AURISA 93
The Annual Conference of the Australasian Urban and Regional Information Systems Association Inc
Adelaide, South Australia
23-26 November 1993

FUTURE DIRECTIONS: GEOGRAPHIC INFORMATION AND ANALYSIS

Michael F. Goodchild

National Center for Geographic Information and Analysis, and
Department of Geography
University of California
Santa Barbara, CA 93106-4060, USA

Phone 1 805 893 8224
FAX 1 805 893 8617
Email ncgia@ncgia.ucsb.edu

ABSTRACT

Further improvements in the speed and storage capacity of computing systems seem assured. The paper examines trends that can be anticipated in GIS and related fields over the next ten years, emphasizing the roles of improved network communications, expansion of GIS applications to the consumer sector, improved decision support, and a more rational approach to data quality. Some final comments are offered on directions in education.

INTRODUCTION

Any attempt to discuss future directions of a fast-moving technology like GIS is fraught with difficulty, even over periods as short as a year or two. No one could have predicted the failures and successes that we have seen in the GIS software industry in the past eighteen months, and while we may each have our own ideas about what is likely to happen in the next eighteen, reality is sure to be just as surprising.

With that concern expressed, however, the purpose of this paper is to explore some of the trends under way that are likely to have significant impact in the next few years, particularly in the areas of geographic information and analysis. The paper covers only a selection of topics of interest to the author (for more general discussion of the future of GIS see Maguire, Goodchild, and Rhind, 1991), but the discussion ranges from networks to the design of databases, and ends with some comments on education.

Just to set the stage a little, it may be helpful to think about the changes that have occurred in GIS in the past ten years, and to realize that change is likely to be at least as rapid in the next ten. In 1983, I wrote papers by hand, gave them to a secretary to type on a typewriter, waited a week for the typescript to come back, edited and corrected it, and had it retyped. I believed strongly that this was the optimum system, and that I could never learn to compose on the screen. The PC had not yet appeared on the market, and I computed on a Control Data mainframe with 64K words of main memory, using a Tektronix terminal and a 1200 baud line, writing much of my own code in Fortran. A software package called ARC/INFO had recently appeared, and had been adopted by the Province of New Brunswick for timber management, on the advice of one Roger Tomlinson. Its capabilities for polygon overlay seemed very exciting, but the cost of software and necessary hardware was far beyond the reach of academic budgets. GIS had no textbooks (Burrough's was published in 1986), no journals (the *International Journal of GIS* began in 1987, and *GeoProcessing* was in the process of folding), and few meetings (the GIS/LIS series began in 1988). To all intents and purposes, GIS had not yet begun.

GIS is a complex application, and much of its growth and development will continue to be driven by advances in hardware. For the past 6-8 years, the speed of desktop workstations has come close to doubling every year, while the price has held roughly constant, or even dropped. I paid $6,000 for my first PC in 1984, and now would pay no more than $2,000 for more than 100 times the power. In fact, in recent years it has been possible to predict the speed of the newest workstations with some accuracy from the formula:

$$mips = 2^{(year-1984)}$$

although with some recent dampening. In 1993, that gives us a cutting edge of 512 mips, compared to 1 mips in 1984.

At the same time, the cost of RAM is dropping dramatically, and the 256Mb chip is here. In major research centers, the marginal cost of computing power is so low that installed capacity vastly exceeds demand, and total CPU utilization over a research network of fast workstations rarely exceeds a few percent. Unlike ten years ago, local availability of computing cycles is no longer a constraint. For GIS, this means continued innovation in the areas of GIS that are inherently compute-intensive, such as building topology on the fly, and easy-to-use graphic interfaces.

Developments in data storage have been equally dramatic. 10 years ago it cost me about $1,000 to store 10Mb of data for a month on a mainframe disk farm. Today, $1,000 will buy a lifetime's use of a 1Gb drive. The Sequoia/2000 project, based at the University of California, is installing a 100Tb drive to handle space imagery for Earth scientists, making the data available to them at various institutions in the State over T3 lines at rates of around 50 Mbits/sec. The concept of 100Tb of data is mindboggling, even though it represents a mere 100 days of observations from the EOS satellites that will be placed in orbit later this decade. The average academic writes perhaps 10Mb in a lifetime, and the entire Agatha Christie corpus is only about 100Mb, yet a 100Tb store has enough capacity to hold the text of all of the books ever published.

