CHAPTER 1

Geographic Information Science: Defining the Field

David M. Mark
Department of Geography, University of Buffalo
Buffalo, NY 14261, USA

1 INTRODUCTION

In the last decade, Geographic Information Science has emerged as a focus of considerable academic attention. To some extent, it is the Earth’s New Science, just as Cognitive Science was the Mind’s New Science a decade or two earlier (Gardner, 1985). But it is not clear how deep or lasting the impact of GI Science will be, either on academia or on the GIS industry. Rather than following the success of Cognitive Science, GI Science could just as easily be the next Regional Science, a similar fusion of disciplines that peaked early and continues today mainly as an internally successful multidisciplinary field of relatively low influence on science, technology, or society. Worse yet, Geographic Information Science could largely be just a pretentious name for Geographic Information Systems, and not really a scientific or intellectual field at all. This paper seeks to explore these issues, and to lay out the intellectual scope of Geographic Information Science.

1.1 What is Geographic Information Science?

What is Geographic Information Science? In the article in which he introduced the term, Goodchild (1992) did not provide a definition, but rather outlined the scope of the field indirectly by listing the major components of the Geographic Information Science research agenda.

A written definition of the field followed when, in December 1994, a group of academics met in Boulder, Colorado, to establish a new organization to represent the GIS basic research committee. There was much debate over the name of the nascent organization, and votes over each word in the title: they settled on calling it the University Consortium for Geographic Information Science (UCGIS) (Mark and Bossler, 1995). Having chosen that name, the group was more or less compelled to provide a definition for the field. Again, though, the definition they provided was somewhat indirect:

“The University Consortium for Geographic Information Science is dedicated to the development and use of theories, methods, technology, and data
for understanding geographic processes, relationships, and patterns. The transformation of geographic data into useful information is central to geographic information science.” (UCGIS, 2002)

A full definition of GI Science was provided in a report on a workshop held in January 1999 at the National Science Foundation, Geographic Information Science:

“Geographic Information Science (GIscience) is the basic research field that seeks to redefine geographic concepts and their use in the context of geographic information systems. GIScience also examines the impacts of GIS on individuals and society, and the influences of society on GIS. GIScience re-examines some of the most fundamental themes in traditional spatially oriented fields such as geography, cartography, and geodesy, while incorporating more recent developments in cognitive and information science. It also overlaps with and draws from more specialized research fields such as computer science, statistics, mathematics, and psychology, and contributes to progress in those fields. It supports research in political science and anthropology, and draws on those fields in studies of geographic information and society.” (Mark, 2000)

However, the community has not fully adopted such a definition of GI Science as a fundamental research field. Recently, at least in the United States, there has been a tendency to use the term “Geographic Information Science” to refer to almost any academic work that involves Geographic Information Systems (GIS), often without changing the content of conventional “GIS” teaching programs at all. Despite this, it appears that the field has considerable depth, and a richness of intellectual challenges that mark it as a legitimate multidisciplinary field and perhaps an emerging new discipline. Thus, the author is comfortable asserting that GI Science is most certainly not just a new name for GIS training and applications. In a purist view of GI Science, even the use of GIS as a tool in scientific research is not GI Science—at a recent NSF workshop, this latter area was termed “Research Using GIS” (Mark, 2000). Research Using GIS, is important to the sciences, and to funding for GIS-related scientific work, but is not GI Science per se. Of course, since current commercial GIS may not have all the capabilities to support scientific research in other fields, Research Using GIS may reveal important research topics for GI Science to address, and contribute to the GI Science research agenda.

1.2 What are Information and Information Science?

Given the name chosen for the field, it would appear that Geographic Information Science would be a branch of Information Science. Does adding the word “geographic” to a good definition of Information Science produce a reasonable definition of GI Science? Unfortunately, it is difficult to find good definitions of information. The Oxford English Dictionary provides several definitions. The definitions that appear closest to what we are dealing with in GI Science are:

3a. “Knowledge communicated concerning some particular fact, subject, or event; that of which one is apprised or told; intelligence, news. spec. contrasted with data.”
3c. “that which inheres in one of two or more alternative sequences, arrangements, etc., that produce different responses in something, and which is capable of being stored in, transferred by, and communicated to inanimate things.”

