NEW HORIZONS FOR THE SOCIAL SCIENCES: GEOGRAPHIC INFORMATION SYSTEMS

Michael F. Goodchild[1]

Introduction

Geographic information systems (GIS) are not exactly new to the social sciences—the UK's Economic and Social Research Council funded a network of GIS-based laboratories at UK universities in the late 1980s and early 1990s, and the U.S. National Science Foundation's Directorate for Social, Behavioral, and Economic Sciences funded the National Center for Geographic Information and Analysis to promote GIS-based research from 1988 to 1996 (both GIS organizations continue to exist). But there are nevertheless good reasons to include GIS in a discussion of the future of the social sciences. The use of GIS has now spread very widely among the sciences, and it is now an accepted tool among all of the disciplines that deal with the surface of the Earth and its human population. Moreover the concept of GIS has evolved substantially, and I propose in this paper to take a deliberately broad view of the term's meaning. GIS also claims to be an *integrating* technology, spanning disciplines and blurring the distinctions between them, both important prerequisites for any broadly useful research infrastructure. Finally, the use of GIS has prompted interest in a number of fundamental issues that are collectively identified as geographic information science.

The paper is organized as follows. The next section explores the nature and history of GIS, and the contemporary meaning of the term. It includes what I hope is an honest assessment of the technology's strengths and weaknesses. The third section includes a personal selection of the key concepts and principles of GIS. This is followed by a brief review of geographic information science. The final major section of the paper discusses the concept of Digital Earth, and its possible value as a motivating force. The paper closes with three final points.

The nature of GIS

Overview

It is very appropriate that this conference is being held in Ottawa, since the city has the strongest claim to be the home of GIS. In the early 1960s, the Federal and provincial governments had funded the Canada Land Inventory, a massive effort to assess the current and potential uses of Canadian land in a belt extending north from the U.S. border well beyond the areas of widespread settlement. The objectives of the project required detailed analysis, to determine the areas in use or available for such activities as forestry, agriculture, or recreation. Maps of different themes were to be overlaid to determine correlations and conflicts, but in manual form both area measurement and overlay are highly labor-intensive and crude operations. Roger Tomlinson was able to show that

[1] National Center for Geographic Information and Analysis, and Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA. Phone +1 805 893 8049. FAX +1 805 893 3146. Email good@ncgia.ucsb.edu

computerization was cost-effective, even in the environment of the mid 1960s, with its primitive and expensive computers and no tools for meeting the special requirements of handling map data. The Canada Geographic Information System (CGIS) was born out of a simple analysis of the relative costs of processing geographic data by hand and by computer.

The subsequent history of GIS has been described in detail (Foresman, 1998). Much important work in the late 1960s and 1970s was conducted at the Harvard Laboratory for Computer Graphics and Spatial Analysis, under the direction of Howard Fisher, William Warntz, and Brian Berry. Many important roots lie in landscape architecture, and in the computer-assisted design (CAD) systems developed at Cambridge and elsewhere. But the beginnings of the modern era of GIS, with its widely available commercial software products, dates to the early 1980s and the dramatic reductions in hardware costs that began then and continue today.

Today GIS is a major computer application with uses that range from the management of natural resources by government agencies and corporations, to the operations of utility companies, to support for scientific research and education. The software market is currently dominated by Environmental Systems Research Institute (ESRI), of Redlands, CA, and particularly by its ARC/INFO, ArcView, and SDE products. ESRI's annual sales are in the region of $250 million, its users number in the hundreds of thousands, and close to 10,000 people attended its most recent user conference in July 1999. Other significant GIS vendors include Autodesk, MapInfo, Smallworld, and Intergraph.

More broadly, GIS is part of a complex of geographic information technologies that includes remote sensing, the Global Positioning System, and geographic information services offered on the World Wide Web (WWW). In a 1993 study the U.S. Office of Management and Budget estimated that Federal expenditures on digital geographic information amounted to $4.5 billion annually, and figures of $10 to $20 billion for annual global expenditures seem reasonable. The term "GIS" is used increasingly to encompass all of these, and phrases such as "doing GIS", "GIS data", the "GIS community" suggest a willingness to see "GIS" as a shorthand for anything that is both digital and geographic in nature. Longley *et al.* (1999) provide a recent review of all aspects of GIS.

