

to combine the potential of e-learning tools and open-source software to strengthen education in GI science, with particular emphasis on fast-changing needs in low-income countries.

Still other software is sold as closed commercial packages by established GI-system vendors, such as Autodesk Inc. (San Rafael, California; www.autodesk.com), Esri, Inc. (Redlands, California; www.esri.com), Intergraph Corp. (Huntsville, Alabama; www.intergraph.com/), or MapInfo Corp. (Troy, New York; www.mapinfo.com). Each vendor offers a range of products, designed for different levels of sophistication, different volumes of data, and different application niches. Idrisi (Clark University, Worcester, Massachusetts, www.clarklabs.org) is an example of a GI system produced and marketed by an academic institution rather than by a commercial vendor (for further information on GI system sources see Chapter 6).

Michael de Smith, along with two of the authors of this book, has produced an online guide (www.spatialanalysisonline.com) and book that is intended to raise awareness of the range of commercial and open software options that are available and the quality of the results that may be produced.

It is not always easy to compare the software solutions suggested by Internet searches.

The third component of a GI system is the data, which provide the foundations for digital representation of selected aspects of some specific area of the Earth's surface or near surface. A database might be built for one major project, such as the location of a new high-voltage power transmission corridor, or it might be continuously maintained, fed by the daily transactions that occur in a major utility company (e.g., installation of new underground pipes, creation of new customer accounts, and daily service-crew activities). As Open Data (see Section 17.4) become more freely available for download, the data that are downloaded for particular projects are frequently obtained from different sources, and thus the constituents of a GI database may have originally been assembled or collected for widely varying purposes and to widely varying standards. We discuss some of the implications of this in Chapters 5 and 17. The size of a project database may be as small as a few megabytes (a few million bytes, easily stored on a DVD) or as large as many terabytes (see Table 1.2).

Big Data is a term that has come to describe individual or linked data sets that are too large and complex to process using standard data-processing software or database-management tools on standard computer servers (see Box 17.2). Geographic databases are often big, not least of all because they include large numbers of location coordinates and

sometimes many raster images. This poses significant challenges of data capture, storage, maintenance, sharing, visualization, and analysis. Scientists regularly encounter limitations in their abilities to manage and analyze Big Data in fields such as genomics (see Box 1.4) and meteorology. These discipline-specific problems are becoming more pervasive as more and more data are gathered using ubiquitous information-sensing mobile devices, remote sensing, radio-frequency identification (RFID) tagging, and wireless sensor networks. The world's technological per capita capacity to store information has roughly doubled every 40 months since the 1980s, and as of 2012, an average of 2.5 petabytes of data were created every day. This poses important management challenges for organizations that need to decide who should own Big Data initiatives that straddle their operation. The management and analysis of Big Data are closely associated with developments in Cloud computing.

Major GI applications also require management. An organization must establish procedures, lines of reporting, control points, and other mechanisms for ensuring that its GI activities meet its needs, stay within budgets, maintain high quality, avoid breaking the law, and generally meet the needs of the organization. These issues are explored in Chapters 16, 17, and 18.

Finally, a GI system is useless without the people who design, program, and maintain it, supply it with data, and interpret its results. The people of a GI system will have various skills, depending on the roles they perform. Almost all will have the basic knowledge needed to work with geographic data—knowledge of such topics as data sources, scale and accuracy, and software products—and will also have a network of acquaintances in the GI community. Most important, they will have a capacity for critical spatial thinking, allowing them to filter the message of spatial data through the medium of a GI system.

1.5 The Disciplinary Setting of GI Science and Systems (GISS)

At this point we review the emergence of GI science, the ways in which it is used, and its relationship with other disciplines. We discuss the significance of its underpinning technologies to business as well as some of the issues arising from its use in government. It should already be apparent that computer science is important because GI systems are computer applications, and its perspective is addressed in Section 1.5.4. Similarly geography, the science and

Biographical Box 1.4

Sir Walter Bodmer, Human Geneticist

Sir Walter Bodmer (Figure 1.13A) is a German-born British human geneticist. He studied mathematics and statistics at Cambridge University, doing his PhD under the renowned statistician and geneticist Sir Ronald A. Fisher before joining Nobel-Prize-winning microbiologist Joshua Lederberg's laboratory in the Genetics Department of Stanford University in 1961. He was an early pioneer of the use of computing to study population genetics, and after a period as a faculty member at Stanford, he left to become the first Professor of Genetics at Oxford University in 1970.

Population genetics is the study of the changing distribution of gene variants (or alleles) under the influence of four important evolutionary processes: natural selection, genetic drift, mutation, and gene flow. Walter was one of the first to suggest the idea of identifying the physical and functional characteristics of the 20,000–25,000 genes of the human genome. This idea was subsequently pursued in the Human Genome Project, which in important respects remains the ultimate investigative analysis using Big Data.

