Michael F. Goodchild

Directions in GIS

Introduction

The purpose of this paper is to discuss research directions in GIS. In doing so I will deliberately broaden the meaning of "GIS" to include a wide range of activities within the broad rubric of digital geographic information, since it seems to me that "GIS" is now generally used in that broad sense, rather than in the narrower sense of a software system designed specifically to store, retrieve, and analyze existing geographic information (Maguire 1991). We seem to be reaching a point where digital technology is encountered in almost all aspects of the communication of information, and the same is true of geographic information. It no longer seems important to ask whether information is digital or not-- rather, the important questions seem to concern the degree to which the digital format imposes itself on the information, forcing us to modify, transform, or otherwise alter the information in order that it can be handled and communicated in digital systems.

GIS research is now very broad, ranging from investigations into how people think and reason with geographic information, so that better systems can be designed that are easier to use, to studies of the legal and intellectual property issues raised by widespread sharing of geographic information. The work of the U.S. National Center for Geographic Information and Analysis has encompassed much of this range (for more information on the work of NCGIA see http://www.ncgia.ucsb.edu), as have recent conferences on GIS research such as the International Symposia in Spatial Data Handling (Waugh and Healey 1994), the international Symposia on Spatial Databases (Egenhofer and Herring 1995), and the Conferences on Spatial Information Theory (Frank and Kuhn 1995). In this paper I can hope to cover only a small part of that range, and the subjects I discuss in what follows reflect certain obvious personal biases, toward NCGIA, the University of California, Santa Barbara, the Alexandria Digital Library project (http://alexandria.sdc.ucsb.edu), and more generally the U.S. University Consortium for Geographic Information Science (http://www.ucgis.org), and the European Science Foundation's GISDATA program (http://www.shef.ac.uk/uni/academic/D-H/gis/gisdata.html).

The paper begins with the concept of a life cycle for geographic data, and discusses the changes that are occurring at various stages along it. The data life cycle is a convenient organizing mechanism for current GIS, given the recent trend to extend digital geographic data handling to all aspects of the cycle from initial observation through to eventual archiving. I then provide a description of the Alexandria Digital Library, an effort to provide the services of a map and imagery library over the Internet, and to exploit the power of geographic location as a means for organizing information. Subsequent sections explore some of the general issues raised by Alexandria, in the areas of metadata, information granularity, and scale. The transition to a digital world is causing us to question many of the traditional ways of doing

things, and to ask what aspects of this legacy should be preserved, and what abandoned. The paper then moves to a discussion of the major impediments to GIS, and directions in which GIS might evolve to avoid them, specifically in the context of data models, which I argue lie at the root of GIS capabilities. This leads to a discussion of interoperability in GIS, which is one of the keys to improvement in the ease of use of the technology.

The data life cycle

A traditional view of GIS, reflected in much early writing in the textbooks of the field, is that its purpose lies in building a digital representation of some existing set of geographic data, thus allowing the data to be subjected to analysis. Spatial analysis is often seen as the primary purpose of GIS, and the results of analysis as being used as the basis for decisions (Cowen 1988). In this view, the key elements of GIS include the ability to convert to digital form, store, manipulate in a fashion analogous to the calculator, and report. Much has been written about the kinds of functions needed to support these activities (Maguire and Dangermond 1991), and on related issues such as the accuracy of each stage. Comparisons have been drawn between these functions of spatial analysis and those of a statistical package--it has been said that GIS is to spatial analysis as a statistical package is to statistical analysis (Goodchild 1987). Underlying the entire framework is the notion that the value of computers lies in their ability to process numeric, and sometimes alphanumeric, information much more rapidly than can a human being. Indeed, the initial Canada Geographic Information System, one of the earliest GIS, had no capabilities at all for visual display, since its intended purpose of numerical analysis of maps could lead only to tabular, numeric output (Tomlinson 1967).

In this view, the source of information for GIS is commonly the map, and the conversion of its contents from paper to digital form through digitizing and scanning is a key GIS operation and one that is ultimately very costly. But once in digital form, the high speed with which the data can be processed normally provides the benefits needed to justify the conversion. However, the few minutes needed to process the data compare poorly to the many months that are often required to prepare the data for processing. Ironically, therefore, GIS-based projects when considered in their entirety are often significantly slower than ones completed using more traditional methods.

