KEYNOTE ADDRESS

SPATIAL INFORMATION SCIENCE

Michael F. Goodchild
National Center for Geographic Information and Analysis
University of California
Santa Barbara, CA 93106, USA
email: good@voodoo.bitnet

Abstract

The papers at this conference address a set of intellectual and scientific questions which go well beyond the implications of its title. The paper reviews the topics which might be included in a geographic information science. Research on these fundamental issues is a better prospect for long term survival and acceptance in the academy than development of technical capabilities.

Introduction

At the opening of the Fourth International Symposium on Spatial Data Handling it seems appropriate to reflect on the nature of the field, on progress to date and on forecasts for the future. The spatial data handling community has come a long way in the six years since the first Zurich meeting, and this is particularly true if one equates spatial data handling with GIS (or regards GIS as a subset or superset). Major research and training programs have been established in a number of countries, new applications have been found, new products have appeared from an industry which continues to expand at a spectacular rate, dramatic improvement continues in the capabilities of platforms, and new and significant datasets have become available. It is tempting to say that this meeting, and this group, is simply a part of this much larger enthusiasm and excitement, but there ought to be more to it than that.

What, after all, is spatial data handling? It may describe what we do, but it gives no sense of why we do it. This was one of the themes behind Tomlinson's keynote address at the first Zurich meeting in 1984 (Tomlinson 1984). It suggests that spatial data is somehow difficult to handle, but will that always be so? It suggests a level of detachment from the data themselves, as if the USGS were to send out tapes labeled with the generic warning "handle with difficulty". It is reminiscent of the name of the former IGU Commission on Geographical Data Sensing and Processing. A quick review of the titles of the papers at this meeting should be enough to assure anyone that we are concerned with much more than the mere handling and processing of data. We are more than the UPS of GIS.

GIS is often accused of being technology driven, a technology in search of applications. That seems to me to be more true of some periods of the 25-year history of GIS than of others. For example, it is difficult to suggest that

Tomlinson and the developers of CGIS were driven by the appallingly primitive hardware capabilities of 1965. On the other hand the prospect of a menu-driven, full color, pull-down menu raster GIS in the 386-based PC on one's desk has clearly sold many systems in the past few years. Technological development comes in distinct bursts, and so does the technology drive behind GIS. It may be the motivation behind the desire to handle spatial data, but it fails to explain many of the diverse research efforts being reported at this meeting.

There have also been phases in the history of GIS when the drive has been from applications. CGIS itself was an application in search of a technology, and the drive was sufficiently strong to lead to the prototyping of the first map scanner, and to numerous other technological developments. McHarg had worked out the principles of the map overlay technique long before Berry and others automated them in MAP and its derivatives. And school bus routing software has been around much longer than the problem's implementation in a standard GIS. But again, much of the subject matter of this meeting lies well beyond any reasonably foreseeable application.

There are also indications that some portions of the GIS industry are data driven. The widespread distribution of Landsat and SPOT imagery, and the availability of DEMs and street files in many countries have certainly led to applications well beyond those used to justify the data's compilation. TIGER, for example, appears to be spawning its own industry of updaters, repackagers and application developers, although it exists in principle only to serve the needs of the 1990 US Census.

But although the GIS driving seat is undoubtedly crowded, I would like to deal in this paper largely with the fourth driver located apparently irrelevantly in the back seat, the "S" word. It seems to me that there is a pressing need to recognize and develop the role of science in GIS. This is meant in two senses. The first has to do with the extent to which GIS as a field contains a legitimate set of scientific questions, the extent to which these can be expressed, and the extent to which they are generic, rather than specific to particular fields of application or particular contexts. To what extent are we at this meeting driven by intellectual curiosity about the nature of GIS technology and the questions that it raises? And if GIS can be motivated by science, then what are its subfields, what are its questions, and what is its agenda? The second sense has to do with the role of GIS as a toolbox in science generally - with GIS for science rather than the science of GIS. What do we need to do to ensure that GIS, and spatial data handling technology, plays its legitimate role in supporting those sciences for which geography is a significant key, or a significant source of insight, explanation and understanding?

