Geographic Information Science: An Introduction

A. Stewart Fotheringham and John P. Wilson

GIS, the acronym for Geographic Information Systems, has been around since the 1980s. Although one can impute a number of characteristics from the use of this acronym, at the heart of the term "systems" lies a computer software package for storing, displaying, and analyzing spatial data. Consequently, the use of the term GIS implies an **object** or tool which one can use for exploring and analyzing data that are recorded for specific locations in geographical space (see Cowen [1988] for an early article articulating this type of definition and Foresman [1998] for a rich and varied account of the history of Geographic Information Systems). Conversely, Geographic Information Science or GI Science, or more simply GISc, represents a much broader framework or *modus operandi* for analyzing spatial data. The term GI Science emphasizes more the **methodology** behind the analysis of spatial data (see Burrough [1986] for what was perhaps the first GIS text to promote such a framework and Chrisman [1999] for an article advocating an extended definition of GIS along these same lines). Indeed, one could define GI Science as: *any aspect of the capture, storage, integration, management, retrieval, display, analysis, and modeling of spatial data*. Synonyms of GI Science include Geocomputation, GeoInformatics, and GeoProcessing.[1]

The Breadth of GI Science

Under this definition, GI Science is clearly an extremely broad subject and captures any aspect connected with the process of obtaining information from spatial data. A feeling for this breadth can be seen in Figure 0.1 which describes a schematic of some of the elements that make up GI Science. At the top level, GI Science is concerned with the collection or capture of spatial data by such methods as satellite remotely sensed images, GPS, surveys of people and/or land, Light Detection And Ranging (LiDAR), aerial photographs, and spatially encoded digital video.[2] The key element here is to capture not only attribute information but also accurate information on the location of each measurement of that attribute. For instance, we might