Definition

GIS is defined most generally as technology for processing a specific class of information—geographic information. *Processing* is understood to encompass creation, acquisition, storage, editing, transformation, analysis, visualization, sharing, and any other functions amenable to execution in a digital domain. *Geographic information* is readily defined as information linking locations on the Earth's surface with specific properties, such as name, feature, population, elevation, temperature. More generally and precisely, it consists of atoms of information or *tuples* of the form <location, time, property>. To be communicable, a scientist would argue that all three components must be well-defined, using terms that are known to both sender and receiver of information. In the case of location, this argument clearly favors general standards such as latitude and longitude over more problematic specifications of location such as place-names. But there are strong arguments for including information in GIS that is poorly defined, vague, or subjective, because of the importance of these forms to human communication, and there has been much interest recently within the research community in the problems of handling vague geographic information.

This definition of geographic information is deceptively simple. Unfortunately the geographic world is continuous and infinitely complex, and there are therefore an infinite number of locations in space and time to be described. In practice, geographic information must somehow approximate, generalize, or simplify the world so that it can be described in a finite number of tuples. There are an unlimited number of ways of doing this, in other words an unlimited number of ways of mapping the real geographic world into the contents of a GIS database, which requires an alphabet comprised of only two symbols, 0 and 1. Many such mappings or *representations* have been implemented in various disciplines and areas of application, and many are implemented in the standard GIS software products as *data models*.

Data models fall into two broad but imperfectly defined categories—*raster* and *vector*. In a raster representation, the world is divided into an array of cells of fixed size (note that some distortion is implied, since the curved surface of the Earth cannot be covered by uniformly sized, non-overlapping square cells). All properties of the surface are expressed as uniform properties of the cells, and all sub-cell information is lost. Moreover, rasters are not convenient for capturing geometric structures larger than the cell, since it is generally difficult to link cells together. In a vector representation, properties are associated with geometric points, lines, or areas, and the locations and shapes of these objects are defined by coordinates. Areas are approximated as polygonal figures by connecting points with straight lines, and curved lines are similarly approximated as *polylines*. Vector representations readily accommodate variable spatial resolution, links between objects, and complex geometric structures, and are strongly favored in applications of GIS to social phenomena.

Large and comprehensive software environments such as GIS are possible because of strong economies of scale in software production. Once a basic framework has been built, by implementing a limited number of basic data models with associated tools for creating, editing, visualizing, and sharing data, additional functions can be added very easily and cheaply. This principle is clearly evident in spreadsheets, word processors, statistical packages, and GIS, all of which are defined by basic data models.

But herein lies one of the fundamental weaknesses of the GIS idea—there are simply too many possible geographic data models. In order to accommodate the needs of new applications, software vendors have repeatedly extended the basic data models of their products. One of the most persistent problems is associated with time, since early GIS were built largely to accommodate the static data of maps, and their data models have been extended with varying success to deal with temporal change (Langran, 1992). The problems of dealing with data on networks, necessary in many transportation applications, has led to the emergence of products specifically targeted to this niche. Today, a vendor such as ESRI offers a suite of products rather than a single, comprehensive GIS. Each product is designed for a particular class of applications, or for a user community with a particular level of sophistication. The products are able to share data, and many of the concepts on which they are based are common. But with the present trend towards unbundling of software in favor of modular code for specific applications, it seems likely that the days of monolithic GIS are numbered. Instead, we are likely to see much smaller software components that can be mixed to service particular applications, held together through common specifications and standards. Efforts are under way through the Open GIS Consortium (*www.opengis.org*) to standardize across the entire vendor community, but whether this will be successful, or whether standardization will be achieved only across the products of each vendor, remains to be seen.

Key concepts

It is not at all clear that a technology designed to process geographic information is of significant value to the social sciences, or that it has potential as research infrastructure. In this section I examine six key concepts, and briefly explore their value to the research enterprise.

Integration

One of the commonest ways of introducing GIS—the basis of the cover design on many textbooks—is the *layer-cake*, a representation of the way a GIS database integrates many properties, variables, and themes through common geographic location. To Tomlinson and CGIS, this was the computational equivalent of overlaying a number of maps portraying different uses of the same area. If a GIS database links properties to locations, it can clearly also link properties to properties through common location. In the literature of GIS, the technology is often presented as the *only* basis for integrating the departmentalized operations of an organization, and the only way of achieving an integrated perspective. For example, the U.S. Geological Survey is organized according to four distinct themes: geology, water, biology, and topography. Yet for many of its users, this organizational structure impedes rather than facilitates, since it makes it difficult to determine all that the USGS knows about a specific place. U.S. socioeconomic data are similarly partitioned among different surveys, agencies, and production systems.