Next, “information science” is defined in the OED as follows:

“(that branch of knowledge which is concerned with) the procedures by which information, esp. that relating to technical or scientific subjects, is stored, retrieved, and disseminated.”

Although the life of philosopher Edmund Husserl (1859–1938) pre-dates the computer age, he provided an intriguing way to characterize information (Smith, 1989). When a mental act of an intelligent agent is directed towards an object in the real world, information is produced. In this view, information is a key part of the content of mental acts, specifically the conceptual or communicable part of the content of mental acts. This definition of information ties it to conceptualisations, intelligent agents, and cognition, which may prove especially valuable.

A practical definition of information science is provided by Shuman. After confirming the impression that “information science is very difficult to define” (p. 15), Shuman states:

“Information science is very difficult to define. ... the field of information science, however, may be defined as one that investigates the properties and behavior of information, how it is transferred from one mind to another, and optimal means for making that transfer, in both natural and artificial systems. Finally, information science is concerned with the effects of information on people and on machines.” (Shuman, 1992, p. 15)

Insert the word “geographic” in front of “information”, and this appears to be a reasonable definition for our field, and the contention that Geographic Information Science is, fundamentally, a branch of information science, seems quite tenable.

2 HISTORY OF GEOGRAPHIC INFORMATION SCIENCE

The history of GI Systems in itself an important topic worthy of much study, and the article by Coppock and Rhind (1991) and the edited book by Foresman (1998) are excellent beginnings that only scratch the surface. Recall, however, that this chapter is about Geographic Information Science, not about Geographic Information Systems, and thus restricts the history discussion to the emergence of the science.

2.1 The NCGIA solicitation

Although geographic information systems and computer cartography generated a number of challenging conceptual and computational problems from their initiation in the 1960s or even earlier, with academics involved since the early days, the idea that there might be an academic field of study, a science, behind GIS software technology, came in the 1980s. A pivotal event in the development of the science came when the field was targeted for
significant funding by the US National Science Foundation (NSF). In 1984, Ron Abler joined the NSF as Program Director for the Geography and Regional Science program. Soon, he saw an opportunity to obtain NSF support for a science and technology research centre specifically devoted to GIS, as an opportunity to direct some of NSF’s interest in research centres and ‘Big Science’ toward the discipline of Geography. Abler wrote the NSF solicitation for a National Center for Geographic Information and Analysis (NCGIA) after meeting with members of the GIS and quantitative geography research communities (Abler, 1987, p. 304). The scientific core of this solicitation was a set of five research topics that an NCGIA should consider addressing. Since these topics were set out as bullet points in the NSF solicitation, they later became known as “the five bullets”:

- spatial analysis and spatial statistics;
- spatial relations and database structures;
- artificial intelligence and expert systems;
- visualization;
- social, economic, and institutional issues

In historical perspective, these five points might now be seen as the first definition of the scope of an emerging new research field that later became known as Geographic Information Science. The successful proposal for NCGIA claimed that the third and fourth bullets, artificial intelligence and visualization, were crosscutting methodological themes, and emphasized three basic areas indicated by the first, second, and fifth bullets NCGIA (1989).

2.2 The Naming of the field

The phrase “Geographic Information and Analysis” is a grammatically awkward yet clever effort to integrate elements of GIS with the spatial analysis tradition from geography. At the same time, academic researchers in GIS were conducting active debates about the intellectual status of the GIS field. In July of 1990, Michael F. Goodchild made a keynote address about GIS-related research priorities to the Spatial Data Handling conference in Zurich, Switzerland, the fourth conference under that title. Goodchild’s talk was entitled “Spatial Information Science”. The address was subsequently published in the International Journal of Geographical Information Science (IJGIS) under the modified title “Geographical Information Science” (Goodchild, 1992); the shift from “Spatial” to “Geographical” in the title was not discussed. In the IJGIS article, Goodchild did not provide a crisp definition of the field, but did discuss what he called “the content of geographical information science”, under the following eight headings:

1. Data collection and measurement
2. Data capture
3. Spatial statistics
4. Data modelling and theories of spatial data
5. Data structures, algorithms and processes
6. Display
7. Analytical tools
8. Institutional, managerial and ethical issues
2.3 The University Consortium for Geographic Information S...