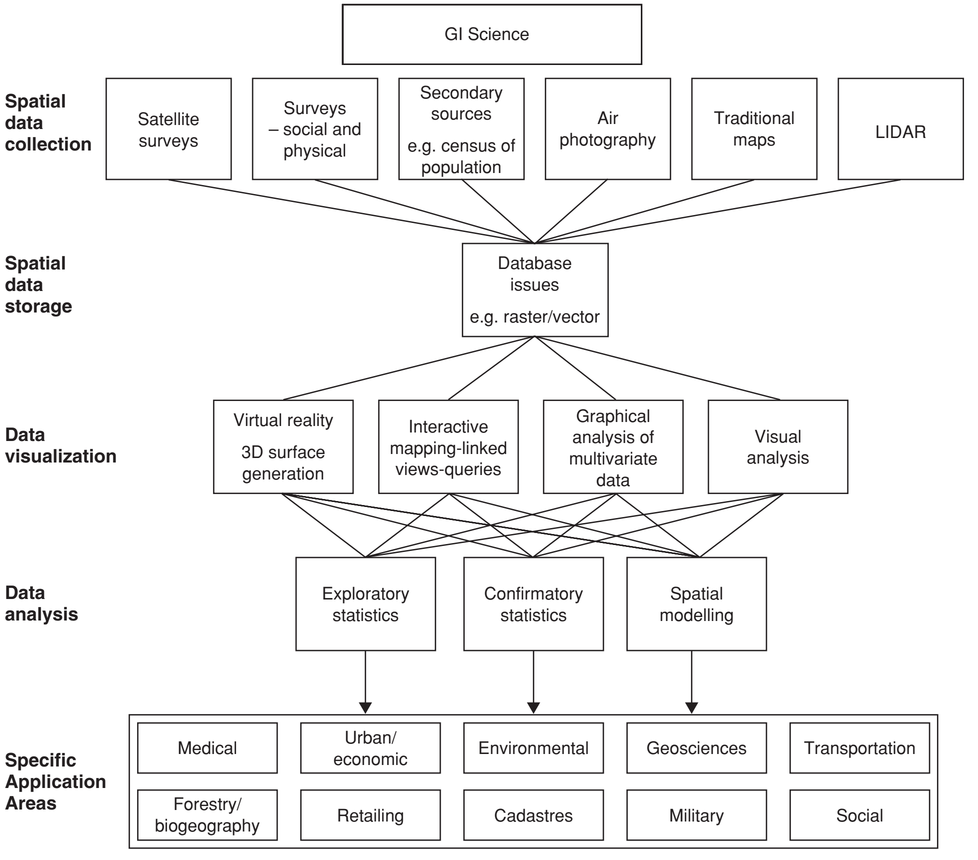

Fig. 0.1 Geographic Information Science (GISc): An Overview

ask people some information on themselves during a survey but we would also like to record some aspect of the location of that individual – this might be the location at which the survey took place, or the person's usual residence or their workplace or some other location. Similarly, if we measure some attribute such as the elevation above sea level and/or the precipitation at a set of points, we also need to know the locations of these points, otherwise the elevation and precipitation measurements are useless (see Corbett and Carter [1996] and Custer, Farnes, Wilson, and Snyder [1996] for two examples of what can be accomplished by combining locational information with elevations and precipitation measurements).

Once spatial data have been captured, they need to be stored and transmitted. This can create challenges as some spatial data sets can be extremely large. A census of population in the USA would contain, for example, records on almost 300 million people with the locational data on each person typically being his or her residence (hence the well-known problem of census data being a snapshot of where people are at midnight rather than during the day). In some countries, a decennial census has been replaced with a more continuous monitoring of the population in the form of a register which can be updated more regularly. Satellite imagery of the Earth's surface can generate terabytes of data and the move towards global data sets can lead to even larger data sets. The data storage and transmission demands of spatially encoded digital video have already been referred to in endnote 2 and pose challenges to current systems. Consequently, spatial data sets can be extremely large and finding ways to store, process, and transmit such large data sets efficiently is a major challenge in GI Science.

The next two levels of operations in the schematic in Figure 0.1 refer to the process of transforming data into information. We are currently living in a data-rich world that is getting richer by the day. In many operations, large volumes of spatial data are being collected and a major challenge in GI Science is to turn these data into useful information. Consider, for example, the following sources of spatial data (which are but a small sample from the complete set of sources):

- Censuses of population which typically occur every five or 10 years and which typically record information on each individual in each household and on the household itself;
- Customer databases held by retail-related companies which hold information on individuals submitted in various application forms or warranty cards;
- Traffic flow monitoring along streets or at intersections;
- LiDAR – low pass fly-overs by plane generating large volumes of detailed data on terrain features or urban areas;
- Digital Elevation Models (DEMs) captured via satellites or the US space shuttle which can be at a global scale;
- Health records either on the location of patients with particular diseases, used to study possible geographic influences on etiology or to assess the level of demand for various services in particular hospitals;
- Satellite remotely-sensed images or aerial photographs used to track land use change over time or to study the spatial impacts of various natural disasters or for various military uses such as tracking missiles or identifying targets;
- Satellite GPS used increasingly for general data capture of vehicles and individuals. This makes possible vehicle tracking, in-car navigation systems, precision agriculture, animal tracking, monitoring of individuals, and general data capture on the location of objects via GPS receivers. It is now possible to contemplate, as the UK is doing, tracking the movement of all vehicles and charging for per mileage road use instead of a flat road tax. Similarly, it is now possible via GPS to monitor a child's movements via a GPS watch linked to a central monitoring system that parents can access remotely via the World Wide Web (see http://www.wherifywireless.com for additional details). The linking of GPS to mobile phones will allow the tracking of friends so that one can query the location of a registered friend at any moment. The use of mobile phones to locate individuals by triangulation from mobile relay stations is already standard police practice in the case of missing persons. Some of these uses of course immediately raise important ethical and legal questions which need to be resolved. Just how much spatial information on ourselves are we prepared to have captured and stored?

Most organizations simply do not have the resources (measured in terms of personnel, knowledge, and/or software) to be able to make full use of all the data they routinely gather. There is a growing need for techniques that allow users to make sense out of their spatial data sets. This mirrors the general transformation of society from one dominated by the industrial revolution with its origins in the eighteenth and nineteenth centuries, to one dominated by the information revolution with its origins in the computer age of the late twentieth century. Two main sets of techniques exist to turn data into information: visualization and statistical/ mathematical modeling.