While this argument is most obviously made about data, it can also be made about process. Demographic and economic processes, for example, interact at common locations. By creating representations that are spatially and temporally explicit, GIS databases permit coupled and integrated modeling of multiple processes that would otherwise be studied only separately and within the domains of different disciplines.

Spatial analysis

Of necessity, much socioeconomic information is collected in *cross-section*, and the construction of *longitudinal* series is beset by problems of continuity, budgets, and changing technology. Spatial analysis, or spatial data analysis, comprises a set of techniques and tools designed to analyze data in spatial context. A GIS database captures not only links between properties at the same place, but also such spatial concepts as proximity, containment, overlap, adjacency, and connectedness. Visualization in spatial context (commmonly, in the form of a map) is an obvious and powerful way of detecting pattern, anomaly, outlier, and even causation. Of course the forms found in cross-sectional data can never confirm cause, since the same forms can always be created by many processes. Nevertheless spatial data can be powerful sources of new insights and hypotheses, and powerful bases for confirmatory tests.

Spatial analysis is often best portrayed as a collaboration between mind and machine, combining the power of the eye and brain to detect pattern and scan complex visual displays quickly, with the machine's power to compare layers, apply statistical tests, and perform transformations. Spatial analysis has undergone substantial change in recent years, as more and more of the power of the desktop computer has been allocated to achieving ease of use through user interfaces that are visual and intuitive. In the early days of computing the machine was expert, and the user *submitted* tasks to it. But today's designs make it much easier to support collaboration, and the concept of spatial analysis has broadened accordingly, to encompass everything from visual examination and exploration of mapped data to complex confirmation of spatial statistical models.

Spatially explicit theory and modeling

A model or theory is *spatially explicit* if it is not invariant under relocation; in other words, if changing the locations of the objects that participate in the theory changes the theory's predictions. For example, spatial interaction models are widely used to predict choices made by consumers among shopping destinations (and a variety of other forms of interaction over space as well, including telephone traffic, migration, and commuting). Distance appears explicitly in the model, and transformation of space, for example by construction of a new transportation link, changes the model's predictions.

Whether space can ever *explain*, or whether it must always be a surrogate for something else (in the case of the spatial interaction model, for the disincentives of travel time or transport cost, for example) is a moot point. Environmental determinism, or the hypothesis that location determines aspects of human behavior, is now largely discredited within the discipline of geography. In other disciplines, notably economics and ecology, there is much current interest in explicit theorizing about space—in its simplest form, through the partitioning of populations into subpopulations whose spatial separation is modeled as a source of imperfect communication. Space is also explicit in many forms of microsimulation, in which intelligent agents representing individual actors are allowed to move and interact according to well-defined rules.

Place-based analysis

The previous argument is taken a little further in the current interest in *place-based* or *local* techniques of analysis. Earlier debates in geography, notably in the 1950s, had pitted the champions of a *nomothetic* approach, whose aim was the discovery of principles that applied uniformly everywhere (*general* geography, in the sense of Varenius) against the champions of *idiographic* geography, whose focus was the description of the unique properties of places (the *special* geography of Varenius). Nomothetic geography was held to be scientific while idiographic geography was *merely* descriptive.

In the past decade something of a middle ground has emerged between these two positions, aided and abetted by GIS. In this new approach the parameters of models are allowed to vary spatially, and their variation is interpreted and used as the basis for insight and further analysis. For example, suppose that some model $p=f(z)$ is hypothesized to apply to human societies. Given the extreme variability of humanity, it seems unreasonable to believe in a single confirmation of the model conducted in a single city—but on the other hand, an experimental design that samples all of humanity's variability is clearly impossible. Instead, place-based analysis focuses on how the parameters of the model vary from place to place, and draws insights and conclusions from those variations. It thus deals explicitly with the problem of *spatial heterogeneity*, or the notion that no geographic area, however large, can be representative of humanity or the Earth's surface unless it encompasses the entire Earth—geography has unlimited variance up to the scale of the Earth.

The set of techniques designed to support place-based analysis includes adaptive spatial filtering, geographic brushing, geographically weighted regression, and local statistics. Fotheringham (1997) provides an excellent recent review.