Geography is central to study of the human genome, for the very good reason that most families remain settled in one part of the world for many generations. In 2005 Walter was appointed to lead a major (£2.3 million/\$3.8 million) project to examine geographic variations in the genetic makeup of the people of the British Isles. The aim of this project is to measure the genetic profiles

of long-established families who can trace their recent ancestry to particular locations and to relate this to historical and archaeological evidence of invasion and settlement. The DNA samples of thousands of volunteers have been analyzed in ways that reveal the biological traces of successive waves of colonizers of Britain—such as the original ancient British settlers, the Anglo-Saxons, and the Vikings. The resulting genetic map (Figure 1.13B) shows, for example, that the Viking invasion of Britain was predominately by Danish Vikings, whereas the Orkney Islands were settled by Norwegian Vikings.

Walter was a pioneering advocate of public engagement with science and technology and remains very active in this field. He was elected a Fellow of the Royal Society in 1974 and was awarded the Society's Royal Medal in 2013 for seminal contributions to population genetics, gene mapping, and our understanding of familial genetic disease.

Check out the People of the British Isles project at www.peopleofthebritishisles.org.

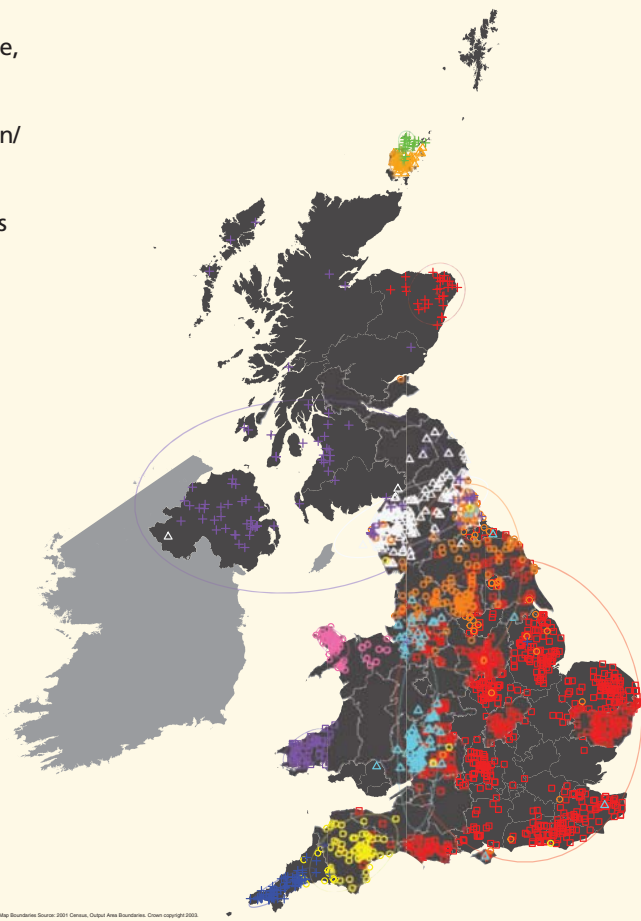


Figure 1.13 (A) Walter Bodmer, human geneticist; and (B) a genetic map of the United Kingdom, with different colors denoting the genetic groups of long-settled residents.

study of phenomena distributed over the surface and near-surface of the Earth, also provides much of the disciplinary context of GI science and is addressed in Section 1.5.5. Finally, we consider some of the ways in which the use of GI is embedded in society.

1.5.1 The Historical Perspective

Although the coining of the term “GI science” can be traced to Goodchild’s paper published in 1992, greater controversy surrounds the emergence of GI systems because parallel developments occurred in North America, Europe, and Australia (at least). Until recently, it was convenient to think of geographic information handling as confined to a freestanding, self-contained, computer-based system, like many other pieces of equipment. Indeed, prior to the innovation of the Internet, or the intranets of large organizations, the system was a physically isolated system of computer hardware, software, and data, such as a desktop computer, with no connections to the rest of the world. It was the extraction of simple geographic measures that largely drove the development of the first GIS to be described as such, the Canada Geographic Information System or CGIS, in the mid-1960s. The Canada Land Inventory was a massive effort by the federal and provincial governments to identify the nation’s land resources and their existing and potential uses. The most useful results of such an inventory are measures of area, yet area was (and still is) notoriously difficult to measure accurately from a paper map (see Section 14.1.1). CGIS was planned and developed as a measuring tool, a producer of tabular information, rather than as a mapping tool.

The first GI system was the Canada Geographic Information System, designed in the mid-1960s as a computerized map-measuring system.