During the 1980s, many sectors of the GIS community formed organizations to promote their interests (Brown et al., 2002). Academic researchers, while among the last, were not immune to this trend, and a series of discussions through 1992 and 1993 led to a founding meeting in Boulder, Colorado, in December 1994. In June 1996, UCGIS delegates met to determine research priorities for the new organization, and at the end of the meeting endorsed 10 research priorities.

- Spatial Data Acquisition and Integration
- Distributed Computing
- Extensions to Geographic Representation
- Cognition of Geographic Information
- Interoperability of Geographic Information
- Scale
- Spatial Analysis in a GIS Environment
- The Future of the Spatial Information Infrastructure
- Uncertainty in Spatial Data and GIS-based Analyses
- GIS and Society

There are four topics on the UCGIS research challenges that were not on Goodchild’s list of topics. One highly applied topic, “The Future of the Spatial Information Infrastructure”, is subsumed under “GIS and Society” in the other schemes. The other three are more basic: interoperability, distributed computing, and scale. Of these, scale is a compelling, crosscutting theoretical and conceptual issue, whereas interoperability and distributed computing appear to be cross-cutting technological themes.

2.4 Project Varenius

In 1995, researchers from the National Center for Geographic Information and Analysis submitted a proposal entitled “Advancing Geographic Information Science” to the National Science Foundation. This proposal defined a new vision of GI Science as a field based on three fundamental research areas:

- Cognitive Models of Geographical Space
- Computational Methods for Representing Geographical Concepts
- Geographies of the Information Society

Seen in the light of the agendas compared in this paper, the Varenius triangle is rather unbalanced, under-playing the computational components of the field. Looking at the research specifics of the Varenius project, the computational component addressed interoperability, ontology of fields, and geographic knowledge discovery and data mining.

3 COMPONENTS OF GEOGRAPHIC INFORMATION SCIENCE

The nature of the field of Geographic Information Science can be characterized by listing its components. These topics are not meant to define a research agenda. It is more like
a curriculum for the topic, the basic components of the field even if research on them is relatively mature or complete. The topics are interlinked, and there is no single linear sequence that would fully preserve their relationships. Thus, the sequence employed below is somewhat arbitrary. Lastly, the headings are neither definitive nor exhaustive, and several important topics that do not readily fit into the scheme are reported under the heading “Other Topics” below.

3.1 Ontology and representation

1. Ontology of the geographic domain

Ontology deals with what exists, and with what may possibly exist. In this sense, it is a branch of philosophy that deals with some of the most fundamental aspects of scientific inquiry, but at a very high level of abstraction. This part of GI Science examines the geographic information and geographic concepts that are used by environmental and social scientists in their research, as well as by people in general. Ontology seeks to provide a consistent formal theory of tokens (instances) and types (kinds) in the real world, their relationships, and the processes that modify them. Overviews of this topic are provided in Smith and Mark (1998, 2001). Recently, ontology was identified as an emerging theme by UCGIS, and was the topic for a special issue of the IJGIS (Winter, 2001).

2. Formal representation of geographic phenomena

More recently, the term ontology has been used in information science and knowledge representation to refer to the specifications of the conceptualisations employed by different groups of users in regard to domains of entities of different types. Characteristically, such specification involves the laying down of a computationally tractable taxonomy of the objects in the given domain of a sort that can support automatic translation from one data context to another. These representations are types or kinds in the digital domain, to be instantiated through data to become digital tokens (instances) that correspond to geographic things in reality. This topic, finding digital formalisms that can capture the essence of geographic phenomena, has traditionally been referred to in GIS as data modelling or as representation, but can also be seen as the applied side of ontology. The DIME encoding system for street networks and census districts (Cooke and Maxfield, 1967) may have been the greatest ever innovation in geographic representation, but there have been many others, and there is certainly still a need for improvements in this area (?).