Several factors have caused this view of GIS to change in the past five years, and have led to its broadening to include many more data-related activities. First, GIS is no longer an activity confined to the desktop. It is now possible to obtain sophisticated processing power and storage in a portable device weighing less than a kilogram, and to think of GIS as an activity that can take place in a car, or almost anywhere. The concept of "field GIS" is now exemplified by many new applications that have emerged in the past five years as digital technology has become smaller and more compact. GIS is now the basis for the new technology of precision agriculture, a term used to describe the use of geographic information technologies to improve greatly the precision with which land is farmed. Rather than make decisions about entire fields, farmers can now determine appropriate levels of seeding, fertilizer, and pesticide applications based on detailed information collected from continuous monitoring of crop yield by harvesters, air photographs, and detailed soil surveys. Precision agriculture is favorable to the

environment, because levels of application of pesticides and fertilizers can often be reduced based on detailed knowledge of soil conditions; and it is also economic, since the improved yields and cost savings are both reflected in improved income to the farmer.

Field GIS is also emerging in the form of new software designed to support collection of data rather than its analysis. The "field notes" form of GIS often includes a "backcloth", typically a georeferenced image, on which additional information can be located by the field scientist. The software is installed in a portable computer, often with an associated GPS receiver for accurate geopositioning, and maps and sketches created by the observer can be uploaded to a conventional desktop GIS either by wire or by radio communication. Such systems are now commonly used to collect information from forestry transects, or in routine maintenance of utility networks. Another form of field GIS, installed in a van and coupled to kinematic GPS, can be used to produce accurate surveys of road networks and to capture associated digital pictures of the state of road pavement or signage (Novak and Bossler 1995). In all of these examples field GIS is allowing us to see GIS more as a tool for data collection and the construction of geographic databases than for their analysis--the database becomes the product of GIS, rather than its input.

Second, GIS is rapidly becoming a technology that impacts not only our professional lives, but our everyday ones as well. Rather than being confined to sophisticated tasks of inventory and analysis, GIS is now appearing in the form of simple mapping capabilities attached to spreadsheets, as exemplified by the recent extensions of Microsoft Excel into simple cartography. Advertisements for simple GIS products appear in airline magazines, and GPS units are available in electronics stores. Car rental agencies now offer in-car navigation systems. In the next few years, we can expect many other applications of GIS for the mass market to appear, including systems that automatically map the locations of accidents based on explosions of car air bags, and trigger appropriate emergency response.

This interest in GIS is being fuelled in part by the availability of low-cost hardware, software, and data; and in part by a growing level of interest in geography, geographic information, and the kinds of thinking and reasoning processes associated with space. It is said that a picture is worth a thousand words; and that maps store and communicate information much more efficiently than text; and that the human eye and brain are marvellously adapted to rapid analysis of visual presentations.

A third point concerns the role of digital geographic technologies in communication. In the traditional data life cycle, technology can be seen as a means of transmitting the knowledge gathered in the field to its eventual users. Decisions must often be made based on knowledge gathered in the field, in circumstances where field scientist and decision-maker are different people, separated by great distances and also perhaps by differences of discipline and experience. In such circumstances the map or spatial database becomes the communication channel, and its contents the means by which knowledge is transmitted. The paper map is a very effective means of communicating certain types of information, but it can severely restrict the communication of other types. For example, paper maps require us to flatten the world; they confine us to a fixed, uniform resolution; and they tend to force information that is frequently vague or fuzzy into precise categories. Such restrictions are a major constraint on the traditional data life cycle, since maps have been the primary repository and communication medium for

geographic information. In the digital era, there is the potential to remove many of these constraints, and thus to widen the communication.

The fourth factor affecting the data life cycle concerns data access, and the changes occurring in the architectures of computing. Unlike the early days, when computing was limited to mainframes, and later developments of interactive access via "dumb" terminals, today's computing environments include local and wide area networks (LANs and WANs), the Internet, high-bandwidth communication channels, clients and servers. Such new architectures allow radical restructuring of the various stages and roles in the data life cycle. With client-server technology, it is possible for the originator of the data to become its permanent custodian, thus ensuring that it is as current as possible, and that no problems arise through the existence of different versions in different locations. Data can also be documented and described at its source, and the documentation updated and maintained digitally through each stage in the data's transformation.

With current object-oriented technology, it is possible for important processes to be encapsulated with data. For example, it may be useful to encapsulate suitable display software, so the potential user can see the data in spatial form. Much data gathered at sample points, such as weather stations, will frequently need to be interpolated for analysis, and encapsulation allows its originator to make decisions about how best to interpolate. Finally, processes must be encapsulated with data if the goals of interoperability are to be achieved, since data must be capable of describing itself, and how it should be processed, to a variety of host systems.

