TITLE: GEOGRAPHIC INFORMATION SYSTEM

BYLINE: Michael F. Goodchild, University of California, Santa Barbara, www.geog.ucsb.edu/~good

SYNONYMS: geospatial information system, spatial information system

DEFINITION:
A geographic information system (GIS) is a computer application designed to perform a wide range of operations on geographic information. Geographic information is defined as information about locations on or near the surface of the Earth, and may be organized in a variety of ways. Thus a GIS includes functions to input, store, visualize, export, and analyze such information. Commercial off-the-shelf GIS software is today capable of virtually any conceivable operation on geographic information, and capable of recognizing hundreds of different formats. GISs are used in a wide range of applications, from the management of the distributed assets of utility companies to emergency response. Their scientific applications are found in any discipline that deals with phenomena distributed over the surface of the Earth, from ecology to criminology. Increasingly GIS technology is encountered by ordinary citizens, in the form of map-making sites based on Google Maps, wayfinding sites such as MapQuest, and hotel-finding sites such as Expedia.

A wide range of introductory textbooks on GIS are available with various levels of sophistication, covering the fundamentals of representation, analysis, and application. Clarke [1] provides a general introduction; Longley et al. [2] give a somewhat more advanced and comprehensive perspective; and Worboys and Duckham [3] provide a computational viewpoint.

HISTORICAL BACKGROUND:
GIS has several roots, and today's technology represents a convergence among several independent developments. In the mid 1960s the Government of Canada developed the Canada Geographic Information System (CGIS), the first to use the term GIS, as a means of generating products from the Canada Land Inventory. CGIS implemented a very small set of functions, including map overlay and area calculation, and in its initial form included no capabilities for map output, yet it solved many fundamental problems. Another set of developments centered around the desire to automate the process of making paper maps, and yet another was stimulated by the need to manage complex sets of data in support of transportation planning and census data collection. By the late 1970s the commonalities in these various threads had been recognized, and the first commercial software products began to emerge. Foresman [4] provides a comprehensive history of GIS.

The early efforts to develop GIS were based on idiosyncratic data models and formats. By the late 1970s the value of relational databases had been recognized, and GIS developers had begun to represent complex geographic features as interrelated collections of elements. A map of counties and their attributes, for example, could be represented as tables of nodes, edges, and faces; and the same approach could be used to represent a road network or a map of land cover types. Nevertheless early relational database software was not easily adapted to the storage of the variable numbers of vertices needed to represent each edge, and so a hybrid approach had to be adopted in which some information was stored in a relational database and the remainder in a unique, proprietary structure. By the late 1980s major problems of lack of interoperability began to impact the industry and led to the formation of the Open Geospatial Consortium.

By the 1990s GIS designers had begun to adopt object-oriented database principles [5], and database technology and computing power had advanced to the point where hybrid architectures were no longer necessary. Greater flexibility in data modeling had opened the possibility of applying GIS tools to phenomena that were never traditionally associated with maps, such as events, transactions, movements and flows, and dynamic phenomena in general. GIS is no longer seen as a container of maps, but as an engine for the representation and analysis of spatio-temporal phenomena. The only limitation is that the information they contain be tied to specified locations on or near the surface of the Earth.

Figure 1 shows a screen shot of a typical application of a contemporary GIS, in this case ESRI's ArcMap.

[Figure 1 about here]

SCIENTIFIC FUNDAMENTALS:
Two fundamentally distinct ways of conceptualizing spatial variation are recognized in GIS. In the discrete-object view, the Earth's surface is analogous to an empty table-top, on which are distributed a countable collection of features. The features may overlap, but between them is emptiness. The features might be represented as points, lines, or areas; and in some cases the third spatial dimension will be important and features may be represented as volumes. In the continuous-field view variation over the Earth's surface is described by a series of functions f of location $\mathbf{x}$, where f may be a measure, as in the case of elevation or temperature, or a class or name, as in the case of maps of soils or counties. Location $\mathbf{x}$ may have two or three spatial dimensions, and may also include time in the case of dynamic phenomena. Phenomena distributed over networks, such as vehicle densities or pavement quality, can be conceptualized as fields distributed over a one-dimensional space that is in turn embedded in two- or three-dimensional space.

These conceptualizations can be implemented in GIS in two ways, as raster or vector structures. In a raster structure the set of possible locations is finite, being defined by a grid that is normally rectangular (though more complex grid geometries are needed when the Earth's curvature is important). In a vector structure, every feature is located using an appropriate number of coordinates. Areas and lines are normally represented as ordered sets of coordinates connected by straight lines, and known as polygons and polylines respectively. These methods are straightforward in the case of discrete objects, but the representation of continuous fields requires another step of discretization, with six commonly used alternatives in the case of two-dimensional variation: 1) a set of irregularly spaced sample points, as used for example in capturing and representing weather observations; 2) a set of regularly spaced sample points, as used in representing terrain; 3) representation of the isolines of the field, as used in contour maps; 4) a triangular mesh, with linear variation within the triangles and continuity of value across triangle edges; 5) a set of rectangular raster cells, as used in capturing and representing images; and 6) a set of irregularly shaped, non-overlapping, and space-exhausting areas represented as polygons, as used in mapping land cover types or aggregated census data.

