PROCESS ON THE GIS RESEARCH AGENDA

Michael F. Goodchild
National Center for Geographic Information and Analysis
University of California
Santa Barbara, CA 93106, USA

ABSTRACT

GIS research is defined as investigation of the generic issues surrounding the effective use of GIS technology. Some of the important scientific questions underlying GIS research are identified. The paper reviews the current state of research in a series of key areas, and speculates on why progress has been so uneven. The final section of the paper looks to the future and to new areas of significant potential in GIS research.

BACKGROUND

It seems appropriate to begin any overview of GIS research with a disclaimer. First, it is not at all clear what GIS research is, or what role it plays in the larger GIS community and industry. What I present in this paper is in many ways my own view, and I would expect it to be challenged. I think my own biases will become clear in what follows. Second, because of the field's diversity and dynamism it is difficult if not impossible for any one individual to attempt a general overview. What follows is therefore almost inevitably incomplete and uneven.

Research is often identified as either pure or applied - driven by basic and innocent human curiosity or by the practical everyday needs of human society. Much of GIS is a response to human needs for information management and analysis, and in that sense one might expect GIS research to be more applied than pure. But one view of pure research is that it is research that has not yet found application; pure research is a long-term investment just as applied research is a short-term investment. From an academic perspective, pure research is often associated with higher prestige, but applied research with greater funding. In the paper I have tried to cover the full range from pure to applied, feeling that both are important to GIS. At the same time "basic research" is the primary purpose of the NCGIA, and the Center is very fortunate in being funded to do research whose applications may lie years or even decades in the future.

The paper is divided into three sections. The first attempts to come to grips with this indefinite nature of GIS research, by clarifying its content and limits. The second reviews progress in key areas. Finally the third section looks at what still needs to be done, and at prospects for the future.

WHAT IS GIS RESEARCH?

During the design phase of the Canada Geographic Information System (CGIS) in the 1960s it became clear that the only practical way to input the large number of maps needed would be by some form of scanning device (Tomlinson, Calkins and Marble 1976). At the time no scanner for map-sized documents existed, and it was necessary to invent one. A prototype drum scanner was built by IBM Canada and successfully tested, at what by modern standards would be regarded as vast expense. Other parts of the CGIS design team were busy

inventing other, equally fundamental and now familiar solutions to technical GIS problems such as the Morton order.

In the almost three decades of GIS development that are now behind us, similar "how-to-do-it" research has produced a large number of algorithms, data structures, spatial indexing schemes, and other technological solutions. Some of these are unique to GIS, but many have been reinvented in several related disciplines. The Morton order, for example, occurs in the literature of several spatial data handling fields under different names (Samet 1989), and descriptions of algorithms for finding Thiessen polygons are spread over a wide range of journals. At the same time there is a growing sense in GIS research that our emphasis has changed, as more and more of the underlying technical problems of GIS are solved. Attention has moved from primitive algorithms and data structures to the much more complex problems of database design, and the issues surrounding the use of GIS technology in real applications.

In a paper given at the Fourth International Symposium on Spatial Data Handling in Zurich in July 1990 (Goodchild 1990), I argued that research had moved from an emphasis on Geographic Information Systems to something approaching a Geographic Information Science. The effective use of GIS technology raises a host of concerns, all of them in some way related to the peculiarities of geographic information. Geography is infinitely complex, but must be represented digitally as a finite collection of discrete objects. Maps also represent geography as a finite collection of discrete features, and humans also discretize the world in order to describe it, learn and reason about it and navigate through it. ULtimately it is the wide range of options available in this process of discretization that makes geographical data modeling so complex.

Here is a sample of these generic issues, all of them important to the successful implementation of GIS, but all of them essentially independent of the technical issues of hardware and software:

- How to compile an accurate representation of geographical variation for input to a database?
- How to convert a hard copy representation to digital form without ambiguity?
- How to structure a given representation for rapid access and processing?
- How to represent the uncertainty or inaccuracy present in a digital representation?
- How to propagate uncertainty from database to GIS products?
- How to handle very large quantities of geographical data efficiently?
- How to display geographical information to facilitate human perception?
- How to make effective use of geographical data and analysis in spatial decision-making?
- How to measure the costs and benefits of GIS?
- How to assess the impacts of GIS technology on organizations?

