

GIScience, Geography, Form, and Process

Michael F. Goodchild

National Center for Geographic Information and Analysis and Department of Geography, University of California, Santa Barbara

The map has evolved over the past few centuries as humanity's primary method for storing and communicating knowledge of the Earth's surface. Topographic maps portray the general form of the surface and its primary physical and cultural features; thematic maps show the variation of more specialized properties such as soil type or population density; and bathymetric maps and hydrographic charts show the characteristics of the sea floor. Maps serve as one of the most important repositories of both the raw data and the results of geographic inquiry, and mapmaking has always figured prominently in the skill set of geographers or their supporting staff. Maps are thus important and indispensable tools in the geographer's search for understanding of how human and physical processes act and interact on the Earth's surface: of how the world *works*.

Geographic information systems (GIS) were devised in the 1960s as computer applications for handling large volumes of information obtained from maps and for performing operations that would be too tedious, expensive, or inaccurate to perform by hand. The Canada Geographic Information System, widely recognized as the first GIS, was built for the purpose of making vast numbers of calculations of area, reporting the results in tables. Over time, the range of functions performed by GIS has grown exponentially, and today it is reasonable to think of a GIS as able to perform virtually any conceivable operation on data obtained from maps (Longley et al. 2001). Geographers have adopted GIS enthusiastically, seeing it as a powerful device for storing, analyzing, and visualizing map information and thus as a much more effective substitute for the paper map (Goodchild 1988).

Over the past decade numerous journals, conferences, academic positions, and programs have adopted titles that combine *information* with *spatial* or *geographic* and with *science* or *theory*. In what follows I will use the term *geographic information science* (GIScience) for simplicity and not enquire into the subtle differences between, for example, spatial and geographic information theory (Goodchild 2001). Geographers have been associated with many of these changes—and, in many cases, have

been at the forefront—and many of the new programs and positions are found in departments of geography. But there has been relatively little general commentary on these trends, or on what they might mean for the discipline of geography as a whole. The first centennial of the Association of American Geographers is an appropriate occasion to reflect on the nature of GIScience and its relationship, if any, to the discipline of geography.

I begin with a discussion of the nature of GIScience, of its relationship to GIS and of its links to the traditional sciences of geographic information. This leads to a discussion of whether GIScience is a *natural* science, concerned with discovering empirical principles and law-like statements about the world; or whether it is a *design* science, concerned with identifying practical principles for achieving human ends, or both. In the third major section I examine how GIScience is positioned with respect to the historic tension in geography between form and process and whether the growth of interest in GIScience has tended to favor form over process. The final section examines a future for GIScience that places greater emphasis on process and discusses the steps that will be needed to make such a future possible.

What Is GIScience?

The University Consortium for Geographic Information Science (UCGIS), a collaboration between approximately seventy academic institutions, private companies, and government agencies and one of the more prominent manifestations of the rise of GIScience in the U.S., is “dedicated to advancing our understanding of geographic processes and spatial relationships through improved theory, methods, technology, and data” (<http://www.ucgis.org>). This theme of tools in the service of science is echoed by Clarke (1997, xi), who defines GIScience as “the discipline that uses geographic information systems as tools to understand the world.” The characteristics of GIScience are thus no more and no less than those one expects of any scientific enterprise: replicability, independence of the observer and the

observed, a shared lexicon of well-defined terms, and a concern for accuracy. One expects results obtained from the use of GIS to be reported with a level of precision that reflects their accuracy and to a level of detail sufficient to allow them to be replicated by others; and one expects GIS procedures to be carefully and fully documented.

But this is only one of two competing definitions of GIScience. Goodchild (1992, 32) defined GIScience as “the science behind the systems,” concerned with the set of fundamental questions raised by GIS and allied technologies, and Mark (2003) has provided a lengthy commentary on definitions. In this interpretation, GIScience is the storehouse of knowledge that is implemented in GIS and that makes GIS possible. It may search for general principles, such as the enumeration of the possible topological relationships between pairs of features by Egenhofer and Franzosa (1991), one of the most-cited papers in GIScience (Fisher 2001). It may discover faster algorithms, or more efficient indexing schemes, or new ways of visualizing geographic information. UCGIS has identified ten “research challenges” (<http://www.ucgis.org>), representing a consensus on the most important long-term components of the GIScience research agenda.