Whether raster or vector, any GIS data set must be referenced to the Earth's surface using some form of coordinate system. In the case of raster data this is usually achieved by registering the corner points of the raster; while in the case of vector data every feature is independently positioned in the coordinate system. Many alternative coordinate systems exist, including latitude and longitude plus numerous systems based on projections of the Earth's curved surface onto a plane. Some have been officially adopted by countries, such as the U.K.'s National Grid, while others have been adopted by international agreement through agencies such as NATO, or by individual states in the case of the U.S. State Plane Coordinate system. Unfortunately this leads to complexity and lack of interoperability, and any GIS must offer a comprehensive set of functions for managing coordinate systems. Moreover the definitions of latitude and longitude are not entirely standard, being dependent on the mathematical figure adopted to represent the shape of the Earth, otherwise known as the datum, which adds another dimension of complexity. One approach to dealing with interoperability has been through the development of metadata standards, or standardized descriptions of data sets that include specification of such properties as coordinate systems. Metadata is now widely used by portals, libraries, and warehouses of geographic data that offer massive information resources to users over the Internet.

A GIS database is normally conceptualized as a collection of layers, each registered to the Earth's surface. A layer may contain the representation of a field, such as a map of land cover type, or it may contain a collection of discrete objects, such as buildings. This layer-based view of the world is often used as an icon, since it is so fundamental to GIS, and has appeared on the cover of more than one textbook. Raster-based layers will likely describe the geographic variation of only a single property, such as county name or land cover type, and because this property will exist for every cell and will have only a single recorded value per cell there is clearly a natural affinity between raster layers and continuous-field conceptualizations. Vector-based layers on the other hand may represent either collections of discrete objects with any number of associated attributes, or the variation of a single property conceptualized as a continuous field. Unfortunately today's GIS designs do not recognize the field/object distinction, with the result that no automatic procedures exist to avoid violating the constraints that apply in the field case. For example, a GIS will treat a set of digitized isolines representing a field as if they were discrete, independent polyline objects, and will not object if a user attempts to edit them in such a way as to make two isolines cross.

Many GIS applications employ some form of tiling, in which the geographic area covered by the database is divided into smaller units, often in the interests of computational efficiency. Tiles correspond to the map sheets of cartography, and are often defined on a rectangular basis. In some cases they may be evident to the user, who must develop explicit strategies for handling queries or analyses that involve multiple tiles; in other cases the tiling scheme may be transparent. In addition, a GIS will employ one or more systems of indexing, again in the interests of computational efficiency. Van Oosterom [6] provides a comprehensive survey of spatial database indexing options.

GIS implementations vary from the stand-alone desktop system designed to support the work of a single user, to departmental solutions that integrate the work of several people around a common database, to enterprise-wide solutions that organize the entire operation of a company or agency around its GIS. Increasingly GIS functions are available on-line from servers, either based on data held at the server, or on data supplied by the user. For example, many Web sites now offer the GIS function of geocoding, which takes a list of postal addresses as input and returns their corresponding geographic coordinates.

A wide range of GIS software products are available to match these various implementation strategies, ranging from versions for hand-held devices through desktop systems to server-side GIS. The use of component-based software architectures has allowed vendors to build different products from the same collection of discrete elements. The trend in recent years has been away from the single-user desktop to integrated client-server architectures, and increasing reliance on data downloaded from Internet portals. Open-source GIS is growing steadily in popularity, and a number of low-cost options have recently appeared to challenge the marketplace dominance of the industry leaders, ESRI and Intergraph. Idrisi, originating from Clark University, continues to provide a software option with very strong support for decision making. Niche products also exist, particularly in transportation where the products of Caliper offer some powerful capabilities, and products from developing countries, particularly China, are increasingly competitive.

KEY APPLICATIONS

As noted above, the earliest applications of GIS were in isolated domains: land resource management; automated cartography; and transportation. By the late 1970s a broader vision of GIS had emerged as a general-purpose software application that could be used to solve a vast range of problems, all dealing in one way or another with the surface and near-surface of the Earth. Nevertheless, the first software products that emerged in the early 1980s found their most important applications in resource management, and were heavily adopted by forest-management companies and agencies, where they were used for compiling inventories, planning harvesting, and management of road access.