In summary, there are many reasons to believe that geographic information technologies will play a role at every step in the life cycle of geographic data, not just in its analysis. The data life cycle concept allows us to see the entire process of data handling, from creation to archiving; to ask questions about the efficiency of communication of knowledge; to look at the roles of various participants; and to ensure that information is passed as effectively as possible between people and techologies that traditionally may have found it very difficult to share and communicate because of differences in terminology, formats, and disciplines.

The Alexandria Digital Library

Many of these issues can be illustrated within the framework of the Alexandria Digital Library project (ADL). Named for the classical library of Alexandria, ADL aims to build a distributed digital library for spatially referenced information, including services for constructing the library's collection (ingest), cataloging, retrieval, and use. All of these functions are to be distributed, and accessible via the Internet. We define spatially referenced information as any information referenced to a two-dimensional frame, of which geographical reference to the surface of the Earth is ADL's primary example. While one can think of ADL as an effort to build a map and imagery library on the Internet, its objectives go further in seeking to exploit geographic location as a method for indexing and retrieving information; and to make maps and images part of the mainstream of the digital library of the future.

Although the idea of a digital library may seem far-fetched, the technology already exists to store the entire collection of humanity's published text in digital form--and almost all text now published is available in digital form. On the other hand maps and images are comparatively voluminous, and a single day's output of the EOS generation of satellites may exceed the information content of many of today's research libraries. Nevertheless, the advantages of universal access to library services via the Internet are very attractive.

ADL is one of six digital library projects being funded by the U.S. National Science Foundation (NSF), the Advanced Research Projects Agency (ARPA), and the National Aeronautics and Space Administration (NASA). The concept of ADL was developed at the Map and Imagery Laboratory at the University of California, Santa Barbara (UCSB), over several years, leading to major funding beginning in October 1994. ADL has many participants and partners, including the Departments of Geography, Computer Science, and Electrical and Computer Engineering at UCSB; the Library of Congress and the U.S. Geological Survey; Digital Equipment Corporation; Environmental Systems Research Institute (ESRI); and many others. Projects like ADL provide exciting opportunities for the GIS community, because they offer links to the vast resources and experience of libraries and library science.

During the first six months of funding, ADL developed its Rapid Prototype (RP) based on existing technology, and designed to act as a straw-person for further discussion. The RP is a standalone system, based on ESRI's ArcView, the interface language Tcl/Tk, and the database management system Sybase. The RP was completed in March, 1995, and has been distributed on CD for evaluation. Following the RP, ADL moved to the development of a universally accessible server implementation based on the Internet and the World Wide Web. A Web prototype was developed by November 1995, and is currently in beta release (for details, see http://alexandria.sdc.ucsb.edu). The number of digital data sets now accessible via the Web prototype is on the order of 105.

The strategy used in the design of the Web prototype is based on some simple assumptions about the nature of searches for spatially referenced information. Almost all such searches start by specifying location; subsequently, the search is refined to certain subjects or themes, ranges of dates, data formats, and levels of geographic detail. Thus a user might request information on Taiwan, covering surficial geology, as recent as possible, in ARC/INFO format, and at a scale of 1:100,000 or better. To accommodate this search strategy, the user of the Web prototype first encounters a world map (other base maps can be used as appropriate). Pan and zoom features are available to focus attention on any area, and additional features and detail become visible on the world map as the user zooms. A query area can be defined, and used to search the ADL catalog for suitable data sets. Other features allow the query to be restricted to certain dates, themes, data formats, etc.

ADL has adopted a standard format for its catalog that is as compatible as possible with both library and GIS communities. On the library side, we have made the format compatible with U.S. MARC, a standard for digital catalogs. On the GIS side, ADL has adopted a subset of the Federal Geographic Data Committee's (FGDC) Content Standard for Geospatial Metadata (http://geochange.er.usgs.gov/pub/tools/metadata/standard/metadata.html), which specifies some two hundred possible descriptive fields. The idea of using a "core" set of metadata elements in the digital library context rather than an exhaustive

description is now widely recognized (see, for example, http://www.oclc.org:5046/conferences/metadata/dublin_core_report.html).

Although both the RP and the Web prototype allow the user to specify the location of a query, typically as a "bounding box" consisting of two latitudes and two longitudes, most users of map libraries specify their needs not in coordinates but through place-names. The great flexibility of place-names, and the hierarchical structure they imply, have made them very difficult to use as bases for library catalogs, and this is one reason why maps and images have never been part of the library mainstream. For example, it is possible that a query about Santa Barbara might be satisfied by information cataloged as "Southern California", or as "Goleta", a suburb of Santa Barbara. These horizontal and vertical linkages make it very difficult to access information by place-name.