In this second sense of the science behind the systems, GIScience builds on the accumulated results of many centuries of investigation into how to describe, measure, and represent the Earth’s surface. The shift to digital technology has revolutionized the older GISciences of surveying, photogrammetry, and cartography, giving new motivation to older research questions and raising new questions related to the greater flexibility and power of digital technologies. Moreover, the older GISciences evolved in an era of distinct, analog technologies—as long as the paper and pen of cartography had little in common with the analytical stereoplotter of photogrammetry or the theodolite of surveying, there was every reason for them to evolve separately, with separate research agendas. But today all three fields have embraced digital technology wholeheartedly. They serve overlapping applications and face similar issues of representation, database design, accuracy, and visualization.

The world of geographic information has also grown more complex, as new questions have arisen that require the skills and principles of other sciences. Remote sensing, the science of Earth observation, is now an important source of geographic information with its own issues and principles. The unique problems of spatial information have begun to intrigue computer scientists, and spatial databases, computational geometry, and

spatial indexing are now recognized subfields of computer science with special significance for GIScience (Worboys 1995; Shekhar and Chawla 2003). Spatial statistics and geostatistics, recognized subfields of statistics, provide important frameworks for the study of accuracy and uncertainty in GIScience (Zhang and Goodchild 2002) and for the development of advanced methods of spatial analysis, modeling, and visualization (Haining 2003; Longley and Batty 2003; O’Sullivan and Unwin 2003). GIScience is a legitimate subfield of information science and particularly attractive to information scientists because of the well-defined nature of geographic information and the comparatively advanced state of knowledge about this information type. Finally, an important section of the GIScience research agenda asks questions of interest to cognitive scientists: how are geographic knowledge and skills acquired by the human brain and how can GIS be made more readily understood and usable by humans?

Is GIScience Experimental?

The Egenhofer and Franzosa 9-intersection (Egenhofer and Franzosa 1991) is a purely theoretical deduction, obtained as results in mathematics are obtained by argument from first principles, rather than inductively by generalization from empirical observation. Other disciplines such as physics or geography build knowledge through a combination of deduction and induction, generalizing from observation to make law-like statements and deducing or hypothesizing principles that can be tested against observation. In this spirit, one might ask whether law-like statements are possible about the subject matter of GIScience: does geographic information display properties about which one can generalize? The practical value of such properties would be enormous since they could guide the design of GIS, leading to efficient choices between representation methods, indexing schemes, and algorithms and to expectations about the volume of information lost due to generalization, for example.

Anselin (1989) has argued in the context of spatial statistics that geographic data display two general properties, both of which must be addressed in any analysis of spatial data: spatial dependence and spatial heterogeneity. The first is the property inherent in Tobler’s First Law of Geography (Tobler 1970, 236; Sui 2004): “All things are related, but nearby things are more related than distant things.” It is no accident that interest in Tobler’s principle has grown rapidly in recent years with the growth of GIS and GIScience, because it is exploited in numerous ways in the design of GIS. All GIS

representations, whether raster or vector, depend for their effectiveness on the comparatively slow variation of properties over the Earth's surface, which makes it unnecessary to record properties uniquely at every distinct point, an impossible task, given the infinite number of such points. Tobler's law is also exploited by all established methods of spatial interpolation and spatial resampling. Every weather map, for example, is prepared from a limited number of sample measurements at points; its contours are generated by following the principle that closer sample measurements provide better estimates of missing values than distant sample measurements.