How will seemingly ever increasing speeds at no higher cost, and ever increasing storage capacity, be used to further the objectives of GIS? The remainder of the paper discusses four specific issues that seem to the author have critical importance for the future of geographic information and analysis at this time.

COMMUNICATION

The change from mainframes to personal computing was largely driven by dramatic reductions in the cost of cycles. Over the past five years, we have seen equally fundamental changes brought about by reductions in the cost of communication, as high speed LANs and WANs and associated software have revolutionized computer communications. The research network Internet (Krol, 1992) now increasingly connects the U.S. GIS community in what is becoming a network of high speed electronic highways - the National Research and Education Network (NREN). Operating currently over a T3 (50Mbit/sec) backbone, Internet is essentially free to most users, who now number over 10 million in the U.S. alone. Gateways extend Internet to most of the world at megabit rates, allowing U.S. GIS users to communicate with their colleagues in Australia as easily as if they were in the same building.

Internet has brought profound changes to the field of GIS in its impacts on personal communication and access to data. I can search library catalogs, the Congressional Record, pending legislation in Sacramento, or the complete works of Shakespeare, without leaving my workstation. With Internet servers like WAIS (Wide Area Information Server) and Gopher, the network becomes one vast data resource, free and at electronic speeds.

Joseph Banks, the 19th Century botanist who described many Australian species, is known for maintaining an extraordinary correspondence of 50 letters a week for all of his adult life, to a large, informal network of colleagues. Using the networks, an average researcher now sends similar numbers of messages by email as a matter of course, to networks of colleagues that are remarkably unconstrained by the traditional boundaries of disciplines. E-mail is fast, and by moving the message rather than the paper, definitely green.

What effect is this having on spatial data handling? Although initially and for many users primarily a means of sending text, Internet also supports graphics, sound, and video, and can be used to send and receive maps or images. NOAA (the National Oceanic and Atmospheric Administration) distributes satellite weather images over Internet. WAIS is being developed by the U.S. Geological Survey as a mode of access to digital spatial data at the Eros Data Center, and similar systems are under development by NASA (National Aeronautics and Space Administration) for its data. Reports, bibliographies, documentation, source code, and endless other GIS-related material is available over Internet in digital form.

Another interesting potential area of application for Internet lies in the area of computer supported collaborative work (CSCW; Bowers and Benford, 1991). We already have the software and hardware technologies that allow two or more researchers or decision-makers, in different cities or even different countries, to work together using a shared computer "desktop", augmented by video and audio. What effect these technologies will have on the process of decision-making, or on its costs, is an interesting area for research.

More valuable, perhaps, is the informal role that Internet plays in personal communication - if you like, the sociology of Internet. NCGIA sponsors an electronic list server, GIS-L, from the Buffalo site, that is probably read by several thousand people daily (Mark and Zubrow, 1993). On the average day some 40-50 messages appear, on topics ranging from data ("does anyone have a DEM for Bolivia?") to technical problems ("looking for a converter from Scitex to ARC/INFO format") to applications ("anyone have experience in the use of GIS to settle native land claims?"). The informality and geographic reach of Internet seems to fill a void that no other medium can, creating a global community that is fractured only by differences of language.

The economics of Internet are also interesting - it has been compared by Vice-President Gore and others to the investments made in the 1950s and 60s in the interstate highway system, and described as "the most important and lucrative marketplace of the 21st Century", with massive potential for stimulus to the economy through a combination of public and private investment. In the next ten years, Internet access will be extended to many libraries, schools, hospitals, and private homes.

The nature of spatial data - shared between many users as a common resource, relatively static, and valuable for many applications - makes it almost uniquely able to take advantage of developments in network communications. Networks will produce a very different GIS community in the next ten years - informal, global, and with geographic information, rather than hardware or software, as the primary commodity.