A second burst of innovation occurred in the late 1960s in the U.S. Bureau of the Census, in planning the tools needed to conduct the 1970 Census of Population. The DIME (Dual Independent Map Encoding) program created digital records of all U.S. streets to support automatic referencing and aggregation of census records. The similarity of this technology to that of CGIS was recognized immediately and led to a major program at Harvard University’s Laboratory for Computer Graphics and Spatial Analysis to develop a general-purpose GIS that could handle the needs of both applications—a project that led eventually to the ODYSSEY GIS software of the late 1970s.

Early GI system developers recognized that the same basic needs were present in many different application areas, from resource management to the census.

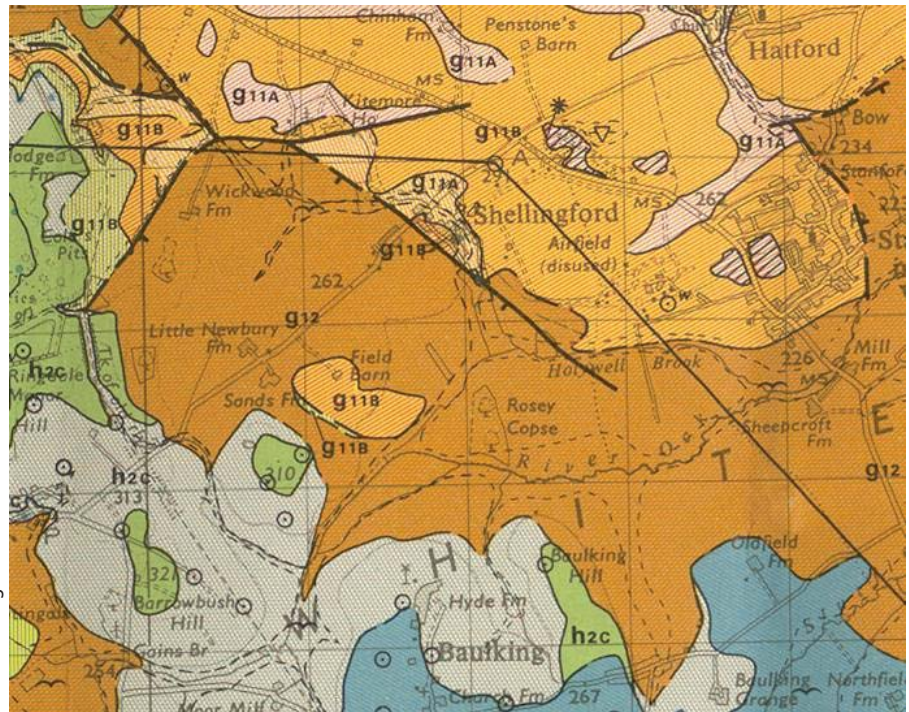
In a largely separate development during the latter half of the 1960s, cartographers and mapping agencies had begun to ask whether computers might be adapted to their needs and possibly to reducing the costs and shortening the time of map creation. The UK Experimental Cartography Unit (ECU) pioneered high-quality computer mapping in 1968; it published the world’s first computer-made map in a regular series in 1973 with the British Geological Survey (Figure 1.14). National mapping agencies, such as Britain’s Ordnance Survey, France’s Institut Géographique National, and the U.S. Geological Survey and Defense Mapping Agency (now the National Geospatial-Intelligence Agency) began to investigate the use of computers to support the editing and updating of maps, to avoid the expensive and slow process of hand correction and redrafting. The first automated cartography developments occurred in the 1960s, and by the late 1970s most major cartographic agencies were already computerized to some degree. But the limits of the technology of the time ensured that it was not until 1995 that the first country (Britain) achieved complete and detailed digital map coverage in a database.

Remote sensing also played a part in the development of GI systems, as a source of technology as well as a source of data. The first military satellites of the 1950s were developed and deployed in great secrecy to gather intelligence, and although the early spy satellites used conventional film cameras to record images, digital remote sensing began to replace them in the 1960s. By the early 1970s civilian remote-sensing systems such as Landsat were beginning to provide vast new data resources on the appearance of the planet’s surface from space and to exploit the technologies of image classification and pattern recognition that had been developed earlier for military applications. The military was also responsible for the development in the 1950s of the world’s first uniform system of measuring location, driven by the need for accurate targeting of intercontinental ballistic missiles, and this development led directly to the methods of positional control in use today (see Section 4.7). Military needs were also responsible for the initial development of the Global Positioning System (GPS; see Section 4.9 and Box 17.7).

Many technical developments in GI systems originated in the Cold War.

GI systems really began to take off in the early 1980s, when the price of computing hardware had fallen to a level that could sustain a significant software industry and cost-effective applications. Among the first customers were forestry companies and natural-resource agencies, driven by the need to keep track of vast timber resources and to regulate their use effectively.