Anselin's second principle of spatial heterogeneity argues that expectations vary across the Earth's surface, with the important consequence that the results of any analysis depend explicitly on the bounds of the analysis. It accounts for the growth of interest in recent years in various forms of *place-based* analysis that allow results to vary spatially, rather than searching for a single universal result. One might see this interest as representing a middle position in the old debate in geography between nomothetic and idiographic science, rejecting the position that all places are unique in favor of general principles whose parameters vary from one place to another. Fotheringham's Geographically Weighted Regression (Fotheringham, Brunsdon, and Charlton 2002) examines how the parameters of a regression model vary geographically, while Anselin's Local Indicators of Spatial Association (Anselin 1995) examine spatial variation in degrees of clustering. One might argue that the spatial heterogeneity principle should rank as the first law and Tobler's as the second, because heterogeneity addresses the properties of places taken one at a time (a *first-order* effect in the statistical sense), whereas spatial dependence compares the properties of pairs of places (a *second-order* effect).

Tobler's First Law and Anselin's concept of spatial heterogeneity are useful, general properties of geographic information and lead immediately to two important questions, neither of which has been extensively researched: first, to what extent do they apply to all spaces, not only geographic space; and second, are there other law-like statements with similar empirical support and similar utility to GIS? I have recently suggested that as many as seven such statements can be identified (Goodchild 2003), including a fractal principle (all geographic phenomena reveal more detail with finer spatial resolution, at predictable rates) and an uncertainty principle (it is impossible to measure location or to describe geographic phenomena exactly) while Montello et al. (2003, 317) have proposed a First Law of Cognitive Geography, "People think that closer things are more

similar," that has utility in the design of such human-centered GIS functions as visualization.

If there can be law-like statements about spatial information, then presumably, it is possible to discover similar properties in spatiotemporal information. Many methods of analysis of spatiotemporal data do indeed employ simple extensions of Tobler's First Law: one need only to generalize "nearby" and "distant" to comparable metrics in space-time. Diffusion processes ensure, for example, that what happens in location x at time t will be related to events in location $x+d$, time $t+e$, where d and e are displacements in space and time that are suitably matched to the rate of diffusion. Law-like statements seem feasible with respect to the space-time behavior of organisms, and, of course, such statements are implicit in models of many natural phenomena, such as weather or topography. The general systems theorists of the 1960s attempted to find very general law-like statements about dynamics, though to date there has been no effort to revive that work in support of temporal GIS.

In summary, it appears that law-like statements about the properties of geographic information are possible and that such statements can be of great value both in justifying decisions made in the past regarding GIS design and in guiding future decisions. Armed with statements about the general properties of geographic information, it is possible to generate data sets that exhibit such properties and to use them as testbeds for new algorithms, data structures, and indexing schemes. It is possible to devise new methods of generalization that respect fractal properties and to make estimates of data volumes following generalization.

Form and Process

A paper map is, of necessity, static, reflecting the state of knowledge at the time it was compiled and printed. The economies of scale of map production led inevitably to an emphasis on the mapping of relatively static aspects of the Earth's surface, such as topography, over relatively dynamic aspects. While paper maps can be annotated individually, the digital environment clearly has massive advantages in the ease with which data can be edited, updated, and redistributed. A GIS database can be used to store frequent changes, or *transactions*, such as those that occur as street networks become more or less congested through the day. Today, it is easy to download such information about congestion from Web sites in the form of dynamic maps whose validity lasts only a few minutes and to obtain similar dynamic maps of recent earthquakes or weather conditions. Increasingly, such services are becoming available through

personal digital assistants and cellphones, despite the limited display area of such devices.

But while such capabilities are becoming commonplace, the maps they generate are still snapshots of the two-dimensional world at a particular point in time: they represent the world as it *looks*, even fleetingly. GIS data models are similarly concerned with representing form, by recording the precise locations of point, line, or area features. There has been some success in extending such models to include the third spatial dimension, to represent the form of geologic, atmospheric, and oceanographic features. There is also a lengthy and rich literature in GIScience on efforts to extend GIS data models to include time and the representation of dynamic phenomena (Langran 1992; Peuquet 2002). Recently, interest has grown in the possibilities of tracking data, or records of the movements of individuals and vehicles in space and time, driven in part by the growing availability of such data as a result of GPS. Kwan (2000) has explored methods of visualization for tracking data, and Miller (2003) has devised interesting new constructs and methods of analysis that extend the early work of Hägerstrand (1970).

