FINDING THE MAINSTREAM

Michael F. Goodchild

Abstract

GIS began as a highly specialized application of information technology, with its own hardware devices for input and output, its own data structures, and its own algorithms for data processing. Through time more and more aspects of GIS have become mainstream, and more and more standard approaches have been adopted to replace earlier specialized ones, taking advantage of the economies of scale inherent in the mainstream. However there are many reasons for treating geographic information as special, and for educating specialists in GIS concepts, principles, and use. The paper enumerates many of these, and presents the arguments against wholesale adoption of mainstream practices. Special attention is paid to metadata standards and the process of search over distributed archives for GIS data sets. The future health of the GIS industry depends on knowing when to generalize and when to specialize.

Introduction

In the movie *Sleepless in Seattle* the two leads, Meg Ryan and Tom Hanks, are linked through a common problem—an inability to sleep at night, and a consequent addiction to all-night phone-in radio. In classic Hollywood style we know the end of the movie—they will meet and live together happily ever after, even though they currently live on opposite sides of the US—but the entertainment lies in the numerous false starts and temporary disappointments of the mating dance. GIS too has its mating dance with the information technology mainstream, as the two players circle around each other, cautious about surrendering autonomy but conscious of the certainty of the conclusion. We want the benefits that being part of the mainstream would bring, but celebrate the special nature of GIS, and are not at all sure that we want to lose our uniqueness. In this paper I explore both sides of this debate: the arguments for joining the mainstream, together with the impediments that remain in the way; and the arguments for a separate identity for GIS, together with the ways in which that separate identity can be preserved and possibly strengthened. For the sake of simplicity I refer to the first set of arguments as those of the *lumpers*, and the second as those of the *splitters*, and the paper is structured as a debate in which both sides first present their cases, followed by the building of a consensus. I use the term *GIS* to refer to the entire geospatial complex of data, systems, services, and community.

Lumpers

The problems weren't so special after all

Some of the earliest roots of GIS are found in the Canada Geographic Information System (CGIS; Foresman, 1998), a massive investment by the Government of Canada in the mid 1960s. CGIS was designed to solve a very specific problem, the compilation of summary statistics from tens of thousands of map sheets. A large part of Canada's land mass had been mapped at a very detailed scale and in the form of several layers, including the capability of the land resource for various kinds of activities, and the current uses of the land. The statistical analysis involved two tasks, the overlay of layers and the measurement of area, that are notoriously difficult, expensive, tedious, and inaccurate when done by hand. Computerization offered the potential for accurate and fully automated analysis, and even though the costs were spectacular they were still less than those of a traditional manual analysis. The CGIS project resulted in the solution of many important problems in geospatial digitization, data modeling, indexing structures, and algorithm design.

I use the example of CGIS because it illustrates the very special nature of early GIS, relative to what was then the computing mainstream. In the mid 1960s almost no-one had thought of using a computer to process the information found on maps; it was far from clear how one could get the information into a computer, and what purpose would be served. Although the project used a standard mainframe and storage devices, the contractors developed a map scanner uniquely designed for large-format maps, with few applications outside GIS. Similarly, the CGIS topological data structure became the basis for many others, including ESRI's coverage model, but has no obvious analogs outside the geospatial field. Through time, however, the uniqueness of these aspects of GIS has become less and less significant. Few GIS projects today use map scanners, because so much geospatial information is already in digital form, and most paper maps are themselves products of digital databases.

The history of the topological data structure provides a typical example of how GIS has become less special. It was originally devised to solve two related problems associated with the digital representation of a certain type of map commonly found in environmental and resource management, and it is not surprising that these turn out to be the most successful areas for the nascent GIS industry of that period. All of the layers of CGIS looked similar: they divided the space into irregular areas using thin lines, and gave each area a uniform class, drawn from a small set of defined classes. This type has been called the *area-class* map (Burrough and Frank, 1996), and it is used to characterize soils, land uses, land cover, and vegetation, among other applications. When such maps need to be digitized, the obvious approach is to regard each area as a separate unit, and to create a polygon by digitizing a sequence of points around its boundary. But when applied to an area-class map, this approach will result in the *double-digitizing* of each internal boundary, adding to what is already a tedious and time-consuming task. Moreover, the two versions of each internal boundary will differ, resulting in a weaving that looks unsightly. The topological data structure of CGIS solved this problem by treating each common boundary between two adjacent areas as the record, rather than each area, and treating areas as collections of such boundaries or *arcs*. With each arc were associated pointers to the areas on each side, and to the junctions or *nodes* at each end.