In the ADL Web prototype, we have implemented a large gazetteer, of about 5 million entries (a gazetteer is defined as an index linking place-names to geographic locations, with place-names as the primary key; most atlases include gazetteers linking place-names to pages in the atlas). This allows the user to search by place-name, to select among various places of the same or similar names, and to see the basemap positioned and scaled appropriately.

The use of the gazetteer gives the Web prototype several powerful features. Besides the ability to specify the location of search by name, the user can also achieve a certain level of "content-based search", or search of the actual contents of the data set rather than the contents of its documentation or metadata. Thus it is possible to search for information on Santa Barbara even though the words "Santa Barbara" do not appear in the data set's catalog entry--the simple geographic intersection between the location of Santa Barbara, as determined from the gazetteer, and the data set's bounding box, are sufficient. The gazetteer can also be used to broaden or narrow a search. For example, by determining from the gazetteer that Santa Barbara is within the State of California, one might broaden a search for information on Santa Barbara to data sets covering the entire state--and similarly search can be extended to Santa Barbara's geographic neighbors if these can be determined from the gazetteer.

There are many problems associated with implementing a gazetteer. First, and perhaps most problematic, is the lack of information on feature extents, since most gazetteers provide only a point location for a feature. This is acceptable for a small, point-like village, but is inadequate for extended features such as long rivers or large political entities. Standard protocols for such features, such as the use of the river mouth as representative point, are not sufficient for the purposes of ADL. It would be impractical to add geometric extents to every feature in a large gazetteer. Instead, we are using various rules, in conjunction with available information such as feature type, to try to infer feature extent in as many cases as possible, and in other cases inviting the user to make appropriate decisions and choices.

While we are far short of having an operating digital library, projects such as ADL allow us to see aspects of what a digital library might offer, and to anticipate some of its problems. Many impediments now stand in the way of offering the full services of a library over the Internet, and one that provides access to all types of information irrespective of the somewhat restrictive legacy of the traditional

library. In the next sections I discuss three of them: metadata, the issue of data granularity, and the specification of level of geographic detail.

Metadata

Metadata is normally defined as data about data--the information that allows us to find, handle, browse, read, and understand the contents of a data set. The concept of metadata embeds a range of metaphors, from the library card catalog, through the written data documentation needed by early generations of computer technology, to the handling instructions that appear on the outside of a mailed package. All three are valid, since metadata must perform elements of all three functions. Metadata must be digital, must travel with or ahead of the data, and must be preserved as the data are transformed by various kinds of processing, and if possible updated appropriately. For example, a change of projection process should cause a corresponding modification of the associated metadata.

While all three metaphors convey a sense of the functions inherited by metadata from the pre-digital era, none of them reflect a world that truly takes advantage of digital technology. One of the most significant advantages of digital technology is the opportunity to rethink the data life cycle, and to add to the value of data by making it shareable and useful to much larger populations of users. In this context metadata provides the opportunity to explain the meaning of data, allowing users who may have had nothing to do with its creation, and may not be familiar with the terminology and practices of the community that created it, to put it to use. Thus Francis Bretherton has defined metadata as "that which makes data useful".

In this sense, the value of metadata to potential users is related to the existence of a common language, and to the uses to which the data will be put. A user who needs information on the positional accuracy of a data set is not necessarily helped by knowing the serial number of the digitizer on which it was created, or the date, unless these items of information can be translated into terms understood by the user, such as the 90th percentile of positional error. Items such as the serial number or date thus reflect a producer's approach to metadata, where the main value of such information is for production control, rather than the needs of the user.

In essence, the metadata needed to support digital libraries must serve as the basis for a dialog between producer and user, in a common language understood by both. It must reflect as much the skills and experience of the user as of the producer. Since the skills of library users range widely, from "spatially aware professionals" (SAPs) to elementary school students, metadata must be able to present many different faces.

In ADL, we are experimenting with studies of the process that occurs when a user enters a map library, in order to gain understanding that can guide the design of the digital library. Ethnographers are recording the process on videotape for analysis, and are finding that ADL is already creating its own culture, with a language that is distinct both from that of the traditional map library and that of the computerized library card catalog.

We have concluded that the only feasible approach to metadata in this context must be hierarchical. SAPs will want to search the library using very cryptic languages that convey very dense meaning; elementary school children use much less compact languages. Correspondingly, ADL must support a hierarchy that ranges from the extreme detail of the FGDC metadata standard to a small number of "core" metadata elements likely to be understood by all. This is a very different approach to those of card catalogs, data documentation, and single-tier metadata standards, but it is necessary if the library is to be as useful as possible to as many as possible.