In the NCGIA research plan (NCGIA 1989), we argued that the absence of solutions to issues like these constituted impediments to the effective application of GIS technology. Other discussions of the GIS research agenda have come to similar conclusions, although with different emphases (Craig 1989; Maguire 1990; Masser 1990). Many are old issues, recognized long before the advent of GIS in fields like cartography, geodesy, and geography. Some may not be unique to GIS. For example, it is not immediately obvious that GIS technology diffuses in a fundamentally different fashion, or shows fundamentally different patterns of adoption from other technologies. Is GIS benefit measurement a unique problem, or an example of the more general problem of measuring the benefits of information technology? Of course these questions are in themselves research issues.

At the same time I think it is very important to point out the areas where GIS has created new and unique issues that are not common to other fields. In the early days of GIS, it was possible to argue that the technology was filling an existing gap, and making possible tasks that had been previously identified, but that were not easy to carry out manually. The use of GIS for suitability analysis, by overlaying layers (Tomlin 1990), mirrors the manual technique popularized by McHarg, while admittedly adding some interesting new capabilities. CGIS was justified on the grounds that the computer was a cost-effective alternative to hand measurement of overlaid areas. But GIS makes it possible to do things with data that the data's gatherers may never have envisioned. GIS technology is producing radical changes in the way geographical data is collected, handled and analyzed, and it will be many years before even the impact of existing technology is felt, let alone the impacts of future developments.

Here are some of the issues that seem unique to GIS:

- How to model time dependent geographical data?
- How to capture, store and process three dimensional geographic data?
- How to model data for geographic distributions draped over surfaces embedded in three dimensions?
- How to explore such data - for example, what exploratory metaphors are useful?
- How to evaluate the geographical perspective on information and processes relative to more conventional perspectives?

These are important issues for GIS, and the GIS community needs a strong commitment to research if it is going to make significant progress on them. As issues that arise within the context of GIS, they are not of major concern in other disciplines. But at the same time the GIS community can benefit enormously from interdisciplinary research. Statisticians can make a very valuable contribution to solving the error problem in GIS, and research in cognitive psychology may be helpful in designing the cognitive aspects of GIS user interfaces.

So to sum up, let me propose a definition of GIS research:

- Research on the generic issues that surround the use of GIS technology, impede its succesful implementation, or emerge from an understanding of its potential capabilities.

Is this "research about GIS" or "research with GIS"? In a sense it is both because these are issues that are both fundamental to the technology of GIS and also issues that must be solved before the technology can be successfull applied. If the problems of doing research with GIS are generic, then the are best tackled as part of the GIS research agenda. However problems tha are specific to the application of GIS in a particular field clearly need t be addressed in the context of that field, and with the benefit of it expertise. Accuracy issues provide a useful example. There are aspects o the accuracy problem that span a wide range of types of geographic data, an need to be solved using generic models of uncertainty, analogous to the rol played by the Gaussian distribution in the theory of measurement error. Bu an analysis of crime data using a GIS will also raise problems of accurac that are specific to that particular application, and need an understandin of the processes operating in criminology and in crime data collection if the are to be understood fully.

With this background, the next section looks at progress already made on the GIS research agenda.

WHAT PROGRESS HAS BEEN MADE?

In pure research, indicators of progress derive from publication, particularly in the refereed journals. Thus progress on the GIS agenda might be assessed by examining recent pages of the *International Journal of Geographical Information Systems*. But because GIS is also an applied research field, the publication of a result is not necessarily progress. From this perspective, the field only benefits from research results if they are actually implemented, and it is to some extent incumbent on the GIS research community to make sure that this happens. In the following discussion I have tried to make this essential distinction, and to identify cases where the set of pure research results is substantially ahead of applied implementation.

The section is organized in a linear fashion, from research on data collection and capture through to research on the decision-making process and the organizational impacts of GIS.

Data collection

The process of discretization, with its implied generalization, abstraction and approximation, takes place as data is collected, interpreted or compiled, and choices are made at this stage that affect the ultimate uses of the data. When those uses change, as they have been doing with the widespread use of GIS, it may be necessary or beneficial to rethink the data collection process. For example, with digital management and delivery of census data, is it still appropriate to conduct a census on a decennial basis? Is the traditional approach to geologic field mapping the most appropriate if the eventual objective is a digital 3D representation of the subsurface? How will topographic mapping change now that it is cost-effective to survey new features using GPS? Geographical data collection is often the domain of specialists in well-established disciplines, so it may be many years before these kinds of questions are investigated or answered. To date the introduction of GIS seems to have had very little effect on the data collection process.