CONSUMER GIS

Ten years ago, the cost of GIS hardware and software limited its applications almost entirely to the public sector, notably in resource management and utilities. More recently, the falling cost of entry-level GIS, now perhaps one tenth of what it had been, has led to numerous new applications in market research, health services planning, and mapping. Spatial data technologies are now increasingly available to the general consumer. GPS (global positioning system) receivers are being sold for $500 in consumer electronics stores to hikers and boat owners. Airline magazines offer desktop mapping systems that provide map-based restaurant and hotel guides for travelers. For $99, one company offers a map database that can be searched for any hotel or restaurant in Los Angeles, one of 100 cities available from this vendor.

Consumer GIS, or spatial data technology offered to the mass market, has a potential number of installations at least two orders of magnitude greater than the current installed base. It has strong links to other, related areas of current research and development, such as Intelligent Vehicle Highway Systems (IVHS), and interactive TV. We have the technology to equip cars with in-vehicle navigation systems, using GPS and dead-reckoning to maintain location with respect to a digital road map. Soon, it will be possible for a businessman planning a visit to a distant city to download a current road map and business directory into his or her PC via the home TV.

From a research perspective, it is important to ask what investments ought we to be making now to prepare for a much broader availability and application of GIS in the future? What is the appropriate mix of public and private investment, and what role can government play in stimulating and coordinating development? In the U.S., there has been much interest recently in the concept of a National Spatial Data Infrastructure (NSDI), as the geographic component of a larger National Information Infrastructure (NII). Using the highway analogy, it is argued that government initiative and public investment should focus on providing the means, in the form of communication networks, standards, and education, while the data and information that travel the network are analogous to highway vehicles, built and owned through private investment. Complicating this is the tradition in the U.S. that geographic data are a public good, at least at the federal level. One possible compromise would leave the digital topographic frame - geodetic control network, locations of topographic features, and road network - in the public domain as a template to which the private sector can add value. But this would be a sharply different solution from that adopted in the UK (Rhind, 1993).

One of the strongest arguments for consumer GIS is its potential for increasing access to geographic information of importance to the general public. Consumer advocates such as Ralph Nader have argued that a GIS in every public library would be a significant step toward an "electronic democracy", in which information is available to all rather than to a select few. On the other hand, it has been argued that the ability to make detailed maps using GIS could be profoundly disruptive. For example, if the average citizen had the ability to map incidence of a life-threatening disease like cancer on a small area basis, public health authorities would be deluged by demands for investigation of apparent "hot spots". Consumer GIS will increase the ability of the general public to look at data from a spatial perspective, with obvious consequences for privacy and for the relevance of spatial thinking in our school systems.

DECISION-MAKING

In the final analysis, the purpose of every GIS is to support some kind of decision-making (Cowen, 1988), although decisions may range from trivial to profound, and the effectiveness of GIS at supporting decisions clearly varies. In the typical GIS project, one first makes a proposal to some funding agency. If the project is funded, data gathering begins, and may involve field work, digitizing, accessing remote file stores, and perhaps reformatting or restructuring. A model is often built as part of the process, and linked to the GIS in some fashion, and model building may require the writing of code, or adaptation of code from another source. Finally, perhaps six months into the project, analysis begins and decisions can be made.

The typical project then moves into decision-support mode. Decision-makers create "what if" scenarios, often in committee. These are passed to analysts, who explore their implications using the GIS and pass the results back to the committee.

With luck, the entire process of preparation might take six months, as suggested earlier. If the system is powerful enough, each "what if" scenario might take 30 minutes to execute. Someone less used to promoting the benefits of GIS might realistically ask why, given this exceedingly sluggish pace, GIS has become so popular as a decision support tool, particularly in the public sector. One might suggest in response that the pace is one with which many public sector decision-makers might feel comfortable, or that the benefits of GIS clearly outweigh the disadvantages of delay. But it seems clear that if developers of GIS are serious about the technology's potential for decision support, something must be done to compress its time scale. We must move from a norm of unique projects that take many months, to one of continuous monitoring and modeling of the environment and its changes.

Faster computing will help, and as noted earlier, computers will continue to compute faster. The time taken to compute a "what if" scenario using complex environmental models will drop, allowing scenarios to be run within the time constraints of committee meetings. Computers will also continue to get smaller, allowing them to be taken directly into the decision-making arena. CSCW, and technologies to share decision-making over networks, will reduce the cost and time taken to bring decision-makers together. Better user interfaces will allow better merging of the roles of analyst and decision-maker. GIS developers will continue to add decision-making tools into GIS.