But whether the phenomena are static or dynamic, these efforts remain focused largely on form. In principle, the study of ontology in GIScience (Winter 2001) includes "the totality of geospatial concepts, categories, relations, and processes" (Mark et al. 2000, 1), but in practice the dominant emphasis in such research is on the objects that form the basis of geographic description and representation, rather than on the processes that are the primary goal of the geographic research enterprise. If studies of ontology are dominated by form, then perhaps there is a need for a parallel research focus on epistemology, with emphasis on process.

Other aspects of the GIScience research agenda are similarly focused on form. The uncertainty problem concerns the degree to which the contents of a database leave the user uncertain about the corresponding contents of the real world, with respect to geometric form, attributes, and topological relationships (Zhang and Goodchild 2002). Spatial analysis and data mining are concerned with discovering patterns, clusters, and trends that may not otherwise be apparent to the user. The need to make easy translations between place names and coordinates has led to increased interest in digital gazetteers and to the processes of naming places, in a revival of the old and largely discredited field of toponymy. Mark and Turk (2003) have recently proposed a new research field of *ethnophysiology* to address the naming of geographic features among different cultures, using the methods of ethnography. All of these trends make good

sense in the context of GIScience, with its concern for the science behind GIS, but their contribution to the ultimate understanding of process is less obvious.

Toward an Emphasis on Process

I have argued that the growth of GIScience has led to an increased interest in form, leaving inference about process entirely outside the system. In this section I suggest ways in which this balance could be reversed in order to provide more effective support in GIS for studies aimed at understanding process.

First, and perhaps most obviously, more rapid progress is needed in representing time in GIS and in the development of methods for the analysis of spatiotemporal data. Process is much easier to infer from longitudinal data, with its representation of the sequence of events, than from cross-sectional data. Valiant efforts have been made, of course, to make the maximum possible use of cross-sectional data when no other data are available, and cross-sectional data can be used to falsify hypotheses about process, even if they cannot often be used to confirm them (Goodchild et al. 2000). But GIS remains poorly equipped to handle dynamic data for a variety of reasons, not all of which are within the control of GIScientists. The persistent use of the map metaphor to conceptualize GIS leads to a focus on static data. Longitudinal series are often difficult to construct, particularly when reporting zones change frequently through time (Frank, Raper, and Cheylan 2001) and when the definitions of variables also change, as they do in the decennial census. Remote sensing snapshots are an abundant and comparatively cheap source of data for GIS and have their own problems of change detection. Most problematic, perhaps, is the difficulty of retrofitting extensions to data models, which require modification of the foundations of software packages, running counter to their basic scale economies. Progress is being made, however, particularly as a result of the widespread adoption of object-oriented data modeling.

Second, there will have to be a much closer coupling between hypotheses about process and the methods of analysis and visualization implemented in GIS. Consider the case of the Modifiable Areal Unit Problem (MAUP; Openshaw 1983). GIS has made it much easier to manipulate the zones used by agencies such as the census to aggregate data, and in a seminal paper Openshaw and Taylor (1979) were able to document the striking effects of manipulating zones, in their case the counties of Iowa, on the results of a simple analysis of the correlation between age and voting behavior. Missing, of course, is any hypothesis concerning process and the relevance of

county boundaries, or any other reporting zone boundaries, to that process. Even if such a hypothesis did exist (for example, that the voting behavior of an individual is related to the individual's age), any test of the hypothesis would be obscured by the use of inappropriately aggregated data. GIS makes it easy to conduct sophisticated and complex analyses without ensuring that these are linked to appropriately formulated hypotheses about process.

In cases like this, methods of analysis are applied to data, but it is left entirely to the researcher to formulate a hypothesis and to understand what the analysis would have revealed had the hypothesis been true. This mental process can be very complex, especially when artifacts such as reporting zone boundaries confuse the outcome. Agent-based modeling and similar methods of massive simulation offer one way out of this dilemma, by making it possible for the researcher to conduct two analyses in parallel: one of the real data and one of a simulated world in which the hypothesis is true and which is identical in all other controllable respects. To make this possible, it would be necessary for GIS to support various forms of simulation.