To do this we must first establish that spatial, or rather, geographical, data is unique, and that its problems cannot therefore be subsumed under some larger field. We must also establish that there are problems which are generic to all geographical data, or at least to establish that it is possible to distinguish those that are from those that are not. For example, the accuracy of attributes on a choropleth map of crime statistics would seem to be very little informed by knowledge of attribute accuracy for geographical data generally, but to require

instead a level of understanding of the specific problems of crime statistics. But the accuracy of population estimates for an arbitrarily defined polygon may well be knowable from, or at least informed by, the general properties of the Modifiable Areal Unit Problem.

What is unique about spatial data?

In many facilities management systems, the role of the GIS is to provide an alternative key to data, a method of access based on geographical location. In essence a spatial database has dual keys, allowing records to be accessed either by attributes or by locations. But dual keys are not unusual. The spatial key is somewhat distinct, as it permits operations to be defined which are not included in standard query languages. For example, it is possible to retrieve all point records lying within an arbitrary, user-defined polygon, an operation which is not defined in standard query languages like SQL. In essence the spatial key is multidimensional, but again multidimensional keys are known from other areas, and analogs of point in polygon retrieval can be defined for non-spatial dimensions.

What distinguishes spatial data is the fact that the spatial key is based on two continuous dimensions. It is possible to visit any location (x,y) in the real, geographical world, defined in principle with unlimited precision, and return a value for a variable, for example topographic elevation z. Terrain is thus characterized by an infinite number of tuples <x,y,z>. In network applications z is defined only for locations on the network, but the number of tuples is still infinite if variation is continuous along this one-dimensional structure of links and nodes. Time series also have continuous keys, but aré rarely conceived, measured or represented as continuous, and there appears to be little commonality of interest in the problems of temporal data handling. By contrast, there is ample evidence of commonality in the spatial data handling disciplines.

Many of our data models, particularly polygon networks and TINs, reflect an underlying view of space as continuous and the need to accommodate the user who wishes to determine z at some arbitrary, and highly precise (x,y). One implication of this is that there exists a multiplicity of possible conceptual data models for spatial data, and that the choice between them for a given phenomenon is one of the more fundamental issues of spatial data handling.

Another distinctive feature of spatial data is what Anselin (1989) refers to as spatial dependence, the propensity for nearby locations to influence each other and to possess similar attributes. Without spatial dependence, there would be no reasonable prospect of creating even approximate views of continuous spatial variation within a discrete, finite machine. It is not uncommon for tuples which have similar values of a key to have similar values of other attributes, but the structure of spatial dependence is unusual, relying as it does on both dimensions of the (x,y) key, with similarity determined by a metric.

Finally, geographical data is distributed over the curved surface of the earth, a fact which is often forgotten in the limited study areas of many GIS projects.

We have worried for centuries about how to portray the earth's surface on a flat sheet of paper, and have developed an extensive technology of map projections. But as a result we have few methods for analyzing data on the sphere or spheroid, and know little about how to model processes on its curved surface. Moreover, we tend to have treated GIS displays as if they were virtual sheets of paper, and insisted on viewing geographical data as if they were projected to a flat surface, instead of exploiting the potential of electronic display to create views of the globe itself. We need to develop the appropriate techniques for working with the globe, and making use of solid modeling rather than conventional 2D graphics, if we are to understand geographical processes at the global scale and contribute effectively to global science. We must rescue the orthographic projection from its present obscurity.

The content of geographic information science

Having established that geographic information has unique properties and problems, we can now review the set of generic questions which might make up a geographic information science. This can be done in a largely linear fashion, from data collection to analysis, although some themes tend to cut across this simple arrangement.

Data collection and measurement

If spatial reality is continuous and subject to complex structures of spatial dependence, then how should it be compiled and measured? More generally, how do people perceive the real world of geographical variation, structure it and learn about it? Although many of these questions are part of the research agendas of remote sensing, photogrammetry, geodesy and cognitive psychology, the lines of demarcation are far from distinct. Should GIS or remote sensing worry about the problems of transfering information from one technology to the other, and more importantly making good sense of it? Is it GIS or remote sensing if ancillary geographic information is used to improve the accuracy of classification or an image is used to update a GIS layer? But ultimately it matters little to which of the many pigeonholes we assign each topic. There are undoubtedly substantive scientific questions here, which require a depth of understanding of the nature of spatial variation, and one person's remote sensing may well be another's geographic information science.