All of this will be familiar to many readers, because it has been one of the cornerstones of GIS teaching for the past two decades (*e.g.*, Bernhardsen, 2002). In the early 1980s ESRI was able to exploit the power of the then-new relational database management systems by storing the pointers of the structure as keys linking tables, though the geometry of each arc still had to be stored in a special proprietary data structure. But by the mid 1990s the basic premises of the structure had changed. First, better digitizing software and procedures, and a general decline in the need for digitizing, removed the original motivations. Second, GIS applications extended far beyond environmental and resource management, to utility systems, military and intelligence activities, and urban planning. In these areas the new data types certainly did not fit the area-class map model, although there were many attempts to fit the new square peg into the old round hole. Third, computer database design was strongly impacted by object-oriented models, which had been devised to address many of the deficiencies in the relational model. GIS vendors quickly recognized the advantages of the new models, and developed systems that could store areas as objects, with geometry integrated as another attribute. The higher computational costs of doing so, including the costs of computing topological relationships on the fly, were more than offset by improvements in computational speed, thanks to Moore’s Law (Longley *et al.*, 2001L).

In summary, GIS technology is now closer to the information technology mainstream, because many of the reasons for developing specialized technology are less valid, and the costs of specialization are no longer worth incurring. The previous discussion used the example of topological data structures, but similar arguments could be made about indexing, or any of the other special innovations of the CGIS era. But this trend has its own costs, because it constantly underestimates the uniqueness of what GIS is trying to do. Again the topological data structure example will serve well as an illustration. The areas of an area-class map, or the counties of a state, are to a degree pieces of a jigsaw that behave as independent, manipulable objects. But the topological data structure implemented several important properties that are not true of pieces of a jigsaw. Areas in an area-class map cannot overlap, by definition, and there cannot be gaps between them. When an object-oriented model is used to implement an area-class map these rules must be added explicitly to the basic model, because they are not implicit in it. Topology, which was built into the basic design of the early coverage model, is now a property that must be explicitly declared by the user of object-oriented GIS. Object orientation is similarly problematic for street and river networks, which must somehow be chopped into persistent pieces. In short, object orientation is ideal for a view of the world as an empty tabletop littered with well-defined and discrete objects, but it is inherently problematic for the alternative view of the geographic world as composed of fundamentally continuous fields (Longley *et al.*, 2001).

Location is just another attribute

In early relational database management systems the tables used to describe the classes of objects in the system could be populated only with simple attributes: numbers, or text. As noted earlier, this was problematic for GIS, because the description of the geometry of lines and areas required a variable number of coordinate pairs and more complex structures. But by the 1990s newer database designs had opened much more interesting possibilities. Today, it is possible to store a wide range of information types in the cells of a database table, including images, hyperlinks, and objects that have their own complex structures. In an object-oriented GIS the geometry of a feature is stored in a column just as any other attribute, an option that was always available for point features but not for lines, areas, or more complex geometries. The rules of access to these geometry attributes are somewhat different from other attributes; and the user's view of the table may have little to do with the actual storage structure used by the system.

This technological advance serves to endorse the view that the geometry of an object is simply another attribute, to be handled and processed by the system in ways that are little different from other more conventional attributes, such as the feature's name. A utility company, for example, might enter the GIS world simply by adding a new attribute to all of its installed equipment; and location-based services (LBS) might be offered by cellphone providers based on adding a new attribute to each record in the subscriber database.

But while this view is compatible with many GIS applications, and has done much to support various new ones such as LBS, it is problematic for many other applications. The location of a feature is often of little interest in itself, but important only as a way of determining the distance or direction between features, or such topological relationships as adjacency or connectivity. In scientific applications of GIS the locations of features, so carefully recorded in the database, are of almost no value to the scientist's understanding of the processes that shape the landscape—location alone can almost never *explain* anything, though it can sometimes act as a surrogate for a truly explanatory factor. Instead, locations allow us to determine proximity, to link layers together, and to create maps. For many scientific purposes the locations of features are much less important than the relationships between them, captured in such devices as the *spatial weights* or **W** matrix (Fotheringham *et al.*, 2000) in which each element measures the potential interaction between features, based on their adjacency or some other property derived from feature locations. A scientist using a **W** matrix will often happily throw the underlying coordinates away.