Finally, it will be necessary to recognize the importance of digital representations of process, or what might be termed *process objects*. These are programs that simulate the actions of real physical and social processes, and, like data, they are digital, but differ from data in being executable rather than static. Unfortunately, such programs exist in many forms, with almost no standards. Some are stand-alone, written in source programming language and executable in standard operating systems. Others consist of scripts written in the specialized language of some simulation environment—a notable example in a geographic context is PCRaster, developed at the University of Utrecht and implementing a language devised by van Deursen (1995). It is also possible to write sequences of commands to standard GIS packages, using the scripting language of the GIS, but performance tends to be poor because such packages are not typically designed for simulation.

In essence, process objects formalize knowledge of process, allowing it to become part of the digital environment and to benefit from the digital environment's capabilities for easy editing, rapid dissemination, reliable preservation, error correction in transmission, and the scale economies of a uniform technology with universal standards. Process objects stand today in much the same situation that faced data objects in the very early days of GIS. There are no format standards, no archives, no widely practiced methods of sharing, and no standards for description (but see Crosier et al. 2003). Instead,

information about processes is confined largely to the text of journal articles and books, shared through the hit-and-miss process of reprogramming. When it is found in digital form, it is as coded text embedded in word-processing documents rather than as executable code. Yet, process objects represent a highly abstracted form of advanced knowledge, with much higher value per bit than raw data.

In summary, I have argued that GIS and GIScience have quite logically led to a renewed interest in form, but that this has moved the field away from the core disciplinary concern with process. I have suggested three ways in which GIS might evolve to provide better support for inference about process and three related items for the GIScience research agenda: better representation of dynamics, and associated improvements in the supply of data and relevant methods of analysis and visualization; a closer coupling between analysis and the conceptualization of process, facilitated by integrated methods of massive simulation; and the development of an infrastructure for sharing digital representations of process. None of these seems particularly difficult, but, taken together, they should ensure that the relationship between GIScience and geography remains strong and vital in the coming decades.

References

- Anselin, L. 1989. *What is special about spatial data? Alternative perspectives on spatial data analysis*, Technical Report 89-4. Santa Barbara, CA: National Center for Geographic Information and Analysis.
- . 1995. Local indicators of spatial association—LISA. *Geographical Analysis* 27:93–115.
- Clarke, K. C. 1997. *Getting started with geographic information systems*. Upper Saddle River, NJ: Prentice-Hall.
- Crosier, S. J., M. F. Goodchild, L. L. Hill, and T. R. Smith. 2003. Developing an infrastructure for sharing environmental models. *Environment and Planning B: Planning and Design* 30:487–501.
- Egenhofer, M. J., and R. D. Franzosa. 1991. Point-set topological spatial relations. *International Journal of Geographical Information Systems* 5:161–74.
- Fisher, P. F. 2001. Editorial: Citations to the *International Journal of Geographical Information Science*: The first ten years. *International Journal of Geographical Information Science* 15 (1): 1–6.
- Fotheringham, A. S., C. Brunsdon, and M. Charlton. 2002. *Geographically weighted regression: The analysis of spatially varying relationships*. New York: Wiley.
- Frank, A. U., J. Raper, and J.-P. Cheylan. 2001. *Life and motion of socio-economic units*. New York: Taylor and Francis.
- Goodchild, M. F. 1988. Stepping over the line: Technological constraints and the new cartography. *American Cartographer* 15:311–9.
- . 1992. Geographical information science. *International Journal of Geographical Information Systems* 6 (1): 31–45.