3.2 Computation

3. Qualitative spatial reasoning

Reasoning about spatial relations and positions in space is a well-established research area in artificial intelligence, and has become important in GI Science as well. Cognitive and linguistic models of spatial relations predominantly involve qualitative topological principles such as contact and containment. There are two domains for spatial relations. When spatial objects are disjoint, their spatial relation is characterized by distance and direction. Distance and direction can be reported quantitatively as metric distances and angles, but
qualitative models of distance and direction correspond well with human reasoning and are appropriate for many purposes (Frank, 1992). The other domain for spatial relations is when the objects touch or overlap. Recent work on relations between non-disjoint spatial objects has been dominated by two formal frameworks. Egenhofer and his colleagues have based their work on point sets (Egenhofer and Franzosa, 1991) and extended this under the 9-intersection formalism (Egenhofer et al., 1994). A competing framework is provided by the RCC family of formalisms by Cohn and his colleagues (Cohn et al., 1997).

4. Computational geometry

Computational geometry provides fundamentals for metric representation of objects and relations in geographic space. Analytical cartography is an alternative term for many aspects of computational geometry applied to the geospatial domain. Computational solutions to geometric problems were required in the very earliest days of computer-assisted cartography and GIS. Computational geometry is challenging, in part because pure Euclidean geometry cannot be implemented in straightforward fashion in a finite-precision digital environment, as discussed in an early paper by Douglas (1974) and discussed in detail by Franklin (1984). Line simplification (Douglas and Peucker, 1973) and many other aspects of map generalization (Buttenfield and McMaster, 1991) fall under the general topic of computational geometry in GI Science, although they also relate to the cross-cutting issue of scale (below). Another set of computational geometry problems relate to the efficient computation of proximity, handled under the conceptual framework variously labelled as Voronoi diagrams, or Thiessen or proximal polygons. Preparata and Shamos (1991) provided a definitive review of these problems, and Gold has done much work to integrate these methods into Geographic Information Science (see Gold, 1994, for example).

5. Efficient indexing, retrieval, and search in geographic databases

Efficient indexing of multidimensional data is an important problem in database research in computer science. Since geographic information is inherently at least two dimensional, these indexing issues have long been important in GIS. The so-called Morton ‘matrix’ approach for ordering map areas on a sequential magnetic tape was a key early innovation in GIS (?). Morton’s index was equivalent to interleaving the bits in x- and y-coordinates expressed as integers. The idea was re-discovered in the context of image processing and retrieval in the 1970s under the label quadtrees, which recursively divide an image into quadrants and subquadrants (see Samet, 1989, for a review). Samet (1989) reviews a number of related indexing schemes such as B-trees, R-trees, k-d trees, etc.

6. Spatial statistics

Spatial statistics is an important research area with strong links to Geographic Information Science. One of the properties that make spatial information special is the frequent presence of spatial autocorrelation or spatial dependence. Spatial statistics (Cressie, 1993)
Foundations of Geographic Information Science

provides formal statistical methods for dealing with spatial autocorrelation, such as measuring it, or controlling for its effects when conducting statistical analyses based on data for spatial units. Spatial statistics can be used to characterize some aspects of data quality, but otherwise this topic appears to stand in some isolation from other components of geographic information science.

7. Other geocomputation topics

A number of additional computational topics are important to GI Science but do not fit under the headings that immediately precede this one. One of these topics is map algebra, a comprehensive conceptual framework for raster-based spatial analysis developed in several articles and summarized in a 1990 book (Tomlin, 1990). This is not just a matter of implementing standard GIS operations based on a different representation of spatial information. Rather, map algebra leads to a different way of conceptualising geocomputational problems based on proximity and local operators that could easily re-cast in a parallel computation environment. Closely related to map algebra are the many spatial operations that can be based on cellular automata (von Neumann, 1966; Couclelis, 1997).

3.3 Cognition

8. Cognitive Models of Geographic Phenomena

This research area involves the study of human perception, learning, memory, reasoning, and communication of and about geographic phenomena. An explicit agenda to examine human cognition of geographic environments was originally introduced into the GI Science agenda as a way to gain insight into the nature of spatial relations, to gain insights into geographic ontology, and to understand and improve human-computer interaction for GIS (Mark and Frank, 1991). There is a large body of work on spatial cognition in psychology and cognitive science, and some of this has dealt with the geographic domain—some benchmark examples include Stevens and Coupe (1978), Talmy (1983) and Herskovits (1986). Studies of human spatial cognition are foundational to several other areas of GI Science. Attention to formalizing common-sense concepts for geographic space was highlighted by Egenhofer and Mark (1995) under the term ‘Naïve geography’.