This viewpoint largely ignores the role of the military in GIS development, however. By the late 1950s the U.S. was heavily engaged in classified programs aimed at acquiring imagery from space, and in developing the systems needed to assemble, interpret, and analyze their products. Civilian satellites were first deployed in the early 1970s, and led quickly to a generation of GIS software, largely raster-based, that found useful applications in agriculture. Today the military and intelligence agencies are among the heaviest users of GIS, and many weapons systems and war-fighting strategies depend on ready access to current, digital geographic information.

One of the most successful areas of GIS application has been in facilities management, and more generally in the management of distributed assets. Public and private utilities, planning departments of governments, transportation agencies, and managers of complex facilities such as university campuses, will all use GIS routinely to inventory assets, manage and schedule maintenance, and address problems. Investments by individual companies may run to the millions of dollars, and involve hundreds of GIS workstations.

Of rapidly growing importance are GIS applications that impact the daily lives of the general public. These include route guidance, provided perhaps by an in-vehicle navigation system or by a Web service, and map-based services such as Zillow that provide current information on property values. The general public is increasingly familiar with GIS data sources, including remote sensing, and with concepts such as geo-tagging that permit information in sources such as Wikipedia to be georeferenced and thus linked to maps. The term neogeography is being used to describe such novel applications of GIS.

FUTURE DIRECTIONS

The days when GIS was an exotic computer application familiar only to a small elite, and accessible only after extensive training, are long gone; today virtually everyone with access to the Internet is familiar with at least some of its capabilities. The meaning of the term itself has become confused: some would reserve it exclusively for the single-user desktop application, while others would extend it to cover what is now a vast array of activities, for which the adjective geospatial has emerged recently as an alternative umbrella term. In discussing future directions, then, it is possible to take either the first, somewhat narrow perspective, or the second broader one.

From the narrow perspective, GIS software will continue to evolve, particularly in the direction of greater support for dynamics. Some application areas are currently in their infancy, and will likely grow in importance in the next few years; they include human health, the understanding of disease transmission and diffusion, and the assessment of health service outcomes; and business, whose mantra “location, location, location” feeds directly into GIS applications. Related areas of logistics, real estate, and insurance are also ripe for greater use of GIS.

A rich research agenda has evolved over the past decade, and will likely produce results that will in turn impact future generations of GIS software. The University Consortium for Geographic Information Science has taken the lead in this context, with short- and long-term agendas that address topics ranging from spatial cognition (How can GIS interfaces be made easier to use?) to the social impacts of GIS technology (Who gains, who loses?). The topic of uncertainty remains a thorny issue, with much useful research completed on the sources and management of uncertainty in GIS but little adoption in mainstream software. As a result, the user is left with little choice but to accept the results of GIS analysis at face value, while knowing that the data on which they were based is inevitably imperfect and uncertain.

From the broader perspective, GIS functions will continue to become more available in Web services. In the past year a very interesting series of developments have occurred that are enabling citizens to create their own geographic information, and to integrate and publish it on the Internet. Sites such as OpenStreetMap are enabling communities to create their own maps; Wikimapia is collecting descriptions of features on the Earth's surface from vast numbers of volunteers; and Flickr is assembling a vast collection of geo-referenced photographs. This process of volunteering geographic information is having a profound impact on our knowledge of the planet, and on the engagement of citizens with geography and GIS.

CROSS REFERENCES: Data Models, Field-Based Spatial Modeling, Spatial Indexing Techniques, Spatial and Multidimensional Databases, Spatial Data Analysis,

RECOMMENDED READING

1. Clarke K. C. (2003): Getting Started with Geographic Information Systems. Prentice Hall.
2. Longley P. A., Goodchild M. F., Maguire D. J., Rhind D. W. (2005): Geographic Information Systems and Science. Wiley.
3. Worboys M. F., Duckham M. (2004): GIS: A Computing Perspective. CRC Press.
4. Foresman T. W., ed. (1998): The History of Geographic Information Systems: Perspectives from the Pioneers. Prentice Hall.
5. Arctur D., Zeiler M. (2004): Designing Geodatabases: Case Studies in GIS Data Modeling. ESRI Press.
6. van Oosterom P. (1999): Spatial Access Methods. In Longley P.A., Goodchild M.F., Maguire D.J., and Rhind D.W., eds., Geographical Information Systems 1: 385-400. Wiley.

FIGURE CAPTIONS

1. Screen shot of ESRI's ArcMap GIS. Three layers are displayed, as listed in the upper left: the county boundaries of the U.S. (black), the railroads (blue), and the interstate highway network (green). Pull-down menus and toolbars provide access to a wide array of manipulation and analysis functions, and many more are available as third-party extensions. The display uses the Lambert Conformal Conic projection.