Granularity

In the traditional library, with its stacks of books, information is organized around the bound volume as the fundamental unit. Very few elements of traditional library catalogs address either groups of books, or portions of a book's contents; the discrete enumeration and cataloging of volumes takes advantage of the physical nature of the book as a unit. Only in the case of serials is there any significant extension to a more general concept of information granularity. Moreover, this same principle can also be applied to a map library's map sheets, standard-sized images, and atlases.

In the digital world, these concepts of granularity are no longer as easily followed. The contents of a single topographic map sheet are likely to be stored in several files, each corresponding to a layer; the full representation of a standard 1:24,000 U.S. Geological Survey topographic map may run to tens of files in certain GIS representations, and there is little reason to store them in one aggregate. Equally, the digital world allows us to escape from the constraints of map sheets, and to explore the feasibility of a continuous, "seamless" view of the world. Why should there be 50,000 individually cataloged 1:24,000 quadrangles, rather than one seamless data set, and why should Landsat imagery be stored in awkwardly shaped "scenes"? Other problems of granularity arise in the case of maps embedded in atlases and books, or geographic data sets in CD collections.

In short, the 1:1 model of the card catalog--one card per book--is overly restrictive and constraining, and not defensible in the digital era. Instead, we need to develop concepts of information abstraction and generalization that span the entire range from the individual feature, through the feature class and map sheet, to the seamless layer or mosaic. The concepts of data and metadata are no longer as separable as they were in the past.

Level of geographic detail

Of the four main dimensions of a search for spatially referenced information, it is the level of geographic detail that has caused us the greatest difficulties. Traditionally, the cartographer has defined the level of detail shown on a map largely through its metric scale: a scale of 1:24,000 defines for the U.S. Geological Survey not only the ratio of distance on the printed map to distance on the ground, but also the positional accuracy, the set of features depicted, and the level of cartographic generalization applied to them. But metric scale has no meaning for data in digital form, since unlike paper maps there are no

distances to be measured in a digital computer's storage.

Although metric scale has no meaning, the other properties related to scale survive the conversion to digital form, including positional accuracy, content, and generalization. Thus metric scale has proved useful, and is often specified in metadata. We know, for example, that data labeled "1:24,000" will show most city streets and name some of the most important of them, but will not include most individual buildings. It will show many streams and creeks as single lines, and will have a positional accuracy consistent with the National Map Accuracy Standards (NMAS) for 1:24,000 data, that is, a 90th percentile of error of approximately 12m. But since the four scale-related elements are now decoupled, there is no way of knowing what it means to say that a U.S. Geological Survey Digital Orthophoto Quad (DOQ) has a "scale" of 1:12,000. Clearly this is not the metric scale. Nor is it indicative of spatial resolution, which is 1m for these data sets, or of the features contained, which are not explicitly identified in this raster data. It turns out that the reason for specifying a "scale" of 1:12,000 has to do with the positional accuracy of a DOQ, which is about 6m, and thus compatible with the NMAS for maps of that scale. But this kind of reasoning is largely arbitrary, and very confusing to inexperienced users, who might ask why positional accuracy is not simply specified in m.

For SAPs, it makes sense to specify level of geographic detail through as many of its dimensions and indicators as possible--a DOQ's metadata record should include its positional accuracy, spatial resolution, and content, as if these are largely independent properties. On the other hand inexperienced users may be happier with a looser definition that exploits a suitable metaphor. The Microsoft Encarta Atlas specifies the scale of a display by using the metaphor of a human eye positioned at some distance above the surface of the Earth--descent produces more detail, and ascent produces less. Of course this is not precise, and it says little about positional accuracy, but it may be fully adequate for the inexperienced user who has not thought much about geographic data and its fundamental concepts.

Summary

In summary, ADL provides us with a prototype or straw-person which we can use to explore some of the implications of digital library technology in the specific context of maps, images, and other spatially referenced information. One of its most powerful concepts concerns the use of geographic location as a means for organizing and retrieving information. In the traditional library geographic location has played little role in search, for several reasons. In fact, one can find information by its location only if a suitable place-name appears in one or more of the author, title, or subject of a book, and then only by some form of keyword search. This is very restrictive--it means, for example, that one cannot use the catalog to find information on Paris in a guidebook cataloged under "France". In the form of search, browse, and retrieval implemented in a digital library such as ADL, however, the geographic key becomes one of the most useful, particularly for information that has a geographic "footprint". At this point we are restricted to footprints that are both "crisp" and singly-bounded, and force them still further into the form of bounding boxes. But in future versions of ADL we expect to be able to deal effectively with "fuzzy" and multiple footprints, and to link information through a common geographic key. In a world that is increasingly in need of new ways of organizing its ever-growing information base, such

extensions offer great potential.