9. Human interaction with geographic information and technology

Human-computer interaction (HCI) for geographic information systems, and the design of user interfaces for GIS, is perhaps the most obvious example of the relevance of cognition to GIS (Mark and Gould, 1991; Medyckyj-Scott and Hearshaw, 1993; Nyerges et al., 1995). The importance of this topic as a part of the GI Science research agenda depends on whether the issues of GIS usability can be separated into general issues of human-computer interaction on the one hand and issues of geographic concepts on the other. If not, then the GI Science research community must address problems in the overlap.
3.4 Applications, institutions, and society

10. Acquisition of geographic data

For all its attention to theory, information, knowledge, and wisdom, data or measurements of positions and attributes of the geographic domain is still central to GIS. This component of GI Science starts with a solid theory of measurement in general, and builds this out to provide an account of how the geographic world and its conceptualisations are measured and converted to instantiations of the aforementioned geographic representations. If narrowly defined, geomatics is a good term for this component of geographic information science. Technologies for the acquisition of geographic data and information, chiefly remote sensing and GPS, are fields in their own rights that dwarf the rest of GI Science in terms of the public and private investments in their infrastructures.

11. Quality of geographic information

Research on the quality of geographic information, including variations under the terms accuracy or error, are important parts of the GI Science research agenda (Goodchild and Gopal, 1989). Data quality may be reduced due to measurement error during data acquisition or by various transformations of the data in GIS processing. Specification errors may occur if the ontology or representation is not done correctly. The impacts of data quality must be judged in terms of the sensitivity of models and the fitness of data for particular uses.

12. Spatial analysis

Spatial analysis is an important topic in quantitative geography, with strong ties to many aspects of Geographic Information Science. Some might argue that spatial analysis is just an application of GI Science principles to problems in environmental or social science, and that the fundamental science underlying spatial analysis is already covered by the ontology, representation, and spatial statistics topics. However, several topics, especially the “modifiable areal unit problem” (MAUP; Openshaw and Taylor, 1981; Openshaw, 1984) seems clearly to be an important class of GI Science research and curriculum topics, yet is not covered well in other parts of the GI Science agenda. Careful deliberation will be needed to divide spatial analysis into topics that are part of GI Science, and topics that are applications of GI Science.

13. Geographic information, institutions, and society

Societal impacts of GIS technology has been a part of the real agenda of GIS since its onset—indeed, before 1980, most technical innovation in GIS occurred in a direct application context, as software was developed at government agencies or by consultants, rather than by academics. The 1980s were a decade of commercialisation of software and data on the one hand and ‘academisation’ of the GIS R&D agenda on the other. Since the end of the 1980s, academic attention returned to the impacts of the technology from a new direction—post-Modern critics of quantification and technology. At the
Foundations of Geographic Information Science

same time, there was also an increase in research on economic and legal aspects of geographic information, its production and sharing, including studies of how the use of GIS and associated technologies by individuals and institutions changes efficiency, effectiveness, equity, and power in society. A particularly active area of research and practice is “Public-Participation GIS” (PPGIS). The main aspects of the post-Modern critique have been presented by Pickles in his edited book (Pickles, 1995) and a more recent review article (Pickles, 1999). A good overview of the economic and legal aspects of geographic information is found in Masser and Onsrud (1993).

3.5 Crosscutting Research Themes

14. Time

Time and temporality, motion and change, are essential to many GIS applications, yet GIS software has been notoriously weak in providing tools for handling temporal dimensions of geographic information. Long ago, Blaut (1961) proposed that the Kant/Newton separation of reality into space, time, and theme, which Berry (1964) proposed as an organizing framework for geography and GIS, made it difficult to deal with process and change. GIS seems ontologically committed to separating space and time, which then impedes certain scientific uses of GIS. If space and time can truly be studied separately and the results later assembled, then time would not be part of the agenda for GI Science. Recent interest in time in geographic space and GIS (Langran, 1992; Peuquet, 1994; Egenhofer and Golledge, 1998) suggests that time is an integral part of GI Science research and one which cuts across most other GI Science topics.