More fundamentally, as GIS and other spatial data technologies mature, it will be easier to avoid the constraints that follow from thinking about a GIS as a container of maps (Goodchild, 1988). This was a useful metaphor in the early days of GIS, as database layers were created by digitizing the contents of maps. But maps are static representations of a world that is in reality constantly changing. Just as the map is the end product of cartography, it is easy to think of the database as the end product of GIS, rather than the decisions that will flow from its analysis. New developments in object oriented databases, temporal GIS, and spatial analysis are helping to move our thinking away from the map metaphor, towards a more flexible concept of a spatial database. We are seeing changes in the way data are collected, geared to monitoring and decision-making rather than mapping, and emphasizing continuous sampling rather than time-slice views of the world.

The problem of excessive database development time in GIS projects will also be helped by the further development of standards, and it is encouraging to see how much progress has been made in standards development for geographic data in various parts of the world in the past few years (Morrison, 1992). Data format standards are helpful, but ultimately they help by minimizing only the time it takes to transfer and reformat data from one database to another. Of much greater importance in the long run to decision-making is the accessibility of data, and the ability of a user to search for and find data that fits the requirements of the study. Communication networks and network servers will help significantly to improve access, as will standards for data documentation and data about data, or metadata. We hope that a metadata standard for spatial data will be in place in the U.S. by the end of this year, and there are indications that the major vendors will respond quickly by adding metadata handling routines compatible with the standard to their systems.

A final factor in reducing the time taken to make decisions using GIS is education. While the decision sciences have developed numerous methods for decision support, GIS developers have only just begun to incorporate these into standard products, and it will be a while before GIS users are familiar with them. Unfortunately the education system is very conservative and slow to change, and it will take time for a generation of GIS users familiar with methods of spatial decision support to have an influence on practice.

DATA QUALITY

The fourth area is the one closest to my own area of research interest, in data quality and models of error in spatial databases. In this short paper I can only make a brief case for the data quality issue, and make some very general comments on future directions in this area. Some very significant basic research on spatial data quality has been completed in the past few years (for a review of NCGIA research on this topic see Goodchild, 1992), and we now know much more than we did about how to build sensitivity to data quality into our spatial data handling systems. We have good models of error, methods for storing their parameters in databases, techniques for propagating the effects of error or uncertainty into the products of GIS operations, and for visualizing their effects.

In ten years, I believe we will have a very different approach to the data quality issue. Using the capabilities of the digital environment, we will no longer present geographic data in ways that mask or hide known uncertainties. Soil or land cover maps will no longer be presented with large areas of uniform class, separated by infinitely thin lines. We will have moved away from the current practice of showing the world as simpler than we know it to be, with precisely positioned crisp objects and double precision coordinates. Traditionally, we describe the quality of a database using surrogate parameters such as the scale of the map from which the data were digitized, and the positional accuracies achievable on maps of given scales are well known, and expressed in national map accuracy standards. Less well known, however, are the consequences of given positional accuracies or misclassification probabilities for measures like polygon area. Unfortunately, simple calculations show that the results of operations like area calculation in a GIS are much less reliable than generally expected, and the principle that results should be reported with no greater precision than is justified by the accuracy of the data is frequently violated by users and designers of GIS software.

Like it or not, this is an issue that will not go away, as GIS becomes less mysterious and its results are challenged to a greater degree, particularly in court. We already have the tools in the research community to take a much more intelligent approach to accuracy, and these are rapidly becoming available for routine application. In ten years, we will no longer be surprised at the high levels of uncertainty associated with results derived from a database of known accuracy.

EDUCATION

As an academic, I would like to finish this paper with some comments on the state of GIS education, and its future directions. The higher education sector was quick to grasp the opportunities of GIS, and according to a recent survey (Morgan and Fleury, 1993), programs in GIS are now offered in well over 500 universities and colleges worldwide. On the other hand, relatively little progress has been made in introducing GIS into primary and secondary education. Our own project in secondary education at UCSB suggests that several impediments stand in the way of GIS in the secondary schools, despite a great amount of interest in many disciplines. Although GIS is seen as an interesting way to teach geography, problem solving skills, environmental studies, and spatial concepts, there is a lack of readily available materials, particularly for the kinds of technical resources found in many schools. Teachers have little access to GIS resources, and face pressure from a host of competing subject matter.