Data capture

Enormous strides have been made in the technology for capturing digital geographic data in the past decade, and the systems now on the market are capable of a high level of intelligence in interpreting scanned map documents. The problem remains the poor quality of the documents, and the ambiguities that are caused by aspects of map design. As a result, manual digitizing remains a widely used approach, despite its high cost, tedium and failure to show significant improvements in efficiency. Two trends may change this situation substantially in the next few years. One is the increasing avoidance of the map document as a step in the data compilation and input process. Surveying and photogrammetry are moving away from compilation using paper maps, and the more interpretive fields such as land use, vegetation or soil mapping are likely to follow suit. The digital total station is likely to be followed by the digital plane table and perhaps even the digital field geology notebook. The other is the long recognized possibility that comparatively minor changes in a map's design can make it vastly easier to scan and interpret (Shiryaev 1987).

Data modeling

Of all the developments in GIS in the past decade, perhaps the most exciting has been the proliferation of data models, and the growing literature on their relative merits. The debate over raster and vector goes back to the earliest days, but has now been joined by debates over objects, layers, the philosophy of object orientation, hierarchical models of complex objects, and the entire range of possibilities inherent in time dependence and three dimensions. Despite the interest, we still do not have a complete and rigorous framework for geographical data modeling, even in the static two-dimensional case, and without one it is difficult to see how GIS can escape the constraints imposed on it by specific system implementations. How much capability is being lost by forcing contemporary applications into the multilayer raster model used by many systems, or the point/line/area coverage model used by many others? This is both a pure and an applied research problem. On the one hand, GIS must develop a comprehensive framework for geographic data modeling, with an associated terminology, to provide the basis for standards and an ideal against which specific systems can be measured. On the other hand, an abstract framework is of little value if it does nòt influence practice, through implementation in vendor products. Here the real issue is whether it is possible to enlarge or "retrofit" the data model underlying an existing product, or whether any attempt to do so is doomed to cause inconsistency and incoherence.

Data structures

The details of schemes for formatting geographical data models in digital files were major concerns in the early days of GIS. As volumes have grown, the need for efficient indexing schemes to access data quickly has also led to productive research. But many systems now handle data through database management systems, and data structure issues have moved more and more into the realm of computer science.

Algorithms

Similar comments can be made about algorithm development, which has also moved out of the limelight of GIS research in recent years. Early papers on Thiessen polygons in the GIS literature were followed by a host of articles in computer science as research in that discipline recognized the significance and challenge of the problem, and there have been similar patterns in other areas as well. Early developments in GIS, as in many other fields, were made using primitive tools like Fortran, but as software has evolved, the

sophistication of the development environment has moved up several levels.

Analysis

Despite widespread recognition that analysis is central to the purpose of GIS, the lack of integration of GIS and spatial analysis, and the comparative simplicity of the analytic functionality of many systems continues to be a major concern. In the early days of the statistical package SAS, there was a very rapid increase in the range of tests and techniques implemented in the system. Unfortunately the same has not been true of GIS, and remarkably little progress has been made at incorporating the range of known techniques of spatial analysis into current products.

There are many reasons for this. One obvious one is the heavy emphasis in the GIS marketplace on information management rather than analysis. The lucrative markets for GIS technology have comparatively unsophisticated needs, emphasizing simple queries and tabulations. Another is the relative obscurity of spatial analysis, a set of techniques developed in a variety of disciplines, without any clear system of codification or strong conceptual or theoretical framework. Even now it is difficult to identify more than a handful of texts (e.g. Haining 1990; Upton and Fingleton 1985). While one might expect that GIS could provide the basis for a system of codification for spatial analysis, the poor level of current understanding of geographic data models is a major difficulty.

At this stage, integration of GIS and spatial analysis is proceeding slowly, in at least three different modes. Some analytic capabilities are being added directly to GIS, for example in the recent expansion of functionality in several network analysis modules. Some progress is being made in loosely coupled analysis, where an independent analysis module relies on a GIS for its input data, and for such functions as display. But still missing is an effective form of tight coupling, in which data could be passed between a GIS and a spatial analysis module without loss of higher structures, such as topology, object identity, metadata, or various kinds of relationships. At present this is impossible, to a large extent because of a lack of standards for data models. Instead, coupling has to occur at a lower level, and higher structures have to be rebuilt on an arbitrary basis.