Knowledge and policy

GIS is widely used both inside the academic research community and also in government agencies, corporations, and NGOs. Its applications thus span the distinction between pure and applied, or curiosity-based and problem-driven research, and it provides a clear bridge between them, echoing the arguments of Laudan (1996) that no effective demarcation exists today between science and problem-solving. General knowledge of the ways human societies operate must be combined with data on local conditions to make effective policy, and this is perfectly captured in the ability of a GIS to combine local detail (the contents of the database) with general principles (the algorithms, procedures, and data models). GIS is used to simulate the operations of processes under local conditions, and to examine the impacts of general principles in explicit spatial context.

One of the compelling attractions of GIS to a government regulatory agency appears to lie in its procedural nature, which to a scientist might seem overly simplistic and naïve. For example, it is easy to write into law that no industry should locate within 1km of a residential area, and easy to implement regulation based on GIS analysis, by computing 1km buffer zones around industries or around residential areas. Scientifically this is naïve, since we have no reason to assume that the effects of industrial pollution are the same upwind as downwind, for example; but the simple procedure stands up well to court challenge, since it can be applied uniformly and diligently. It raises the interesting question of whether effective policy can ever be based on good but complex and often equivocal science.

Place-based search

Narrowly defined, geographic information provides a representation of spatial variation of phenomena over the Earth's surface. It includes maps and images, which provide *exhaustive* representations of the entire surface within their limits, and sampled data that provide information only about a selection of places. Recently, however, there has been much interest in a third class of information that is not strictly geographic but nevertheless can be geographically referenced—that has some form of geographic *footprint*. For example, a tourist guide to Paris has such a footprint, and while it may contain maps it also contains many other forms of information, some strictly geographic and some not.

Interest in such geographically referenced information arises because of the potential for using geographic location as a basis for search over large and possibly distributed collections of information. For example, geographic location is one of two primary organizational keys for the Electronic Cultural Atlas Initiative (*www.ecai.org*), an international effort to make primary data in the humanities accessible over the WWW (the other key is time). The University of Southern California is building a major digital archive of its collection of historic photographs of Los Angeles, using the same principle.

A *geolibrary* is defined as a library whose primary search mechanism is geographic. Location has not fared well as a basis of search and organization in the traditional library for largely technical reasons, but there are no technical reasons that prevent a digital library being organized to respond to the query "what have you got about *there*?". A recent report of the U.S. National Research Council (NRC, 1999) elaborates on the concept and describes many current prototype implementations.

Geographic information science

GIS has developed as complicated and sophisticated technology for support of science and policy-making, but it has done so largely in the absence of a coherent body of theory or language. In this it stands in sharp contrast to the statistical packages, which developed to support an existing and widely used set of techniques underpinned by well-defined theory. If the statistical packages are implementations of statistical theory, then where is the theory that GIS implements?

One consequence of this lack of pre-existing theory is the diversity of languages and standards that have emerged from a largely uncoordinated GIS software industry. GIS products appear to their users as highly intuitive and pragmatic, rather than as implementations of some universally accepted set of principles, which perhaps explains their popularity. But it means that the GIS community is deeply divided into distinct *information communities*, each with their own set of norms, standards, and terms. There are very high costs associated with moving data from one product to another, or retraining staff.

Geographic information science seeks to develop the science behind the systems, and to address the fundamental issues raised by GIS. Its focus is well described by the research agenda of the University Consortium for Geographic Information Science, an organization of major U.S. research universities that now includes some 60 members (*www.ucgis.org*). The agenda was developed by consensus at the Annual Assembly of UCGIS in Columbus in 1996 (UCGIS, 1996), and contains ten topics:

- *Extensions to representations*, or research to elaborate the set of data models that form the basis of GIS, notably to include time, the third spatial dimension, and level of detail.
- *Scale*, or research on the characterization of level of detail, transformations that aggregate or disaggregate, and the role of scale in modeling process.
- *Uncertainty*, or research on the characterization of data quality, its impacts on the results of modeling and analysis, and its visualization and communication.

- *Cognition,* or research on the ways humans understand, reason about, and work with geographic information.
- *Spatial analysis,* and the development of new techniques and tools for analysis of spatial data.
- *Distributed and mobile computing,* and the opportunities offered by new technology for new uses of GIS in the field and distributed over electronic networks.
- *Interoperability,* or research on the problems caused by lack of standard protocols and specifications, and the development of new theory-based terminology.
- *Acquisition and integration*, or research on new sources of geographic information, and their integration with existing sources.
- *Spatial information infrastructure*, or policy-oriented research on the production, dissemination, and use of geographic information.
- *GIS and society,* or research on the impacts of GIS on society, and the context provided by society for GIS.