The importance of legacies

Technologies like ADL are caught in a fundamental contradiction, between the need to create digital environments that seem familiar, by making use of familiar metaphors, on the one hand, and the need to exploit fully the power of the digital environment on the other. For example, it may be important to provide the "look and feel" of the card catalog--but this is basically incompatible with the ideas expressed in the previous section. Old ways of thinking provide comfort in the digital world, and GIS and ADL are full of them--GIS is often explained as "having a map in a computer", and digital libraries as the digital equivalent of books on shelves. In ADL and in GIS generally we face the need to find a balance, by using the legacy of previous technologies to provide familiar landmarks, while trying to escape the constraining aspects of legacies. Sometimes this dilemma is resolved by selecting one direction over the other, as when automated cartography is used to produce maps that are indistinguishable from their pre-digital parents. In other cases new technologies exploit new concepts to such a degree that they appear hostile to the user, and require lengthy training prior to use.

This argument is usually presented in the context of the transition from analog to digital, but it can apply equally to the replacement of old digital technologies by new ones. Early versions of GIS technology were very restrictive, and we will continue to find new ways of broadening the base of GIS and removing some of the earlier constraints. In this sense GIS will never be a perfect technology, and must continuously reinvent itself in order to overcome the constraints of the past. In the next four sections I present four ways in which I believe GIS can reinvent itself, and indeed must do if it is to continue to grow. Each of these four reflects a current area of GIS research, where recent results suggest the possibility of substantial progress.

One of the great advantages of GIS is its ability to store and manipulate a wide range of types of geographic data. Any such software environment that seeks to provide a range of capabilities and functions is fundamentally driven by the data model that it implements--by the set of options provided by the underlying database for the storage and retrieval of information. Thus a word processor offers a vast assortment of capabilities and functions for written text, but would be of little use for processing raster images. Similarly a spreadsheet implements the functions that operate on simple rectangular tables of information. A data model is defined as the set of entities and relationships used to build a representation of some real-world phenomena, and for GIS these are the entities and relationships needed to represent variation on the Earth's surface.

A growing literature has drawn attention to the role of the data model, and to possible ways in which GIS data models might be extended, made more versatile, or otherwise broadened to support a wider range of GIS operations. But ultimately it is the choice of data model by the GIS designer that determines what the GIS can be made to do. Goodchild (1992) has identified three broad classes of data models--those needed to represent continuous fields, discrete objects embedded in a two-dimensional space, and variation over linear networks embedded in two-dimensional space. But although these three

classes capture almost all data models currently implemented, the level of user interaction is with a lower conceptual level, where geographic variation is represented in terms of geometric objects--points, polylines, polygons, and raster cells--and their attributes and relationships. No less than six distinct data models are used to implement the concept of a continuous field--a raster of cells, regularly spaced sample points, irregularly spaced sample points, digitized contours, polygons forming an irregular tesselation of the plane, and triangulated irregular networks (TINs).

Interoperability between data models

This problem of a plethora of data models addressed in the previous paragraph is one reason why current GIS has the reputation of being difficult to use. For example, a set of points stored in a GIS with associated attributes may represent very different things to the user, as representations of very different data models, and the associated meanings of different functions and processes are different as well. An irregularly spaced set of points may represent a sample of weather stations, and an attribute such as mean temperature may be used to create a continuous surface by using a process of spatial interpolation. But the same function of spatial interpolation would be meaningless if the points represented cities, and the attribute was population, or average income. Because interaction with current GIS commands is almost always with the primitive element, such as the point, rather than with what the point represents, it is necessary for the user to maintain a high level of conceptual involvement with the application. The GIS is incapable of warning the user when something meaningless is being done, or of taking steps automatically when no user intervention is needed.

One way to make GIS easier to use would be to raise the level of interaction to that of the conceptual data model--the continuous field, class of discrete objects, or linear network--rather than the level of implementation as is currently done. One uniform command language could be applied to all six representations of continuous surfaces, for example, hiding the details of the internal representation except when it is necessary that the user be aware of them. This would effectively reduce the number of data models currently implemented to three rather than at least ten, and would greatly reduce the complexity of current command languages and scripts.

For example, consider a simple request to a GIS to compute the mean rainfall for each U.S. state. Rainfall is conceived as a field, with a single value of the variable at each point in the U.S., and with the variable measured on a continuous (ratio) scale. Similarly "state" is conceived as a discrete or multinomial or nominal field, with a single value at each point in the U.S., measured on a scale which has only 50 values (51 if the District of Columbia is included) and in which value conveys no sense of order, ratio, or difference. At this point the query is fully defined, and there is no need for the user to supply any additional information for it to be resolved, assuming of course that the necessary data is available.