GIS has developed in an information-rich environment, in which it has often been sufficient to build a database from already compiled data. As a result we have rarely had to deal with raw data, or to confront the issues of data compilation. Yet information as compiled for maps may not be ideal for GIS purposes - it is not at all clear that the stylized representation of soils on a soil map is best for the varied applications being developed in the GIS environment, for example the modeling of soil loss and runoff. To date, GIS users and developers have been comparatively unaware of the issues involved in spatial data collection and compilation.

Spatial statistics

Because spatial data is always an approximation or generalization of reality, it is full of uncertainty and inaccuracy. A change of data model or scale can introduce a loss of information, as can digitizing or scanning. Processing in a finite machine also inserts its own form of uncertainty, although this is often insignificant in relation to the errors inherent in the data themselves. Many human geographical constructs are implicitly uncertain, including spatial objects ("Indian Ocean, Europe") and their relationships ("in", "across"). Whether we think of uncertainty in set theoretic terms through notions of fuzziness or in statistical terms through the calculus of probabilities, the study of spatial data uncertainty, its measurement and modeling, and the analysis of its propagation through the processes of spatial data handling is undoubtedly part of geographic information science.

Geographical data brings its own special set of problems to spatial statistics. Whereas in medical imaging the problem may be to determine the true location of objects from "dirty" pictures (Besag 1986), in geographical images there is often no clear concept of truth, since objects are often the products of interpretation or generalization. We need much better methods of measuring and describing uncertainty, particularly in the complex spatial objects common in GIS. We need better methods for dealing with the world as a set of overlapping continua, instead of forcing the world into the mold of rigidly bounded objects. Most of the answers to these questions will have to come from spatial statistics, but geographical information specialists must provide the motivation and the examples, and define the overall objectives and constraints.

Theories of spatial data

Data models are the logical frameworks which we use to represent geographical variation in digital databases. Since each must be an approximation, the choice between alternative models not only constrains the functions available, but also the accuracy of products. Data modeling is perhaps the most persistent and challenging issue in GIS. What was once a simple choice between a regular tesselation and a space littered with points, lines and areas has become a complex of entities and relationships, object orientation, TINs and DEMS, layers, shared primitives, etc. If anything, the old raster-vector debate has become an issue of internal representation, ultimately to be hidden from the user, since it is clear that a given view of the world can be represented in either way with little real loss of information, and that the terms are poorly defined at best. But the effects of the development of object-oriented databases and object-oriented programming systems in computer science are likely to be much more profound. Within GIS, they are precipitating lively discussion over the entire question of the degree to which we view, analyze, represent and model the world as discrete or continuous, as a collection of objects or a set of fields. Do we think in terms of variables with defined values everywhere in space, or of an empty space littered with possibly overlapping objects? In essence, these issues have brought GIS debate from the comparative obscurity of internal data structures to the much more general issues of how we understand geographical variation. Everyday human experience sees a world of objects, but the science of natural processes deals more with continuous variation. Thus the object oriented debate threatens to pit the New Agers against the embattled remnants of the Enlightenment, and what could be more stimulating than that?

Data structures, algorithms and processes

Much of the basic research results which have accumulated over the past two and a half decades in GIS concern internal representations of data, and the algorithms which operate on them. The quadtree, band sweep algorithms for overlay, analysis of computational complexity and the arc-node data structure are all intellectual breakthroughs of lasting significance. Many challenging problems remain, for example in the design of efficient algorithms to minimize overposting and in other areas of cartographic design, or in developing better methods for converting between various terrain data models. We seem, though, to have reached a point where all of the simpler, more generic problems have been solved, and where what remains is a set of difficult, context-specific problems. It seems clear, for example, that further advances in terrain data model conversion (for example from contour to TIN) will require much better understanding of the nature of terrain, and will perhaps have to be specific to terrain type (e.g. fluvial vs. glacial).

Display

GIS have often been criticized for failing to give adequate attention to principles of cartographic design, or for regarding the map as a simple store of information rather than a tool for communication. If we think of the database as the truth, then a map is no more than a store, since there is often a simple correspondence between objects in the database and objects on the map. But if the database is seen merely as an approximation of the geographical truth, then the design of output displays is critical, since it can affect the user's view of the world. Such simple things as the choice of background color, or the contrast between adjacent polygons can have a significant effect.