15. Scale

Scale has multiple meanings relevant to GI Science. In cartography, scale refers to size on the map divided by size in the world—small-scale maps show large regions. Map scale interacts with geometry of the world to require map generalization (Buttenfield and McMaster, 1991). In the physical sciences such as meteorology or geomorphology, the term scale is used to indicate the size, extent, or characteristic length for physical processes. Micro-, meso-, and global-scale atmospheric processes are familiar terms. In biology and geomorphology, interactions between size, shape, and function have been variously explored under the heading of allometry—sometimes shape must change systematically in order to maintain function as size changes (for a review, see Church and Mark, 1980). Thirdly, the term scale is used to summarize resolution, the smallest entities that can be detected or represented, both for display and analysis (see Quattrochi, 1997). More recently, cognitive aspects of scale have been highlighted (Montello, 1993; Montello and Golledge, 1999). The issues surrounding the term scale are important to the GI Science curriculum and to the research agenda, and cut across most of the other topics described in this paper.
4 COMPARISON OF TOPICS

Of course, the list of components of geographic information science proposed in this paper was prepared in full knowledge of the lists proposed by Goodchild (1992) and by the UCGIS. Still, it is instructive to compare the lists.

Topics on all three lists

Four topics appear on all three lists:

- data acquisition
- representation (data modelling)
- spatial analysis
- societal issues surrounding GI

The last three of these also correspond well with core issues in the bullets proposed in NSF’s NCGIA solicitation.

Topics on two lists

Three topics from Goodchild’s (1992) list also were identified as key topics in this paper, but were not on the UCGIS list:

- computational geometry
- indexing for spatial databases
- spatial statistics

In fact, Goodchild combined the first two of these topics under his “Data structures, algorithms and processes.”

Three additional topics were on the UCGIS list and in this paper but not singled out by Goodchild:

- ontology
- cognition
- data quality

Another research topic, scale, is a research topic on the UCGIS list but identified as a crosscutting research theme here.

Lastly, one topic, visualization and display, was on Goodchild’s (1992) list, and recently was identified as an emerging research theme by UCGIS, but is not on the list of GI Science fundamentals described in this paper; perhaps it should join scale and time as cross-cutting themes.

Topics on only one list

All of Goodchild’s topics from 1992 appear on one or both of the other lists. However, three of UCGIS’s original research challenges, and one of their emerging themes, appear on neither of the other lists:
Foundations of Geographic Information Science

- distributed computing
- interoperability of geographic information
- the future of the spatial information infrastructure
- geospatial data mining and knowledge discovery

Proposed in this paper, but not on either of the other lists, are four additional topics:

- human interaction with GI and technology
- qualitative spatial reasoning
- time in geographic space
- other geocomputation topics

The human-computer interaction topic is partially covered by UCGIS under its version of the “Cognition” topic, and dealing with time is part of the material in UCGIS topic, “Extensions to Geographic Representations”.

5 DISCUSSION

How can we put these proposed components of Geographic Information Science in perspective? One rather narrow but revealing way to look at the prospects for progress in GI Science is to compare the research priorities outlined above to the priorities of funding agencies. The programs of the Information and Intelligent Systems (IIS) Division of the Computer and Information Science and Engineering (CISE) Directorate at the US NSF are very interesting in this regard. As of the summer of 2000, NSF’s IIS Division was divided into 8 programs:

- Computation and Social Systems
- Human Computer Interaction
- Information and Data Management
- Information Technology Research
- Knowledge and Cognitive Systems
- Robotics and Human Augmentation
- Special Projects (IIS)
- Digital Libraries

The similarity of this subdivision of information science aspects of computer science research to the GI Science agenda is striking. This provides good evidence regarding the claim that Geographic Information Science should be regarded as a branch of Information Science. Of course, Geographic Information Science has an intimate relationship with the discipline of Geography, since they address the same aspects of reality. However, Geographic Information Science is concerned with ontology, representation, and computational issues, whereas Geography attempts to explain and predict geographic phenomena.

Material presented in this paper confirms that Geographic Information Science has considerable depth, and a richness of intellectual challenges that mark it as a legitimate multidisciplinary field and perhaps an emerging new discipline. In the writer’s opinion, progress in the field would be aided by a consensus among leading researchers on the nature, scope, and elements of the field; this paper attempts a step in that direction.
REFERENCES


Foundations of Geographic Information Science