Current GIS vastly complicates this situation, by requiring its user to specify significantly more. We need to know whether the "state" coverage is raster or vector, and how the rainfall data is stored. Operations to convert raster to or from vector may have to be implemented, followed by the overlay of

two layers, and a cross-tabulation. The associated script may stretch to ten or even 20 statements. But all of these operations can be anticipated, and none add to the specification of the original query. Parameters such as the cell size and numbers of rows and columns should be available in the associated metadata, though at this time there is very little implementation of metadata in current GIS. With a system such as this we would expect the results to include an estimate of the uncertainty associated with each value of rainfall, and again this should be computed without further intervention or specification by the user.

The research needed to allow this development to occur is largely complete, and experimental systems of this nature have already been built. Before they can be implemented, however, we will need to confront the kinds of issues raised earlier in connection with legacies--such systems will lack many of the familiar landmarks of GIS, including the function of "overlay", which is often regarded as the most important element of the GIS toolkit, but which is fully redundant under the scenario sketched here.

Expanding the domain

I argued earlier that data models ultimately constrain the degree to which software functionality can be integrated in any one package. Because of their historical roots, GIS data models emphasize the representation of maps and images, while word processing data models emphasize linear text, and spreadsheets focus on numeric tables. In this sense they have inherited many of the characteristics of paper maps. While it is easy to develop a full range of functions for this type of data, it is very difficult to extend a data model once it has been chosen as the basis of a package, and thus it has proven difficult for GIS to escape its heritage.

Specifically, paper maps are flat, requiring the use of complex map projections in order to represent the curved surface of the Earth. They are two-dimensional, and it has been difficult to adopt GIS technology to the representation of three-dimensional data, or objects embedded in three-dimensional space, except in cases where the third dimension can be regarded as a function of the other two (e.g., in representation of terrain). They are static, and it has been difficult to add the temporal dimension to GIS. Finally, they are of uniform scale, and it is not easy in GIS to link information across scales.

Early research focused on the representation of each of these extensions of the basic data model. Goodchild and Yang (1992), for example, have discussed structures for the sphere that have some of the characteristics of quadtrees; there is an extensive literature on true three-dimensional GIS (Turner 1992); Langran (1992) has discussed the implementation of time-dependence in GIS; and many solutions have been proposed for handling spatial hierarchies.

The next generation of GIS may take a different perspective on these issues, based on concepts of interoperability. Rather than a single, monolithic package, future GIS is likely to feature interoperable components. Thus three-dimensional capabilities may exist in one package, and may communicate with a traditional two-dimensional GIS by sending and receiving appropriately structured objects, using open object protocols, and through remote procedure calls. The Open GIS Consortium (http://www.ogis.org/

overview.html) is developing the protocols to support this kind of interoperability.

In addition to these technological problems, many other issues combine to create impediments to any expansion of the domain of GIS. While it is easy to imagine a three-dimensional GIS, for example, we lack many of the essential ingredients to a three-dimensional GIS application. There is very little truly three-dimensional geographic data. We lack well-known, exemplary applications, except in limited areas. Finally, there are no familiar metaphors, such as the paper map, to guide our thinking about these extensions. In short, the paper map creates a powerful legacy.

Additional directions

The previous section identified four possible extensions to GIS data models, each of which has been the subject of recent research (see, for example, Molenaar and De Hoop 1994). But this analysis raises the question of whether other directions might also exist. The paper map has a uniform scale, and also a uniform quality of coverage, except where validity diagrams indicate variation in currency or other aspects of quality. Programs such as national topographic mapping have encouraged us to believe that uniformity of quality is a natural condition of geographic data. But in the future it seems unlikely that that assumption will remain true. The U.S. National Spatial Data Infrastructure (NSDI), for example, will be created through a series of partnerships, and its framework data will likely vary in scale, with large-scale data for urban areas and small-scale data for rural. The new satellite sources now coming on line acquire data on command rather than in fixed swaths, so their output is likely to be distributed in the form of mosaics rather than as fixed scenes. Global databases that have been compiled from national sources are also likely to show significant variation in quality.

At this time we have very little in the way of data models for data of varying uncertainty. In such circumstances metadata is likely to take the form of a map, rather than a set of records, more akin to the validity diagram of a topographic sheet than the card catalog. We lack analytic methods for data of variable quality, and have a poor level of understanding of the implications of varying quality for the results of analysis.

