outreach

51

Towards a science of geographic information

Michael Goodchild
University of California, Santa Barbara

Introduction

It has often been said that GIS is technology-driven; that the interest we see expressed in conferences, software sales, textbooks and magazines is motivated not so much by the ability of GIS to solve real problems, as by a fascination with a field that combines the attraction of maps and geography with novel high technology. In this somewhat cynical view GIS is a technology in search of applications, and the ability to process geographic information and create colourful, impressive products may explain some of its current popularity.

At another level GIS is clearly application-driven, with an industry committed to defining and exploiting specific market niches. Many of the current range of GIS software products were developed for early applications in resource management, and later extended to meet the needs of utilities or municipal governments. Many of these applications emphasize the management of facilities in a spatial context, and rightly see GIS as a problem of providing geographical access to vast numbers of records on customers and facilities.

Neither of these views addresses the relationship between GIS and universities, or the nature of GIS research, or the role of GIS as a scientific tool. Yet the use of GIS raises a host of issues regarding the computerized handling and analysis of geographical data, and at the same time offers tremendous analytical power. In this article I explore these two roles of GIS in science—first as an enabling technology for a range of scientific disciplines, and second as a focus for research on a set of generic issues that together might constitute a science of geographic information.

GIS as a scientific tool

A useful starting point for a discussion of the value of GIS as a tool for science is to compare it to statistical packages, e.g. SAS, SPSS, S, or Systat. All were developed to satisfy a market in the scientific research community for integrated software to perform a variety of forms of statistical analysis. Statistics as a discipline has developed a vast array of techniques, and integration is desirable to avoid duplication of effort, in keeping track of data formats, variable names, etc., and

to allow procedures to be linked together easily. Integration at this scale (many of these packages contain literally thousands of procedures) is possible because the techniques share a very simple, flat-file data model.

Initially packages like BMD and SAS were developed and marketed entirely within the scientific community, as general purpose scientific tools. Although the customer base has broadened considerably, there is no similarity to the dependence of the GIS industry on municipal, utility and government markets. The statistical packages are primarily tools for science, and very effective methods of delivering the theory and methodology developed in the parent discipline of statistics. From a marketing perspective, GIS by contrast is a tool for information management, with no obvious parent discipline and with only a minor interest in supporting scientific analysis and modelling. Only a small fraction of GIS installations currently support scientific research, and only a fraction of GIS vendors market actively to the research community.

Against this we must balance the fact that geographical relationships and the geographical perspective are enormously powerful devices in research. A map showing incidence of disease is invaluable in suggesting possible causal factors or processes of contagion. Geographical proximity and adjacency are powerful factors in processes as distinct as consumer shopping behaviour and migration. Maps are valuable ways of visualizing the results of modelling, for example in atmospheric science. Geographical space has a clear role in explanation not only in the earth sciences but also in social science research, in understanding distributions of artifacts or prehistoric settlement patterns in archaeology, or in regional economic modelling or sociology. Ultimately, GIS has a role as a display, analysis and modelling tool in all those disciplines that deal with geographically located information and processes on the earth's surface. Perhaps it may even be that the abundance and importance of geographic information is actually greater than the tabular, numeric information accommodated by the statistical packages—certainly the flat tabular file is a subset of the relational model so popular in current GIS.

What will it take to exploit the potential of GIS as a tool for science, rather than a system for information management? On the hardware side, many GIS vendors have already moved to the scientific workstation as the platform of choice, and UNIX offers enormous power for integrating GIS with more advanced modelling and analysis, and for networking. Workstation capabilities are expanding dramatically, and we can expect to see spectacular improvements in the next few years, particularly in 3D graphics.

However, the story on the software side is more complex, and the scientific market is not currently viewed as particularly significant by many vendors. The market is not large, and academics have an annoying habit of expecting major price reductions, if not outright gifts. GIS courses and programmes are proliferating, but many are oriented to training students for jobs in the GIS industry, rather than to educating future scientists in the use of an important tool. In this sense again there is little comparison with offerings of statistics courses.

To develop the role of GIS as a technology for science we must ensure that vendors see support of analysis and modelling as important. We need institutions, journals and texts that draw out this role for GIS, and courses that provide introductions to GIS and geographical analysis for scientists in the relevant disciplines. One of the critical factors in the early success of SAS was the ease with which specialized methods of analysis could be contributed by scientists and incorporated within the general framework, even though their applications might be limited to a few projects. We have yet to see the same sort of arrangement in GIS, although there is a similarly enormous range of forms of spatial analysis that might be contributed, and GIS could provide a very effective method of delivery.