There is a vast array of techniques that have been developed for visualizing spatial data and it remains a very intense and fruitful area of research (see Dykes, MacEachren, and Kraak [2005] for one such treatment). Spatial data lend themselves to visualization because the data are geocoded and can therefore be represented easily on maps and map-like objects. Simply mapping spatial data can shed so much more light on what is being studied than if the data are presented in tabular form.[3] However, maps can also deceive (see Monmonier [1991] for a popular treatment of this topic) and there are many GI Science issues that need to be considered if spatial data are to be displayed to provide reasonably accurate information content. The development of algorithms for continuous cartograms, software to create pseudo-3D virtual reality environments, and hardware that allows digital video to be linked to a GPS is providing us with the means towards much more sophisticated visualization of spatial data than traditional 2D maps and there are great advances in this area yet to come. Because spatial data contain attribute and locational information, the data can be shown together as on a simple map of the distribution of an attribute or they can be displayed separately in different windows. The use of multiple windows for displaying spatial data is now commonplace and it can sometimes provide useful information to display a map in one window and a non-spatial display of the data (such as a histogram or scatterplot for example) in another window and to provide a link between the two (see GeoDA™ [Anselin, Syabri, and Kho 2006] and STARS [Rey and Janikas 2006] for examples of such systems). In this way, only data highlighted on the non-spatial display need to be mapped to show the spatial distribution of extreme values, for example. Alternatively, all the data can be displayed on a map and only the data points selected on the map need to be highlighted in the non-spatial display. Finally, many spatial data sets are multivariate and it is a major challenge to try to represent such complex data in one display.

The other set of methods to turn spatial data into information are those involving statistical analysis and mathematical modeling. Statistical analysis was traditionally dominated by what is known as "confirmatory" analysis in which a major objective was to examine hypotheses about relationships that were already formed. The typical approach to confirmatory analysis would be to develop a hypothesis about a relationship from experience or the existing literature and to use statistical techniques to examine whether the data support this hypothesis or not. Confirmatory statistical analysis generally depends on assessing the probability or likelihood that a relationship or pattern could have arisen by chance. If this probability or likelihood is very low, then other causes may be sought. The assessment of the role of chance necessitates the calculation of the uncertainty of the results found in a set of data (if we had a different data set, would the results perhaps be substantially different or pretty much the same?). In classical statistical methods, this calculation typically assumes that the data values are independent of each other. A major problem arises in the use of this assumption in spatial data analysis because spatial data are typically not independent of each other. Consequently, specialized statistical techniques have been developed specifically for use with spatial data (see Bailey and Gattrell [1995] for an informative and accessible summary of some of these techniques) and a great deal more research is needed in this area.

More recently, and probably related to the recent explosion of data availability, "exploratory" statistical techniques have increased in popularity. With these, the emphasis is more on developing hypotheses from the data rather than on testing hypotheses. That is, the data are manipulated in various ways, often resulting in a visualization of the data, so that possible relationships between variables may be revealed or exceptions to general trends can be displayed to highlight an area or areas where relationships appear to be substantially different from those in the remainder of the study region. A whole set of localized statistical techniques has been developed to examine such issues (for example, Fotheringham, Brunsdon, and Charlton 2000).

Finally, spatial modeling involves specifying relationships in a mathematical model that can be used for prediction or to answer various "what if" questions. Classical spatial models include those for modeling the movements of people, goods, or information over space and the runoff of rainwater over a landscape. There is a fuzzy boundary between what might be termed a mathematical model and what might be termed a statistical model. Quite often, models are hybrids where a formulation might be developed mathematically but the model is calibrated statistically. Where models are calibrated statistically from spatial data, one important issue is that it is seldom clear that all, or even most, relationships are stationary over space, usually an assumption made in the application of various modeling techniques. For instance, the application of traditional regression modeling to spatial data assumes that the relationships depicted by the regression model are stationary over space. Hence, the output from a regression model is a single parameter estimate for each relationship in the model. However, it is quite possible that some or all of the relationships in the model vary substantially over space. That is, the same stimulus may not provoke the same response in all parts of the study region for various contextual, administrative or political reasons – people in different areas, for example, might well behave differently. Consequently, specialized statistical techniques such as Geographically Weighted Regression (GWR) have been developed recently to allow for spatially varying relationships to be modeled and displayed (Fotheringham, Brunsdon, and Charlton 2002).