Integration between GIS and spatial analysis might also take the form of a language, whose primitive elements would represent the fundamental operations of spatial analysis. The beginnings of such a language already exist in the macro languages of many of the current generation of GIS, and in various attempts to extend SQL to spatial operations. But all of these are specific to, and heavily dependent on limited data models, and there is remarkably little similarity between them at this time. At Santa Barbara we have been attempting to define a common language from an analysis of the languages used by a variety of current GIS, but a more satisfactory solution would begin with the conceptual framework provided by a comprehensive data model.

Another problem in integrating GIS and spatial analysis is that in the former discretization of space is explicit, whereas in many forms of spatial analysis it is often either implicit, or unspecified. Many forms of spatial analysis are written on continuous fields, and fail to deal with the uncertainties introduced by the inevitable process of discretization. For example, in GIS there can be no measure of slope that is independent of discretization, and similarly the length of an area object's boundary is dependent on its digital representation. Yet slope and length commonly appear as unqualified parameters in spatial models. In this sense, integration of GIS and spatial analysis is a two-way process, in which the inadequacies of both GIS and spatial analysis must be addressed.

Accuracy

Because of discretization, all geographical data is uncertain to some degree. However all of the current generation of GIS follow the common practice in cartography, and represent geographical objects as if their positions and attributes were perfectly known - data quality may or may not be addressed in a separate statement. The consequences of uncertainty for GIS products are never estimated.

Recent research has followed several different and productive lines in attempting to address the data quality problem. One is to match precision to accuracy. In a locational sense, this means using limited precision in data representation and processing, most often through the use of a raster whose size is determined by data accuracy. Various forms of quadtree structure have also been used to fit locational precision to known levels of accuracy. There have been several recent papers on finite resolution processing in GIS (e.g. Franklin 1984; Dutton 1989), and finite resolution geometry is an active research area in mathematics.

Another productive approach has been to incorporate techniques from geostatistics, notably Kriging, since the statistical basis of these techniques makes uncertainty explicit. We now have several useful models of digitizing error, and its consequences for estimated measures such as area (e.g. Chrisman and Yandell 1988; Keefer, Smith and Gregoire 1988). Finally, there have been several successful efforts to model geographical data sets as random fields, or derivatives of random fields, and to use this approach to model uncertainty in GIS objects (e.g. Goodchild 1989). Between all of these methods, we probably now have an adequate set of models of accuracy from which to build an error-tracking GIS.

However spatial statistics is not an easy field, and many of these techniques go well beyond elementary statistics in their conceptual sophistication. Particularly in error modeling, and to some extent in spatial analysis generally, it is easy to go far beyond the understanding of the average GIS user. We run the risk of destroying a powerful argument for GIS - its ability to deliver spatial analysis. If we are successful at integrating spatial analysis and GIS, then we will create a large demand for courses in the essential underlying methods and concepts. We need to think of GIS courses as providing much stronger grounding in these areas, and as much less concerned with the technical details of operating a system. The user of a GIS needs to be thoroughly sensitized to the issues and problems of geographical data.

Lineage

GIS analysis is often a multistage process, as different layers and sets of objects are processed and combined. Much recent research has gone into developing tools for tracking multistage analyses, notably in propagating error and in managing intermediate products. The results are just beginning to appear as practical tools for spatial decision support.

Institutional and managerial issues

Research is just beginning to appear on the issues involved in implementing and managing GIS, especially in large institutions. This is difficult research, and generalizations are not discovered easily. But the success of several large projects in the US, and the discussions surrounding several large acquisitions by federal agencies, have created the opportunity for a number of useful case studies. Many more are needed, particularly given the importance of such research for improving the institutional environment in the future.

WHAT REMAINS TO BE DONE?

Looking back over nearly three decades of GIS research, it is clear that the greatest progress has been made on the best-defined and easiest problems, where solutions lay in advances in the technology itself. Rapid progress was made on algorithms and data structures in the 1970s and 1980s, but many of the hard problems of data modeling, error modeling, integration of spatial analysis, and institutional and managerial issues remain. Some of these may be unsolvable - for example, there may simply be no generalities to be discovered in the process of adoption of GIS by government agencies, however easy it may be to pose the research question.