A science of spatial information

The current range of GIS software and hardware products incorporates an impressive array of technological breakthroughs. Concepts such as the TIN and quadtree are the direct result of GIS research, and are only two among the many innovative ideas to have emerged over the past three decades. Any technologically based field must be constantly supplied with new ideas if it is to thrive, and needs to be supported by an active research and development community.

However, there is a strong feeling at the present time in the GIS community that the most important issues confronting the field are not necessarily technological. The GIS community seems to be converging not around a single, uniform software product (a standard GIS) or a single application, or around the technology itself, but around a set of generic issues that emerge from the technology. Whatever the application or EDP solution, every user of GIS faces the same set of problems in dealing effectively with digital geographical data, and these problems in turn form the agenda for discussion at GIS meetings—the true glue of the GIS community. Some of the more prominent are:

- data capture—how to convert data from raw to digital form in an efficient, cost-effective manner;
- data modelling—how to represent the infinite complexity of the real world in a discrete, digital machine—whether to use raster or vector, layers or objects, how to model complex objects;
- accuracy—how to cope with the uncertainty present to varying degrees in all geographical data;
- volume—how to deal with the fact that demands for geographical data will often exceed the space available for storage;
- access—how to design data structures, indexes and algorithms to provide rapid access to large volumes of geographical data;
- analysis—how to link GIS databases with advanced modelling capabilities;
- user interfaces—how to present the GIS database to the user in a friendly, comprehensible, readily used fashion;
- costs and benefits—how to measure the benefits of GIS information and compare them to the costs;
- impact on organizations—how to introduce GIS successfully into a complex organization.

All of these issues transcend the technology itself, and all of them in one way or another affect the technology's usefulness, whatever the application and whatever the platform. In recent years they have emerged in various guises as the basis of the research agendas of the NCGIA (NCGIA, 1989), URISA (Craig, 1989) and the Regional Research Laboratories (Masser, 1990; Maguire, 1990). Goodchild (1990) has argued that together they constitute a science of geographic information, and that the future of the GIS community lies in recognizing a common interest in geographic information science rather than in the technology of geographic information systems.

Once one begins to see the generic issues that underlie GIS, and transcend the particulars of its technology and its applications, then one can begin to understand how GIS can affect one's view of the world. Traditionally, information about places on the earth's surface has been stored and transmitted in the form of maps, images, text and to some degree sound. The focus of early GIS was on the digital database as a store of maps—maps were the input, the output and the metaphor of GIS applications. But increasingly GIS is seen as a means of access not to maps but to the real world that those maps represent. The purpose of the database must be to inform the user accurately about the contents of the real world, more than about the contents of a source document. A DEM (digital elevation model), for example, should be assessed on its ability to return the elevation of any point on the earth's surface, not the position of an abstract contour line.

GIS has also affected the role of geographic information within an organization. It encourages the notion that geographic information is a commodity that flows through the organization, and that has a value determined by its accuracy, currency, accessibility, etc. In fact it may be the central commodity in some organizations, for example forest resource management agencies. Geographical data needs careful planning and budgeting if it is to be collected and updated on a regular basis, and made accessible to the organization's analysts and decision-makers. Finally, if information is important, then it is rational to use different types of information as the basis for the organization of departments and systems.

In summary, GIS is a diverse collection of interests, software and hardware solutions, and applications. Two software products applied to the same problem (e.g. ESRI's ARC/INFO and IBM's GFIS applied to management of a utility company's facilities) can produce entirely different solutions; and the needs of forest resource management and school bus routing appear to have very little in common. But there is a growing sense that the issues that hold the GIS community together, and produce convergence rather than divergence, are the generic issues of dealing with geographic information—with representing it in a digital computer, and working effectively with it to produce answers to problems.

GIS as a discipline

The current growth of GIS shows no signs of abating and is likely to continue for some time into the 1990s. New magazines are appearing, and existing ones, such as *GIS World* and *Mapping Awareness*, are growing and increasing their circulation. Conferences are numerous and successful, offering workshops on increasingly specialized topics and access to the latest vendor products. New software vendors are entering the market with exciting and innovative products. GIS is finding new applications and strengthening its penetration into existing markets. GIS courses are proliferating at universities and colleges, and are finding increasing interest from students anxious to acquire useful skills.