Platform interoperability

Finally, I suggest interoperability between platforms as a fourth area for GIS development. Earlier I discussed the potential for interoperability between the six distinct field data models, since each implements the same basic conceptual model of a continuous field. In principle, we ought to be able to extend this argument to the case of data models operating in different computing environments. For example, it ought to be possible to resolve the query "determine the average rainfall by state in the U.S." even though rainfall is stored in an ARC/INFO coverage on a Unix platform, and the state coverage is in IDRISI on a PC running under Windows 95. The same argument applies--the extensive script that would need to be written in today's GIS environment adds nothing to the specification of the original query that could not be resolved automatically by the computing environment without user intervention.

This example serves to emphasize once again the essential role played by metadata in achieving interoperability within the data life cycle. Metadata allows data to describe itself to other systems, in addition to allowing us to find data, share it, and assess its fitness for use. It adds value to data by making it possible for others to use it, thus ensuring better return on our original investment.

Concluding comments

As I noted at the outset, this view of GIS research directions is strictly my own, and I have undoubtedly omitted mention of many interesting research areas that are equally important to the future of the field. The pages of the journals, notably the *International Journal of Geographical Information Systems*, and the proceedings of recent conferences give a much more complete picture.

In this paper I have tried to stress several general and interlocking themes. One is the influence of new thinking, and its relationship to the development of technology--the developments in digital technology that have occurred over the past few years have stimulated an enormous amount of new thinking on the future, and the likely form of the advanced technological society that is now emerging. Another is the transition of GIS from a technology of analysis to one that underlies and facilitates the entire data life cycle, including data collection and library search. A third is the need for GIS to become easier to use, and easier to integrate with other technologies.

Acknowledgment

References

Cowen, D.J. (1988) GIS versus CAD versus DBMS: what are the differences? *Photogrammetric Engineering and Remote Sensing* 54: 1551-1554.

Egenhofer, M.J., and J.R. Herring, editors (1995) *Advances in Spatial Databases: Fourth International Symposium, SSD '95, Portland, Maine, August 6-9, 1995*. Lecture Notes in Computer Science 951. Berlin: Springer-Verlag.

Frank, A.U., and W. Kuhn, editors (1995) *Spatial Information Theory: A Theoretical Basis for GIS: International Conference, COSIT '95, Semmering, Austria, September 21-23, 1995*. Lecture Notes in Computer Science 988. Berlin: Springer.

Goodchild, M.F. (1987) A spatial analytical perspective on GIS. *International Journal of Geographical Information Systems* 1(4): 327-334.

Goodchild, M.F. (1992) Geographical data modeling. *Computers and Geosciences* 18(4): 401-408.

Goodchild, M.F., and S.R. Yang (1992) A hierarchical spatial data structure for global geographic information systems. *CVGIP-Graphical Models and Image Processing* 54(1): 31-44.

Langran, G. (1992) *Time in Geographic Information Systems*. London: Taylor and Francis.

Maguire, D.J. (1991) An overview and definition of GIS. In D.J. Maguire, M.F. Goodchild, and D.W. Rhind, editors, *Geographical Information Systems: Principles and Applications*, Vol 1, pp. 9-20. London: Longman Scientific and Technical.

Maguire, D.J., and J. Dangermond (1991) The functionality of GIS. In D.J. Maguire, M.F. Goodchild, and D.W. Rhind, editors, *Geographical Information Systems: Principles and Applications*, Vol 1, pp. 319-335. London: Longman Scientific and Technical.

Molenaar, M., and S. De Hoop, editors (1994) *Advanced Geographic Data Modelling: Spatial Data Modelling and Query Languages for 2D and 3D Applications*. Publications on Geodesy, New Series, No. 40. Delft: Netherlands Geodetic Commission.

Novak, K., and J.D. Bossler (1995) Development and application of the highway mapping system of Ohio State University. *Photogrammetric Record* 15(85): 123-134.

Tomlinson, R.F. (1967) *An Introduction to the Geographic Information System of the Canada Land Inventory*. Ottawa: Department of Forestry and Rural Development.

Turner, A.K. (1992) *Three-Dimensional Modeling with Geoscientific Information Systems*. Dordrecht: Kluwer.

Waugh, T.C., and R.G. Healey, editors (1994) *Advances in GIS Research: Proceedings of the Sixth International Symposium on Spatial Data Handling*. London: Taylor and Francis.

Michael F. Goodchild
Director, National Center for Geographic Information and Analysis
Professor, Department of Geography
University of California
Santa Barbara, CA 93106-4060, USA
+1 805 893 8049
FAX +1 805 893 7095
good@ncgia.ucsb.edu