It's all about economies of scale

CGIS was a stand-alone system, developed for a specific application, and never intended to solve any other problem, or to be widely distributed. Today's GIS vendors, on the other hand, rely on the massive economies of scale that exist in computer applications. By developing a platform that provides basic housekeeping, editing, and other management facilities for a given type of data, it is possible to create highly integrated and very powerful systems to handle a multitude of tasks at very little additional cost. Thus GIS is to geographic data as Excel is to simple tables, or

Word is to text—an integrated system for performing virtually any kind of operation on a specific type of information.

Economies of scale underlie all of today's information technologies. The Internet is successful because its packet technology can be used for any type of information irrespective of its meaning, and a single packet might contain codes for text, numbers, music, or images. Digital technology is successful because its binary alphabet can be adapted, using simple coding systems, to represent virtually anything. As a result the unit costs of digital technology are low, and falling lower.

Thus it benefits GIS vendors to be able to adopt technologies developed for generic applications; to fit GIS applications into more generic models; and to favor those that can be so fitted over those that cannot. GIS has been influenced in this way by a long list of technologies, from the original programming languages to relational and object-oriented database management systems, CASE tools for database design, reusable software component technologies such as COM, and many more. GIS is being drawn into the information technology mainstream by an inexorable economic process, which ensures that it is cheaper to do GIS this way, to give less emphasis to its more specialized needs, and to attempt to fit GIS applications into the generic forms adopted by the mainstream.

The splitters

Spatial is special

To the splitters, the needs of GIS are so unique as to require special programming, special hardware, special courses to introduce the fundamental concepts, special training—in fact, everything that we identify with the GIS professional, or the Spatially Aware Professional (SAP; Longley *et al.*, 2001). Only the SAP knows about map projections and datums, knows the peculiar language of GIS and the specialized meanings that it gives to such terms as buffer or topology, and knows how to decode the numerous three-letter acronyms starting with D: DOQ, DEM, DLG, DRG, *etc*.

Several people have written about what exactly makes spatial special. Anselin (1989) identifies two characteristics that are almost universally true of geographic data: spatial dependence and spatial heterogeneity. Spatial dependence refers to the tendency for geographic data to exhibit spatial autocorrelation, the subject of Tobler's First Law of Geography (Longley *et al.*, 2001): "all things are related but nearby things are more related than distant things." Spatial dependence is the basis of all GIS data models, since it allows the kinds of simplifications that are essential to the description of what is in reality an infinitely complex world. Spatial dependence also creates a headache for any statistical analysis in GIS, because it invalidates one of the most commonly made assumptions of hypothesis testing, that samples were drawn independently from a population. Spatial heterogeneity refers to the tendency for conditions to vary from one geographic location to another, such that the results of an analysis in one area tend to differ from the results of the same analysis in another area. This behavior, which statisticians would call non-stationarity, plagues any attempt to generalize about processes on the geographic landscape. Taken together, the two characteristics ensure that it is never easy to use standard statistical methods in GIS, and leads to the need for specialized courses in spatial statistics and geostatistics.

Of course many other things are special about spatial data besides these rather abstract and esoteric properties. Among them are:

- *Large volume*. While 1 MB of text would make a medium-sized novel, a typical remotely sensed image occupies hundreds of MB, and the EOS series of satellites generates in excess of 1 TB of data per day.

- *Uncertainty*. It is impossible to measure position on the Earth's surface exactly, and attributes of geographic features are often uncertain, because of limited measurement accuracy, vague definitions of classes, errors introduced during processing, and for many other reasons (Zhang and Goodchild, 2002).
- *Applications*. Geographic information has a vast number of uses, in virtually all areas of human activity.
- *Production arrangements*. In many countries, it has been traditional for geographic information to be produced by the national government, through a national mapping agency. This tradition is of course currently evolving rapidly into a much more complex series of arrangements, since it is now possible for almost anyone to produce geographic information.
- *Impacts on society*. GIS is unique in its ability to threaten individual privacy, through the creation of massive, high-resolution databases and their linkage through street address and other geographic keys, and in its importance to surveillance (Curry, 1998).