But the capabilities of electronic display go far beyond those of conventional cartography. We need research on the design of animated displays, 3D, the use of icons and metaphors in user interfaces, continuous gradation of color and tone, zoom and browse, multiple media including voice and pointing devices, multiple windows which allow simultaneous access to spatial and temporal series of multivariate data. We need to use the electronic medium to think far beyond improvements to the design of choropleth maps. All of these are fundamental problems to a science of geographic information.

Analytic tools

A GIS is a tool for supporting a wide range of techniques of spatial analysis, including processes to create new classes of spatial objects, to analyze the locations and attributes of objects, and to model using multiple classes of objects and the relationships between them. It includes primitive geometric operations such as calculating the centroids of polygons, or building buffers around lines, as well as more complex operations such as determining the shortest path through

a network. The functionality of leading products continues to grow, with no obvious end in sight.

Tomlin (1990) has made one of the few attempts to add some sort of structure or framework to the proliferation of GIS functions, which in the case of ARC/INFO is already around 10^3. We badly need a taxonomy of spatial analysis, developed perhaps from an enumerated set of data models, but going well beyond the primitive geometric operations.

Most of the current generation of GIS provide some sort of macro or script facility, allowing the user to define products from complex sequences of operations, but to invoke them with a single instruction. Although these often include the ability to construct customized environments and interfaces, they do not as yet provide tools which are specific to the needs of spatial analysis. One limited exception is Prime/Wild's ATB, a set of tools constructed on top of System/9 which allows the user to work with complex analyses, visualize their sequences and manage intermediate results. Tools like this will be needed increasingly if GIS is to move into an era of more sophisticated analysis and decision support, because it is not uncommon for relatively simple GIS products to involve processing tens of layers through similar numbers of primitive steps. We need to research methods for keeping track of data lineage and error propagation, backtracking to recover intermediate results, and preventing the user from combining operations in incorrect or meaningless ways. And we need research on ways of incorporating this sort of analysis into the GIS acquisition and planning process.

This emphasis on complex multistage analysis and the generation of products from a multilayered database seems very different from the research being reported at this meeting on knowledge-based systems, spatial reasoning and spatial query. One of the attractions of the GIS field is its breadth of applications, and the correspondingly extreme variety of environments for user interface design. In data modeling, the important question is not whether extended relational or object oriented models are better for geographic data, but what types of geographic data are best modeled by each approach. Similarly the important research issue in user interface design is to determine the optimal environment for each of the many types of GIS application. What is best for a vehicle navigation system may be entirely different from what is best for a forest resource manager with a deeply seated fear of keyboards and VDUs, either color or monochrome.

Tests of commonality

The six sections above have looked at various candidate areas for inclusion in a geographic information science. In each case there are clearly challenging scientific questions to be posed and researched. I have no reason to believe that the list is complete, or that there are not additional and substantive questions in other related areas. In each case the spatial context appears to be distinctive, although clearly it is more so in some cases than others. For example, we might debate whether the spatial context was distinctive in the area of decision theory, but the issue seems clearcut in the case of data modeling.

But mere existence of scientific questions is far from an adequate basis for a science. Is there a commonality of interest here? Can these subfields find sufficient basis for interaction that they will develop the lasting accoutrements of a science - journals, societies, texts, philosophers etc.? Will they behave as a group of scholars? Is there a valid analogy between the systems and science of geographic information on the one hand - tools supporting researchers - and statistical packages and statistics on the other? Statistics is a highly formalized discipline, but more technologically oriented groups can be found in such areas as exploratory data analysis (EDA), statistical visualization, and applied statistics. Certainly the relationship between science and tools is a stormy one at times, but nevertheless vital to the success of both. The ongoing debate over the value of statistical software in teaching statistics has interesting implications for the same issue in GIS.

I would like to look briefly at the arguments for a commonality of interest in geographic information science, first in principle and then in practice. The field is small - rhetoric about growth in the industry aside, no one would suggest that GIS is a major discipline. It is distinct, with its own reasonably unique set of questions. And it is certainly challenging and innately appealing. On the negative side, it is multidisciplinary, competing with longstanding cleavages and rivalries. It lacks a core discipline, unlike the statistical analogy, where there has been a steady growth in the number and size of academic departments for the past few decades. One of its claimants to the core, Geography, has traditionally been a non-technical field, and in some areas of social geography there is a strong and fundamental antipathy to technological approaches.