Other issues have already been solved in a pure research sense, but implementation remains a major applied research question. In accuracy, for example, a substantial set of techniques has been defined, but the problem of moving them into actual application remains. The academic research environment is set up to pursue significant areas of research, but is generally poor at providing the means of implementation. For that we need a software industry that is tightly coupled to the research community, but able to find the resources to motivate development. But more importantly, we need an education system that responds rapidly to new research, and is able to build new concepts quickly into its programs. Unfortunately the higher education sector is too often characterized by conservativism, and it may take many years for new ideas to work themselves into the curriculum.

GIS research is like geographical data - the more closely one looks, the more interesting issues appear. GIS research has only begun to tackle the important issues in the research agenda. We are in an enviable position, working in a field with such strong motivation and such a strong underlying industry, and with such an interesting set of problems spanning so many disciplines and fields.

ACKNOWLEDGMENT

The National Center for Geographic Information and Analysis is supported by the US National Science Foundation, grant SES 88-10917.

REFERENCES

Chrisman N R, Yandell B (1988) A model for the variance in area *Surveying and Mapping* 48: 241-246

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

Dutton G (1989) Modeling locational uncertainty via hierarchical tesselation. In Goodchild M F, Gopal S (eds.) *Accuracy of Spatial Databases* Taylor and Francis, London, pp 125-140

Franklin W R (1984) Cartographic errors symptomatic of underlying algebra problems *Proceedings, International Symposium on Spatial Data Handling, Zurich* University of Zurich, pp 190-208

Goodchild M F (1989) Modeling error in objects and fields. In Goodchild M F, Gopal S (eds.) *Accuracy of Spatial Databases* Taylor and Francis, London, pp 107-114

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

Haining R (1990) *Spatial Data Analysis in the Social and Environmental Sciences* Cambridge University Press

Keefer B J, Smith J L, Gregoire T G (1988) Simulating manual digitizing error with statistical models *Proceedings GIS/LIS 88* ASPRS/ACSM, Falls Church VA, pp 475-483

Maguire D J (1990) A research plan for GIS in the 1990s *The Association for Geographic Information Yearbook 1990* Taylor and Francis, London, pp 267-277

Masser I (1990) The Regional Research Laboratory initiative: an update *The Association for Geographic Information Yearbook 1990* Taylor and Francis, London, pp 259-263

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

Samet H (1989) *The Design and Analysis of Spatial Data Structures* Addison-Wesley, Reading MA

Shiryaev E E (1987) *Computers and the Representation of Geographical Data* Wiley, New York

Tomlin C D (1990) *GIS and Cartographic Modeling* Prentice Hall, Englewood Cliffs NJ

Tomlinson R F, Calkins H W, Marble D F (1976) *Computer Handling of Geographical Data* UNESCO Press, Paris

Upton G J, Fingleton B (1985) *Spatial Data Analysis by Example* (2 volumes) Wiley, New York

---

DATA INTEGRATION IN FUTURE VERSIONS OF 'ILWI

Ben Gorte
ITC
P.O. Box 6
7500 AA Enschede
The Netherlands

ABSTRACT

Since the beginning of 1986, an interdisciplin
has been working on the development of a Geograp
System software, which to date is known as th
From the beginning the aim has been to dev
Geographic-data storage and manipulation, toget
for information extraction.

Because of its multidisciplinary nature, the em
on the support of a variety of different data
vector and non-spatial) on the combination of
different types, and on the goal of making the to
purpose" as possible. Indeed, the system has
suitable for a large number of applications in d
of earth science.

As a consequence, however, the system has
complicated to operate in certain circumstances. L
it turned out that the capability to perform com
complicates performing "routine" tasks.

Currently, design considerations are being made
of removing this drawback while preserving the f
solutions are found in a more consistent Data
System approach for Geographic-data, including
concepts such as Data Dictionary and User Views
allowing the "combination" of data of different ty
should allow the "integration" of these data,
in a uniform manner using (at least from the
user) a single Geographic-Data Base Mana

otho
surveys dat
raster emphas.
special the effic
So far Howe
successful. version
a future import
of the most
expert, not only in