Digital Earth

In a speech written for presentation at the opening of the California Science Center in Los Angeles in January 1998, U.S. Vice President Al Gore proposed "a multi-resolution, three-dimensional representation of the planet, into which we can embed vast quantities of georeferenced data" (*www.opengis.org/info/pubaffairs/ALGORE.htm*). In the speech, Digital Earth is an immersive environment through which a user, particularly a child, could explore the planet, its environment, and its human societies. It might be available at museums or libraries, and a more modest version might be available through standard WWW browsers running on a simple personal computer.

Digital Earth is interesting for several reasons, and the concept has attracted widespread interest (the first International Symposium on Digital Earth will be held in Beijing in December 1999). First, it has some of the properties of a *moonshot*, or a vision that can motivate a wide range of research and development activities, in many disciplines. It challenges our state of knowledge about the planet, not only in terms of raw data, but also in terms of data access, and the ability to communicate data through visualization. How, for example, would one portray state of human health or quality of life to a child? Moreover, it challenges our understanding of process in the invitation to model, simulate, and predict, since the concept should not be limited to static portrayal.

Second, Digital Earth is interesting because of its implications for the organization of information. The prevailing metaphor of user interface design is the office or desktop, with its filing cabinets and clipboards. Many prototype digital libraries employ the library metaphor, with its stacks and card catalogs. But Digital Earth suggests a much more powerful and compelling metaphor for the organization of geographic information, by portraying its existence on a rendering of the surface of the Earth. The idea can be seen in limited form in many current products and services, including Microsoft's Encarta Atlas.

Finally, Digital Earth is a fascinating example of a *mirror world* (Gelernter, 1991). Just as a map, it captures a particular state of understanding of the planet's surface, and the data and infromation available to its builders. But since it cannot be a complete representation, it is interesting in what it leaves out, and in how it reveals the agendas of its builders.

Closing comments

I would like to make three brief points in conclusion.

First, GIS seen narrowly is an important and growing application of computing technology. It includes software, today largely developed and marketed by the private sector; data, increasingly available in large quantities through the medium of the WWW; and tools for analysis and modeling that focus on the spatial aspects of data, and increasingly on the temporal aspects. As such, GIS is of increasing importance to those social sciences that deal in one way or another with activities and phenomena that distribute themselves over the surface of the Earth, and with understanding the processes that lie behind them.

Second, GIS seen broadly raises a number of challenging and fundamental issues that range from human spatial cognition to the modeling of complex spatial processes. Collectively, they motivate a multidisciplinary effort to advance what can be termed geographic information science, and many of these issues intersect and engage the social sciences.

Finally, GIS seen broadly is intimately related to the concept of Digital Earth, or the development of an accessible, unified emulation of the surface of the planet and the processes that affect it, both human and physical. As a vision, Digital Earth may or may not be achievable, depending on the assumptions one is willing to make about future technologies, the availability of information, and our ability to characterize and understand process. But as a moonshot it is an idea that can motivate a broad spectrum of activities across many disciplines.

References

Foresman, T.W., editor (1998) *The History of GIS: Perspectives from the Pioneers*. Upper Saddle River, NJ: Prentice Hall PTR.

Fotheringham, A.S. (1997) Trends in quantitative methods. 1. Stressing the local. *Progress in Human Geography* 21(1): 88–96.

Gelernter, D. (1992) *Mirror Worlds: Or the Day Software Puts the University in a Shoebox. How it will happen and what it will mean.* New York: Oxford University Press.

Langran, G. (1992) *Time in Geographic Information Systems*. London: Taylor and Francis.

Laudan, L. (1996) *Beyond Positivism and Relativism: Theory, Method, and Evidence*. Boulder, CO: Westview Press.

Longley, P.A., M.F. Goodchild, D.J. Maguire, and D.W. Rhind, editors (1999) *Geographical Information Systems: Principles, Techniques, Management and Applications*. New York: Wiley.

National Research Council (1999) *Distributed Geolibraries: Spatial Information Resources*. Washington, DC: National Academies Press (also at *www.nap.edu*).

University Consortium for Geographic Information Science (1996) Research priorities for geographic information science. *Cartography and Geographic Information Science* 23(3): 115–127.