On the other hand there are increasing signs of diversification, and this trend is likely to continue to strengthen in the next few years. GIS applications such as facilities management fall under the spatial information paradigm, whereas scientific and resource analysis applications fall under the spatial analysis paradigm. The former emphasizes the database and query aspects of GIS, whereas the latter tends to focus on modelling. The split is illustrated by the case of two Canadian companies—TYDAC and GeoVision—one marketing 'spatial analysis systems' with the very successful SPANS product, the other marketing "geographic information systems'. Within the PC marketplace, there is increasing divergence between products aimed at GIS applications in resource management, facilities management, or market research (compare, for example, PAMAP, TYDAC's SPANS, Facility Mapping Systems' FMS/AC and Strategic Mapping's ATLAS*GIS).

This trend to diversification is appropriate and rational, as it matches software and platforms with different functions and applications. The complex modelling and analysis of resource management require a very different solution from intensive digitizing or the management of large facility inventories. In time, we can expect this trend to lead to more and more specialization within the GIS industry, as it becomes less and less possible to offer a single software solution for all platforms and all applications. One vendor may specialize in digitizing stations using PCs, another in database maintenance using large mainframes and terminals, another in spatial analysis using advanced personal workstations, and another in 3D applications.

Despite this sense of growing diversity in the GIS community, there is evidence of convergence. The past few years have seen the emergence of several series of conferences aimed at the full GIS community. In the US, the annual GIS/LIS series sponsored by a consortium of five societies (AAG, ASPRS, ACSM, AM/FM and URISA) has grown quickly to over 3,000 attendees. In Canada, the Ottawa meetings in early March have been similarly successful. The lone textbook of 1986 by Peter Burrough (*Principles of GIS for Land Resources Assessment*, Oxford) has now been joined by several others (e.g. Aronoff's *GIS: A Management Perspective*, WDL, Ottawa; Star and Estes' *Introduction to GIS*, Prentice-Hall), and many more are on the way. New organizations have appeared (the Association for Geographic Information seems to be a particularly successful example), and GIS now has its own journals.

All of these would be recognized in the sociology of science as symbols of an emerging scientific community—in short, a discipline. But unlike physics or biology, GIS has no fundamental problems to solve of the magnitude of the origins of the universe or the basis of life. One view holds that it is merely a tool, and that the GIS research community must wither away as the tool reaches perfection. Another, presented at some length above, holds that there are fundamental issues in GIS, not so much in the tools themselves as in the use of the tools. Or perhaps GIS is like statistics—a tool to most scientists, but a set of fundamental research problems to its parent discipline.

If GIS is a discipline, then it is clearly not 'owned' by any traditional one. Geography, cartography, surveying, photogrammetry and engineering have all been accused from time to time of trying to dominate GIS, but with little success, as GIS is fundamentally an interdisciplinary field. Whether it develops the institutional structures of a discipline in its own right, like statistics, or continues as an interdisciplinary consortium of interests like remote sensing remains to be seen.

Prospects for the future

There seem to be two contrasting views of the prospects for GIS in the coming decade. The first is negative and the second positive, and it seems more likely that the second will prevail. However there are actions that can be taken to strengthen the odds.

In the negative view, GIS will fragment and disappear, and by the end of the decade will be nothing but a memory. Geographers often draw a parallel between GIS and the introduction of quantitative methods to geography in the late 1960s (Taylor, 1990), and comment on the lack of interest in quantification, at least in human geography, in the 1980s. GIS will fragment because it is too loose to hold together, and because the glue is too weak and abstract. Users of IBM's GFIS, Intergraph's TIGRIS and Map/Info will cease to see any reason to attend the same conferences. The consortium of five organizations responsible for the North American GIS/LIS conference series will break up and each will concentrate on its own agenda. GIS will be seen as the Edsel of EDP, too awkward, complex and expensive except in some specialised applications.

In the positive view, the GIS consortium will continue to converge. A constant supply of better tools seems assured, particularly in computing speed, software integration, network communication, graphics and storage capacity. The infrastructure of the GIS community will continue to improve, with better magazines, organizations, textbooks, meetings, and all of the symbols of an emerging specialty. Less assured but essential is a constant supply of new players in the industry, since the pattern has been that they are the source of a disproportionate share of technological innovation. New players such as Prime/Wild with System/9, Small World Systems, or Caliper's GIS-PLUS bring new ideas to the industry.