The final layer of Figure 0.1 represents some of the application areas of geocomputation which gives an indication of why it is such an important and rapidly growing area of study. Spatial data can be found in most areas of study and include many different types of data, such as:

- Geodetic – coordinate reference systems for locating objects in space;
- Elevation – recording heights of objects above mean sea level;
- Bathymetric – recording the depth of water bodies;
- Orthoimagery – georeferenced images of the earth's surface;
- Hydrography – data on streams, rivers and other water bodies;

- Transportation networks – roads, railways, and canals;
- Communication networks – the transmission of ideas and data across space;
- Cadastral – precise positioning of property boundaries;
- Utilities – the locations of pipes, wires and access points;
- Boundaries – electoral, administrative, school and health districts;
- Medical – the location of incidents of disease and patients with respect to the location of services;
- Crime – the location of police incidents;
- Environmental – habitats, pollution, and the impacts of natural disasters;
- Urban – the location of areas of high priority for social and economic intervention;
- Planning – the spatial impacts of locational decisions;
- Retailing – the location of consumers with respect to the location of services;
- Biogeography – the location of one species with respect to the location of one or more others.

We now turn to a brief discussion of the topics covered in this book. Given the enormous breadth of GI Science, it is clear that not everything can be included in this volume. However, in order to be as comprehensive as possible, we have tried to solicit contributions which have a fairly general application as opposed to being strictly about the use of GI Science in one particular field.

What Follows Next!

The remainder of this book is organized under six headings. We start each section with a brief description of the chapters that follow and the chapters, themselves, offer stand-alone treatments that can be read in any order the reader chooses. Each chapter includes links to other chapters and key references in case the reader wants to follow up specific themes in more detail.

The first group of six chapters looks at some of the recent trends and issues concerned with geographic data acquisition and distribution. Separate chapters describe how the production and distribution of geographic data has changed since the mid-1970s, the principal sources of social data for GIS, remote sensing sources and data, the possibilities of using spatial metaphors to represent data that may not be inherently spatial for knowledge discovery in massive, complex, multi-disciplinary databases, the myriad sources of uncertainty in GIS, and the assessment of spatial data quality.

The second section of the book explores some of the important and enduring database issues and trends. Separate chapters describe relational, object-oriented and object-relational database management systems, the generation of regular grid digital elevation models from a variety of data sources, the importance of time and some of the conceptual advances that are needed to add time to GIS databases, and new opportunities for the extraction and integration of geospatial and related online data sources.

The third section of the book consists of seven chapters that examine some of the recent accomplishments and outstanding challenges concerned with the visualization of spatial data. Separate chapters describe the role of cartography and interactive multimedia map production, the role of generalization and scale in a digital world, the opportunities to display and analyze a variety of geographical phenomena as surfaces, fuzzy classification and mapping in GIS, predictive rule-based mapping, multivariate visualization, and the ways in which digital representations of two-dimemsional space can be enriched and augmented through interactivity with users in the third dimension and beyond.

The fourth section of the book contains three chapters looking at the increasingly important task of knowledge elicitation. These chapters examine the role of inference and the difficulties of applying these ideas to spatial processes along with the process of geographic knowledge discovery (GKD) and one of its central components, geographic data mining, and the prospects for building the geospatial semantic web.

The next group of four chapters examines spatial analysis. The links between quantitative analysis and GIS, spatial cluster analysis, terrain analysis, and dynamic GIS are discussed here.

The six chapters of Part VI examine a series of broader issues that influence the development, conduct, and impacts of geographic information technologies. Separate chapters examine institutional GIS and GI partnering, public participation GIS, GIS and participatory decision-making, several participatory mapping projects from Central America to illustrate the dynamic interplay between conceptions of people and place and the methods used to survey them, the relationship between GIS, personal privacy, and the law across a variety of jurisdictions, and the major developments and opportunities for educating oneself in GIS.

