M. M. Thompson, *Maps for America,* 2nd ed. Reston, VA: U.S. Geological Survey, 253, 1981.
2. Association for Geographic Information, *AGI GIS Dictionary,* 2nd ed., http://www.geo.ed.ac.uk/agidexe/term638, 1993.
3. R. G. Healey, Database management systems, in D. J. Maguire, M. F. Goodchild, and D. W. Rhind (ed.), *Geographical Information Systems,* Harlow, UK: Logman Scientific Group, 1991.
4. S. C. Guptill, Desirable characteristics of a spatial database management system, *Proc. AUTOCARTO 8,* ASPRS, Falls Church, VA, 1987.
5. D. J. Maguire and J. Dangermond, The functionality of GIS, D. J. Maguire, M. F. Goodchild, and D. W. Rhind (ed.), *Geographical Information Systems,* Harlow, UK: Logman Scientific Group, 1991.
6. B. P. Buttenfield and W. A. Mackaness, Visualization, in D. J. Maguire, M. F. Goodchild, and D. W. Rhind (ed.), *Geographical Information Systems,* Harlow, UK: Logman Scientific Group, 1991.
7. M. F. Worboys, *GIS: A Computing Perspective,* London: Taylor & Francis, 1995 p. 2.
8. Digital Cartographic Data Standard Task Force, The proposed standard for digital cartographic data, *The American Cartographer,* **15**: 9–140, 1988.
9. Reprinted for S. C. Guptill and J. L. Morrison, *Elements of Spatial Data Quality,* Copyright 1995, with kind permission from Elsevier Science Ltd, The Boulevar, Langford Lane, Kidlington 0X5 1GB, UK, p. 10, 1995.
10. Ref. 9, p. 11.

11. R. W. Schmidley, *Framework for the Control of Quality in Automated Mapping,* unpublished dissertation, Ohio State University, Columbus, OH: 1966.

J. RAUL RAMIREZ
The Ohio State University

GEOMETRIC CORRECTIONS FOR REMOTE SENSING. See REMOTE SENSING GEOMETRIC CORRECTIONS.

sign parameter) and its cross-sectional area t (an independent design variable or decision variable). In particular, then, the capital cost is CLt, where C (a design parameter) is the cost per unit volume of the material making up the line. Also, suppose the operating cost is simply proportional to the power loss, which is known to be proportional to both L and the line resistivity R (a design parameter) as well as to the square of the carried current I (a design parameter) while being inversely proportional to t. In particular, then, the operating cost is $DLRI^2/t$, where the proportionality constant D (a design parameter) is determined from the predicted lifetime of the line as well as the present and future unit power costs