In the positive view, the public agencies will promote and develop standards for data exchange formats, structures, models, and data quality. Training and education programmes will develop through cooperation between vendors and institutions, and lead to the emergence of a strong set of core concepts. Funds will be available through cooperative agreements to support the development of teaching facilities, and to ensure that these keep pace with developments in the technology.

The results of research currently under way will emerge in improved products. Of particular significance will be:

- data models to handle 3D and time dependence, and complex interactions between objects;
- support for complex analytical applications, including tracking of data lineage, tools for visual interaction with the stages in the analysis process, and propagation of uncertainty;
- support for quality assurance and quality control (QA/QC) especially in GIS applications where litigation is a constant problem;
- support for multiple media—unstructured images, both digital and NTSC, text and sound;
- integration of GIS with the capabilities of GPS for data collection and compilation;
- tools for visualizing 3D and time-dependent data;
- tools for data compilation, particularly in 3D;
- improved techniques for conducting functional requirements studies, evaluating costs and benefits, benchmarking and other aspects of the GIS acquisition and project management process.

Finally, the GIS community will converge around a common concern not only for the technology of GIS, but more importantly for the common issues that transcend the technology and pervade all of its applications. GIS can survive by constantly developing new and exciting capabilities, or by constantly finding new applications, but the really fundamental issues in GIS are those that are common to all users of geographic information—how to capture a complex and dynamic world in a digital database and provide access to it in a useful, accurate and cost-effective manner.

References

Craig, W.J. (1989). URISA's research agenda and the NCGIA. *Journal of the Urban and Regional Information Systems Association*, 1(1): 7–16.

Goodchild, M.F. (1990). Keynote address: spatial information science. *Proceedings*, Fourth International Symposium on Spatial Data Handling, Zurich, *1*: 3–14.

Maguire, D.J. (1990). A research plan for GIS in the 1990s. In Foster, M. and Shand, P.J. (Eds) *AGI Yearbook* 1990, London: Taylor & Francis, 267–277.

Masser, I. (1990). The Regional Research Laboratory initiative: an update. In Foster, M. and Shand, P.J. (Eds) *AGI Yearbook* 1990, London: Taylor & Francis, 259–263.

NCGIA (1989). The research plan of the National Center for Geographic Information and Analysis. *International Journal of Geographical Information Systems*, 3(2): 117–136.

Taylor, P.J. (1990). Editorial comment: GKS. *Political Geography Quarterly*, 9(3): 211–212.

Based on the First Midlands GIS Lecture delivered at the University of Leicester, October 6th 1989, and sponsored by The Association for Geographic Information, The Midlands Regional Research Laboratory, and The Institute for Engineering, Surveying and Space Geodesy.

52

An object-oriented future

Hilary Hearnshaw
ESRC Midlands RRL, Leicester

The object-oriented approach

The last twelve months or so have seen a growth in interest in t the object-oriented approach in the world of geographical inform interest reflects similar, developing, interest in the principles of in computing science in general. "Object-oriented" is a descripti been applied to a range of activities, including:

- *Object-oriented data modelling* (OODM). This is the abstr world to a form appropriate for use by a digital system, properties of an object and its operators to be associated sulated), and for specified passing of properties between obj This allows a closer similarity in the model to the complexit than previous techniques for data modelling, such as e modelling, are able to do. They are not able to include the easily as object-oriented data models can.
- *Object-oriented programming languages* (OOPLA). La Smalltalk, C + +, Actor, are high level languages which allow among other things, efficiently to represent objects which not own internal data model, but also contain their own operatio
- *Object-oriented database management systems* (OODBMS). developed in response to the need to handle highly comple Previous relational, or hierarchical, databases were failing to d are more suitable than previous DBMS in such fields as multi complex engineering databases or geographical informatio Attempts at implementing successful object-oriented data mo also motivated the development of appropriate database man

As the development, and evolution, of object-oriented techniq there is no question that they will, so the application to geographical cessing will grow. It is, of course, realised that neither object-orie ing, nor object-oriented database management systems, will solve of, the problems in geographical information science. However, obje ques seem to provide more appropriate solutions than previous at quirement for handling, and using, large amounts of spatially refer