In practice, commonality of interest is evident in the proliferation of GIS meetings, and we are beginning to see a supply of texts and journals. But the scientific track at GIS meetings is often small. People who attend GIS meetings need a constant supply of novelty, whether in scientific research or vendor products, and will soon desert if the supply dries up.

Options for the future

I hope I have shown that handling spatial information presents a range of intellectual and scientific challenges of much greater breadth than the phrase implies - in effect, a geographic information science. I see "geographic" as essential - much of what we are about concerns the geographical world and our relationships with it, and is much richer than "spatial". The change in meaning of the "S" word - from systems to science - seems to be going well, as evidenced by the success of this series of conferences, the move of the AutoCarto series to fully refereed papers, the new texts, subscriptions to the International Journal, and GIS paper submissions to such established journals as *Geographical Analysis*, *Computers and Geosciences*, *Computer Vision, Graphics and Image Processing*, and publications of the Regional Science Association and the IEEE.

I hope I have also shown that a strong scientific program serves not only

itself, but also the needs of industry and GIS users. GIS needs a strong scientific and intellectual component if it is to be any more than a commercial phenomenon, a short-lived flash in the technological pan. It is too easy to see current GIS as a hardware and software technology in search of applications, and to see the field of GIS as defined by the functional limits of its major vendor products. We need to move from system to science, to establish GIS as the intersection between a group of disciplines with common interests, supported by a toolbox of technology, and in turn supporting the technology through its basic research. But as currently perceived, GIS is about as close to a science as FORTRAN is to algebra.

In recent years we have seen a growing cleavage in GIS between two traditions, that of spatial information on the one hand and that of spatial analysis on the other. The spatial information tradition stresses large inventory databases, and gives geography the role of an access mechanism. The spatial analysis tradition stresses rich functionality and a range of data models, and gives geography a fundamental role in analysis and modeling. The cleavage is illustrated well by the relationship between two companies with the same parent - GeoVision, which markets "Geographic Information Systems" and TYDAC, which sells "Spatial Analysis Systems". The two traditions share common data structures and algorithms, and rely on the same sources of data and hardware. But this is not enough to convince the academy of the existence of a scientific field. To claim this we need to take a broader view, and to include data modeling, accuracy, cognition, reasoning, human/computer interfaces (HCI) and visualization, and to show how these are integral parts of both traditions.

Without such arguments, the GIS field will fragment, and the GIS storm will blow itself out. Associations as fundamentally disjoint as the AAG and AM/FM will find it impossible to justify joint sponsorship of conferences. Vendors will specialize in data input workstations, spatial analysis workstations or facility management systems with little potential for interaction or integration. This would be tragic.

How can we ensure a lasting future for GIS, both systems and science? Disciplines are like tribes, with their own totems, symbols and membership rules, languages and social networks. The GIS tribe is currently very cohesive - it is well funded, the field is exciting, and much useful research is being done. But in the longer term the field has not done well at behaving as a science, and the academy is still doubtful about whether it needs to be taken seriously. Science is hard, and places heavy obligations on its practitioners. We have been too busy, and technology has been moving too quickly. Too much of our literature is in conference proceedings, which bring fast exposure but only to limited audiences, and lack sufficient quality control. Few people have had the time to write the textbooks or to identify the intellectual core, or to publish the good examples.

I believe we ensure the future of GIS by thinking about science rather than systems, and by identifying the key scientific questions of the field and realizing their intellectual breadth. Geographic information systems are a tool for geographic information science, which will in turn lead to their eventual improvement. We need to speak to the academy, both directly and through key articles and texts on the philosophy, methodology and foundations of the field, and by placing GIS papers in strong journals. All three communities - users, vendors and researchers - have vital and symbiotic roles to play, and we will serve all three best by playing ours in the fullest possible sense.

Acknowledgment

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

References

Anselin, L., 1989. What is special about spatial data? Alternative perspectives on spatial data analysis. Technical Report 89-4, National Center for Geographic Information and Analysis, Santa Barbara, CA.

Besag, J., 1986. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society B* 48:259-302.

Tomlin, C.D., 1990. *Geographic Information Systems and Cartographic Modeling.* Prentice Hall.

Tomlinson, R.F., 1984. Keynote address: Geographical Information Systems - a new frontier. *Proceedings, International Symposium on Spatial Data Handling* (Zurich) Vol. 1 pp. 2-3.

Feature Recognition