Educating the SAP

If spatial is indeed special, then it follows that SAPs require special education, in the form of courses, training programs, textbooks, societies, conferences, and all the other apparatus of a discipline. No one could play a grand piano without extensive training and practice; by analogy, no one can do GIS without an understanding of its basic principles, a knowledge of its terms, and a familiarity with its systems. In some jurisdictions, such as the State of South Carolina, laws now require professional qualifications for anyone providing GIS services.

The question of whether GIS is a discipline, or the basis for a discipline, has been debated at length (Wright *et al.*, 1997). The consensus seems to be that there are fundamental issues associated with GIS, and underlying its development and application, that do indeed constitute a branch of science, and specifically a branch of information science, and that GIS is better seen in this light than as an application of generic methods of computing. The shift from GIS to GIScience has turned out to be a fairly extensive and successful trend, with several journals, numerous courses and degree programs, and growing conferences. GIScience has roots in a number of disciplines that address issues related to mapping, including cartography, photogrammetry, surveying, and remote sensing; interest in GIS has reinvigorated many of them, and motivated new forms of collaboration. At the same time other disciplines have discovered their relevance to the GIS enterprise, including statistics, cognitive science, and of course computer and information science.

But there are distinct differences between the history of GIScience, and the history of comparable disciplines. The theory and principles of statistics were established long before the invention of computers, hand calculators, and even slide rules, and today some instructors still insist on students demonstrating their abilities to perform analysis by hand before being allowed to use computers and statistical packages. As a result packages could be developed to use a common language and commonly accepted conceptual and theoretical framework. But GIS has in a sense reversed this process—it is only after the development of large packages that the community is beginning to realize the need for a common language and common principles, and for courses that instruct students in the conceptual frameworks of the field. The efforts of the Open GIS Consortium (http://www.ucgis.org) are in a sense retrofitting standards onto a field that would have benefitted enormously had they existed earlier.

Metadata

The question of whether spatial is special comes to a head in the approach taken by the GIS community to the design of suitable metadata standards. Metadata (often described as data about data, or data that make data useful) are essential to the effective sharing of data sets, because they provide the catalog records that can support searching, the instructions needed to open and use data sets, and information essential to an informed assessment of fitness for use. In the early 1990s the U.S. Federal Geographic Data Committee sponsored the development of a metadata standard for geographic information, known as the Content Standard for Digital Geospatial Metadata or less formally as the FGDC Standard (http://www.fgdc.gov). It includes descriptions of the basic characteristics of the data set, including its spatial resolution and the geographic extent of its coverage, information on how and when it was created, information on its format, and descriptions of its quality. Since its inception the standard has been widely adopted, and has become the basis of a proposed international standard.

The FGDC standard is elaborate, and it is quite common for a full description to exceed the size of the data set itself. It has been described as a *producer's* approach to description, because its content is dominated by things the producer of the data would know, rather than things the user of the data might need to know. It is also unique to geographic information, and it would make no sense to attempt to use it for data sets that are not geographic.

By contrast, the library community at about the same time developed what is now known as the Dublin Core, a minimal set of metadata descriptors that could be applied to any information. They include the standard library search keys—author, title, and subject—together with other information that a library would record about a book, such as its publisher. The 15 elements of the Dublin Core contrast sharply with the several hundred elements that can be included in the FGDC standard.

The contrast between the two standards suggests an obvious question: does GIS need its own standard, or would a more generic standard suffice? A more generic standard would allow catalogs of GIS data sets to be readily interoperable with catalogs of other types of data sets. The limited scope of the Dublin Core standard makes it much easier and cheaper to implement; by contrast, the FGDC has found it necessary to encourage the use of its standard by offering funding to agencies willing to adopt it. On the other hand the original version of the Dublin Core does not contain the two most obvious characteristics of any geographic data set, its spatial resolution and spatial extent, although it is easy to extend the standard to include them. Should the GIS community move into the metadata mainstream, or continue to invest in a highly specialized approach to metadata?

Search engines

A related issue concerns the engines such as Google and Altavista that provide essential search services on the WWW. These engines crawl the vast resources of the WWW, examining pages of text and extracting keywords that appear to characterize content. The catalogs that result are used by millions of users every day to find information resources appropriate to their needs, by matching keywords. As every SAP knows, search engines are not an effective way of finding GIS data sets, because in most cases they use special formats that are not recognized by search engines looking for text (although the growth of text-like representations, such as GML, may help in this respect). MapFusion, a product of Global Geomatics Inc., is one example of a specialized engine capable of detecting a large number of GIS data set formats, though the domain of its search must be defined explicitly. Google Image is another example, capable of detecting a range of standard image formats and building a catalog from the contents of pages containing the images. But geographic information is unlikely to be found embedded in WWW pages.