- . 2001. A geographer looks at spatial information theory. In *Spatial information theory: Foundations for geographic information science: Proceedings of the International Conference COSIT 2001, Morro Bay, CA, September*, ed. D. R. Montello, Lecture Notes in Computer Science 2205, 1–13. Berlin: Springer.
- . 2003. The fundamental laws of GIScience. Paper presented at the Summer Assembly of the University Consortium for Geographic Information Science, Pacific Grove, CA, June. http://www.csiss.org/aboutus/presentations/files/goodchild_ucgis_jun03.pdf (last accessed 26 July 2004).
- Goodchild, M. F., L. Anselin, R. P. Appelbaum, and B. H. Harthorn. 2000. Toward spatially integrated social science. *International Regional Science Review* 23 (2): 139–59.
- Hägerstrand, T. 1970. What about people in regional science? *Papers of the Regional Science Association* 24:7–21.
- Haining, R. P. 2003. *Spatial data analysis: Theory and practice*. Cambridge, U.K.: Cambridge University Press.
- Kwan, M. -P. 2000. Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: A methodological exploration with a large data set. *Transportation Research C* 8:185–203.
- Langran, G. 1992. *Time in geographic information systems*. New York: Taylor and Francis.
- Longley, P. A., and M. Batty, eds. 2003. *Advanced spatial analysis: The CASA Book of GIS*. Redlands, CA: ESRI Press.
- Longley, P. A., M. F. Goodchild, D. J. Maguire, and D. W. Rhind. 2001. *Geographic information systems and science*. New York: Wiley.
- Mark, D. M. 2003. Geographic information science: Defining the field. In *Foundations of geographic information science*, ed. M. Duckham, M. F. Goodchild, and M. F. Worboys, 3–18. New York: Taylor and Francis.
- Mark, D. M., M. J. Egenhofer, S. Hirtle, and B. Smith. 2000. UCGIS emerging research theme: Ontological foundations for geographic information science. http://www.ucgis.org/priorities/research/research_white/2000%20Papers/emerging/ontology_new.pdf (last accessed 26 July 2004).
- Mark, D. M., and A. G. Turk. 2003. Landscape categories in Yindjibarndi: Ontology, environment, and language. In *Spatial information theory: Foundations of geographic information science*, ed. W. Kuhn, M. Worboys, and S. Timpf. Lecture notes in computer science 2825, 28–45. Berlin: Springer.
- Miller, H. J. 2003. A measurement theory for time geography. Forthcoming, *Geographical Analysis*. http://www.geog.utah.edu/~hmiller/papers/time_geog_measure.pdf (last accessed 26 July 2004).
- Montello, D. R., S. I. Fabrikant, M. Ruocco, and R. S. Middleton. 2003. Testing the first law of cognitive geography on point-display spatializations. In *Spatial information theory: Foundations of geographic information science*, ed. W. Kuhn, M. Worboys, and S. Timpf. Lecture notes in computer science 2825, 316–31. Berlin: Springer.
- Openshaw, S. 1983. *The modifiable areal unit problem. Concepts and techniques in modern geography (CATMOG)* No. 38. Norwich, U.K.: Geo Books.
- Openshaw, S., and P. J. Taylor. 1979. A million or so correlation coefficients: Three experiments on the modifiable areal unit problem. In *Statistical applications in the spatial sciences*, ed. N. Wrigley, 127–44. London: Pion.
- O'Sullivan, D., and D. J. Unwin. 2003. *Geographic information analysis*. New York: Wiley.
- Peuquet, D. J. 2002. *Representations of space and time*. New York: Guilford.
- Shekhar, S., and S. Chawla. 2003. *Spatial databases: A tour*. Upper Saddle River, NJ: Prentice Hall.
- Sui, D. 2004. Forum: Tobler's first law of geography: A big idea for a small world? *Annals of the Association of American Geographers* 94: 269–77.
- Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46:234–40.
- USGIS. Brochure. <http://www.ucgis.org/aboutucgis/ucgisbrc2003.pdf> (last accessed 27 September 2004).
- van Deursen, W. P. A. 1995. *Geographical information systems and dynamic models: Development and application of a prototype spatial modelling language*. Utrecht, Netherlands: Koninklijk Nederlands Aardrijkskundig Genootschap/Faculteit Ruimtelijke Wetenschappen Universiteit Utrecht.
- Winter, S., ed. 2001. Special issue: Ontology in the geographic domain. *International Journal of Geographical Information Science* 15 (7): 587–90.
- Worboys, M. F. 1995. *GIS: A computing perspective*. New York: Taylor and Francis.
- Zhang, J. X., and M. F. Goodchild. 2002. *Uncertainty in geographical information*. New York: Taylor and Francis.

Correspondence: Department of Geography, University of California, Santa Barbara, CA 93106-4060, e-mail: good@geog.ucsb.edu.