Finally, Part VII examines future trends and challenges. Separate chapters examine the role of the World Wide Web in moving GIS out from their organization- and project-based roles to meet people's personal needs for geographic information, the emergence of location-based services (LBS) as an important new application of GIS, and two views of the challenges and issues that are likely to guide GI Science research for the next decade or more.

Closing Comments

This handbook seeks to identify and describe some of the ways in which the rapidly increasing volumes of geographic information might be turned into useful information. The brief introductions to topics offered in the previous section give some clues as to what we think is important here – the rapid growth in the number and variety of geographic data sets, finding new ways to store, process, and transmit these data sets, new forms of visualization and statistical/mathematical modeling, etc. To the extent that this book has helped to clarify the current state of knowledge and indicate profitable avenues for future research, it will have helped to educate and inform the next generation of geographic information scientists and practitioners. This generation will need to be more nimble than its predecessors given the rapid rate of technological change (innovation) and the tremendous growth of geographic information science, geographic information systems, and geographic information services that is anticipated in the years ahead. With this in mind, we hope the reader will tackle the remainder of the book with an opportunistic and forward-looking view of the world around them.

ENDNOTES

1 In some circumstances, Geomatics also is used synonymously with GI Science as in Geomatics for Informed Decisions (GEOIDE), the Canadian Network Centre of Excellence headquartered at Laval University (www.geoide.ulaval.ca). However, in other circumstances, such as in the naming of academic departments in the UK, the term Geomatics has been used to "re-brand" Departments of Surveying where its scope and purpose are much more restricted.

2 While we are used to seeing orthophotographs, photographs with associated files giving information on the location of each pixel so that operations can be carried out on the spatial relationships within the photograph, spatially encoded digital video allows the user to perform spatial queries and spatial analysis on video images. As one can imagine, the volumes of such data that need to be stored and transmitted create special challenges.

3 See Ian McHarg's 1969 book entitled "Design with Nature" for an influential book that documented how maps could be overlaid and used to evaluate the social and environmental costs of land use change. This book has been reprinted many times and still serves as an important text in many landscape architecture courses and programs.

REFERENCES

Anselin, L., Syabri, I., and Kho, Y. 2006. GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis* 38: 5–22.

Bailey, T. C. and Gattrell, A. C. 1995. *Interactive Spatial Data Analysis*. New York: John Wiley and Sons.

Burrough, P. A. 1986. *Principles of Geographical Information Systems for Land Resources Assessment*. New York: Oxford University Press.

Chrisman, N. R. 1999. What does "GIS" mean? *Transactions in GIS* 3: 175–86.

Corbett, J. D. and Carter, S. E. 1996. Using GIS to enhance agricultural planning: The example of inter-seasonal rainfall variability in Zimbabwe. *Transactions in GIS* 1: 207–18.

Cowen, D. J. 1988. GIS versus CAD versus DBMS: What are the differences? *Photogrammetric Engineering and Remote Sensing* 54: 1551–4.

Custer, S. G., Farnes, P., Wilson, J. P., and Snyder, R. D. 1996. A comparison of hand- and spline-drawn precipitation maps for mountainous Montana. *Water Resources Bulletin* 32: 393–405.

Dykes, J. A., MacEachren, A. M., and Kraak, M. J. (eds). 2005. *Exploring Geovisualization*. Amsterdam: Elsevier.

Foresman, T. W. (ed.). 1998. *History of Geographic Information Systems: Perspectives from the Pioneers*. Englewood Cliffs, NJ: Prentice Hall.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. E. 2000. *Quantitative Geography: Perspectives on Spatial Data Analysis*. Thousand Oaks, CA: Sage Publishers.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. E. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: John Wiley and Sons.

McHarg, I. L. [1969] 1994. *Design with Nature* (25th anniversary edn). New York: John Wiley and Sons.

Monmonier, M. 1991. *How to Lie with Maps*. Chicago: Chicago University Press.

Rey, S. J. and Janikas, M. V. 2006. STARS: Space-time analysis of regional systems. *Geographical Analysis* 38: 67–86.