In short, the task of searching for geographic information on the WWW is special. So too is the task of searching for any other kind of data set that makes use of highly specialized formats, as GIS data sets must. But what makes the GIS community unique in this respect is its belief in the widespread sharing of data sets, based on the observation that GIS data sets are often useful for a wide range of purposes, and the tradition of centralized production of GIS data sets as public goods.

Towards consensus

This discussion leads me to a series of conclusions about the nature of the mating dance between GIS and the information technology mainstream—the equivalent of a prenuptial contract between Tom Hanks and Meg Ryan, aimed at identifying areas of agreement, while reserving certain areas of distinct identity.

1. Some applications of GIS are more compatible with the mainstream than others. They include applications that deal with features that can be conceptualized as discrete objects, and where the GIS is used to maintain and manage an inventory of objects distributed over the landscape. Utility applications are ideal in this respect, because a GIS can be used to add geographic location to the existing attributes of such features as transformers, poles, manholes, and buildings. LBS applications are similarly well-suited, because they rely on adding location to the existing attributes of customers. Many of these applications are extremely widespread, and potentially lucrative for the GIS industry. Object-oriented data models are very powerful frameworks for addressing these applications.
2. Other applications are much less compatible with the mainstream. They include the use of GIS in scientific research, because of the dominance of the field view that emphasizes continuously varying attributes of location, rather than discrete and persistent objects. They also include network applications, because a network does not naturally partition into discrete objects. Object-oriented data models are problematic for these applications.
3. Economies of scale will continue to pull GIS into the mainstream. SAPs interested in applications that are less compatible with the mainstream, notably scientific applications, will have to pull hard in the opposite direction, arguing for example for data models that are more appropriate to geographic phenomena conceptualized as fields.
4. GIS data sets require highly specialized approaches to search and retrieval. A concerted effort to develop a new generation of search engines would have great potential benefits for the field.
5. The special characteristics of GIS will continue to foster an emerging GIScience, and courses aimed at educating new generations of SAPs. There will always be areas of GIS that require highly trained users, though other applications such as geocoding, mapmaking, or navigation are already accessible to the general public.
6. Because of the wide spectrum of applications and skills associated with GIS, it will always be difficult to bound and regulate the field; attempts to put professional GIS use on a similar footing to surveying make little sense in this very diverse environment.

In summary, the mating dance will continue for at least the foreseeable future. There are major benefits to joining the mainstream, but there are also major reasons why GIS is special, and why the specialized knowledge, concepts, and skills of the SAP are essential to its continued health and success.

References

Anselin, L., 1989. *What Is Special about Spatial Data? Alternative Perspectives on Spatial Data Analysis*. Technical Report 89-4. Santa Barbara, CA: National Center for Geographic Information and Analysis.

Bernhardsen, T., 2002. *Geographic Information Systems: An Introduction*. 3rd Edition. New York: Wiley.

Burrough, P.A. and A.U. Frank, editors, 1996. *Geographic Objects with Indeterminate Boundaries*. London: Taylor and Francis.

Curry, M.R., 1998. *Digital Places: Living with Geographic Information Technologies*. London: Routledge.

Foresman, T.W., editor, 1998. *The History of Geographic Information Systems: Perspectives from the Pioneers*. Upper Saddle River, New Jersey: Prentice Hall PTR.

Fotheringham, A.S., C. Brunsdon, and M. Charlton, 2000. *Quantitative Geography: Perspectives on Spatial Data Analysis*. Thousand Oaks, CA: Sage.

Longley, P.A., M.F. Goodchild, D.J. Maguire, and D.W. Rhind, 2001. *Geographic Information Systems and Science*. New York: Wiley.

Wright, D.J., M.F. Goodchild, and J.D. Proctor, 1997. Demystifying the persistent ambiguity of GIS as 'tool' versus 'science'. *Annals of the Association of American Geographers* 87(2): 346–362.

Zhang, J.X. and M.F. Goodchild, 2002. *Uncertainty in Geographic Information*. New York: Taylor and Francis.