Another sector poorly served to date by GIS educators is continuing education. GIS has grown quickly, and many people encounter it for the first time long after they have left the formal education system. Their ability to retrain is often limited by geography, family, and financial constraints.

Recently, a consortium of European universities known as UNIGIS has announced a program of GIS through distance learning. Aimed directly at the professional sector, UNIGIS offers GIS in a manner that can take advantage of new communication technology and at the same time avoid many of the traditional impediments. What could be more appropriate to GIS than a global course supported through electronic communication?

SUMMARY

To summarize briefly, this paper has argued that continued improvements in hardware, in the form of computing speed and capacity of workstation platforms, is assured over the next ten years. GIS will continue to grow, and it has already shown itself to be remarkably recession-proof.

In response, four significant trends are anticipated. Further development of communication networks will provide a very exciting series of developments that will improve accessibility of spatial data; lead to the further growth of informal networks in the worldwide GIS community; and subvert many of the existing impediments and barriers to GIS applications. Vertical organizational structures will continue to decline, and horizontal linkages will continue to strengthen.

Second, the application of spatial data technologies will expand rapidly in the consumer sector, leading to an increase of at least two orders of magnitude in the installed base of GIS. GIS will begin to touch our daily lives, through routine tasks such as vehicle navigation, or finding restaurants in strange cities. Spatial data will be a significant part of the revolution in household electronic communication.

Third, the role of GIS in decision-making will continue to evolve, as the GIS project life cycle is compressed, and GIS moves into a mode of true decision support not confined to government agencies. The prospect of GIS wars, as GIS appears in support of both sides in litigation, will become increasingly likely.

Fourth, the next decade will see significant progress in the data quality issue, through improved techniques of data quality description, standards, error propagation, and visualization.

Finally, the next decade will see the introduction of GIS concepts - spatial data, spatial thinking, and spatial problem solving - into the primary and secondary curricula. Professional or continuing education in GIS will emerge as an important sector, based on new technologies exploited in distance learning.

ACKNOWLEDGMENT

The National Center for Geographic Information and Analysis is funded by the U.S. National Science Foundation, grant SES 88-10917. Further information on its activities can be obtained from the Publications Office at the author's address given above, or by writing to any of its three sites - the University of California, Santa Barbara; the State University of New York at Buffalo; and the University of Maine. Many of the center's reports are available by anonymous ftp from ncgia.ucsb.edu.

REFERENCES

Bowers, J.M., and S.D. Benford, eds. 1991. *Studies in Computer Supported Cooperative Work: Theory, Practice, and Design*. Amsterdam: North-Holland.

Burrough, P.A., 1986. *Principles of Geographical Information Systems for Land Resources Assessment*. Oxford: Clarendon.

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

Goodchild, M.F., 1988. Stepping over the line: technological constraints and the new cartography. *The American Cartographer* 15: 311-320.

Goodchild, M.F., 1992. *Research Initiative 1: Accuracy of Spatial Databases Closing Report*. Santa Barbara, CA: National Center for Geographic Information and Analysis.

Krol, E. 1992. *The Whole Internet: User's Guide and Catalog*. Sebastopol, CA: O'Reilly and Associates.

Maguire, D.J., M.F. Goodchild, and D.W. Rhind, editors, 1991. *Geographical Information Systems: Principles and Applications*. London: Longman Scientific and Technical.

Mark, D.M., and E.A. Zubrow, 1993. Join the GIS-L electronic community! *GIS World* (May) pp. 56-57.

Morgan, J.M., III, and B.B. Fleury. 1993. Academic GIS education. *Geo Info Systems* 3(4): 32-40.

Morrison, J.L., 1992. Implementing the spatial data transfer standard - introduction. *The American Cartographer* 19(5): 277.

Rhind, D.W., 1993. Data data: Another view: what's next for national mapping agencies? *Geo Info Systems* 3(4): 26-29.