Courtesy: British Geological Survey and Ordnance Survey
© NERC. All rights reserved. IPR/59-13C.



on the collection of geographically referenced data or activities that build on this, such as value-added reselling of data, consulting, app development, training, system integration, software design, and so forth. Many new opportunities (particularly for start-up companies) are arising from use of open-source software or Open Data (see Section 1.5.3), where the underpinning costs of data or software production are met through volunteer activity or are underwritten by research institutions such as universities.

Such activity has mushroomed with the ever-wider use of the Internet to disseminate software and data, as well as the innovation and use of search engines, blogs, and social networking sites to spread news about what is available. This section looks at the diverse roles that people play in the business of GI systems, and is organized by the major areas of human activity associated with it.

1.5.2.1 The Software Industry

Many of the roots of the development of GI systems can be traced to the commercial development of off-the-shelf software packages that began in the 1980s. Today's commercial solutions are manifest in a wide range of forms. Thus, for example, the major software vendor Esri, Inc. (Redlands, California) today sells a family of GI system products under the ArcGIS brand name to service the disparate needs of its diverse user base. At its core are three niche desktop systems: Basic, for viewing spatial data, creating layered maps, and performing rudimentary spatial

1.5.2 The Business Perspective

GI systems provide the underpinning technology for GI science, and many people play many roles in their development. Important activities range from software development to championing the importance of the spatial dimension in the activities of an organization through spatial data analysis. The business activities of many established companies are based

Table 1.4 Major events that shaped the development of GI systems.

The Era of Innovation	
1963	Canada Geographic Information System is developed by Roger Tomlinson and colleagues for Canadian Land Inventory. This project pioneers much technology and introduces the term <i>GIS</i> .
1964	The Harvard Laboratory for Computer Graphics and Spatial Analysis is established under the direction of Howard Fisher. SYMAP, the first raster system for automated cartography, is created by Harvard researchers in 1966.
1967	The U.S. Bureau of Census develops DIME-GBF (Dual Independent Map Encoding—Geographic Base File), a data structure and street-address database for the 1970 U.S. Census.
1969	ESRI (Environmental Systems Research Institute) Inc. formed by Jack Dangermond, previously at the Harvard Lab, and his wife Laura.
1969	M&S Computing (subsequently renamed Intergraph Corp.) formed by Jim Meadlock and four others, who worked on guidance systems for Saturn rockets.
1969	Publication of <i>Design with Nature</i> by Ian McHarg; introduces many of the basic concepts of geographic analytics, including the map overlay process (see Section 13.2.4).
1972	Landsat 1 launched—the first of many civilian remote-sensing satellites.
1973	First digitizing of maps by a national mapping agency in a production system (Ordnance Survey, Great Britain).
The Era of Commercialization	
1981	ESRI ArcInfo launched—the first major commercial GI system based on the vector data structure and a relational database.
1985	The Global Positioning System gradually becomes a major source of data for navigation, surveying, and mapping.
1986	MapInfo Corp.'s software develops into first major desktop GI system. It defines a new standard for GI systems, complementing earlier software.
1986	The British Broadcasting Corporation (BBC) launches the Domesday machine to mark the 900th anniversary of the original Domesday Survey. Based on a microcomputer system with some GI system functionality, it held data and information provided by a million volunteers. The 21,000 GI files comprised maps, millions of words of text, and photographs of all of Britain, all cross-referenced and accessible by location and theme.
1988	TIGER (Topologically Integrated Geographic Encoding and Referencing), a follow-up from DIME, is described by the U.S. Census Bureau. Low-cost TIGER data stimulate rapid growth in U.S. business GI systems.
1992	The 1.7 GB Digital Chart of the World, sponsored by the U.S. Defense Mapping Agency (now NGA), is the first integrated 1:1 million-scale database offering global coverage.
1994	Executive Order 12906, signed by President Clinton, leads to creation of U.S. National Spatial Data Infrastructure (NSDI), clearinghouses, and the Federal Geographic Data Committee (FGDC).
1994	OpenGIS (subsequently Open Geospatial) Consortium of vendors and users established to improve interoperability between software.
1995	Complete conversion of the 240,000 topographic maps in the Great Britain national map coverage at 1:1,250, 1:2,500 and 1:10,000 scales into digital form.
1996	Innovation of Internet GI system products by Autodesk, ESRI, Intergraph, and MapInfo.
1996	MapQuest Internet mapping service launched.
1999	New generation of commercial satellites launched with submeter resolution capability (e.g., IKONOS and Quickbird).
The Era of Openness and Pervasive Use	
2000	The United States ceases the deliberate degradation (or “selective availability”) of U.S. Global Positioning System (GPS) signals for national security purposes to encourage commercial and civilian applications using GPS technology.
2003	U.S. Federal e-government initiative provides “One-Stop” access to geospatial data and information (now part of geo.data.gov/geoportal/).

Table 1.4 (continued)

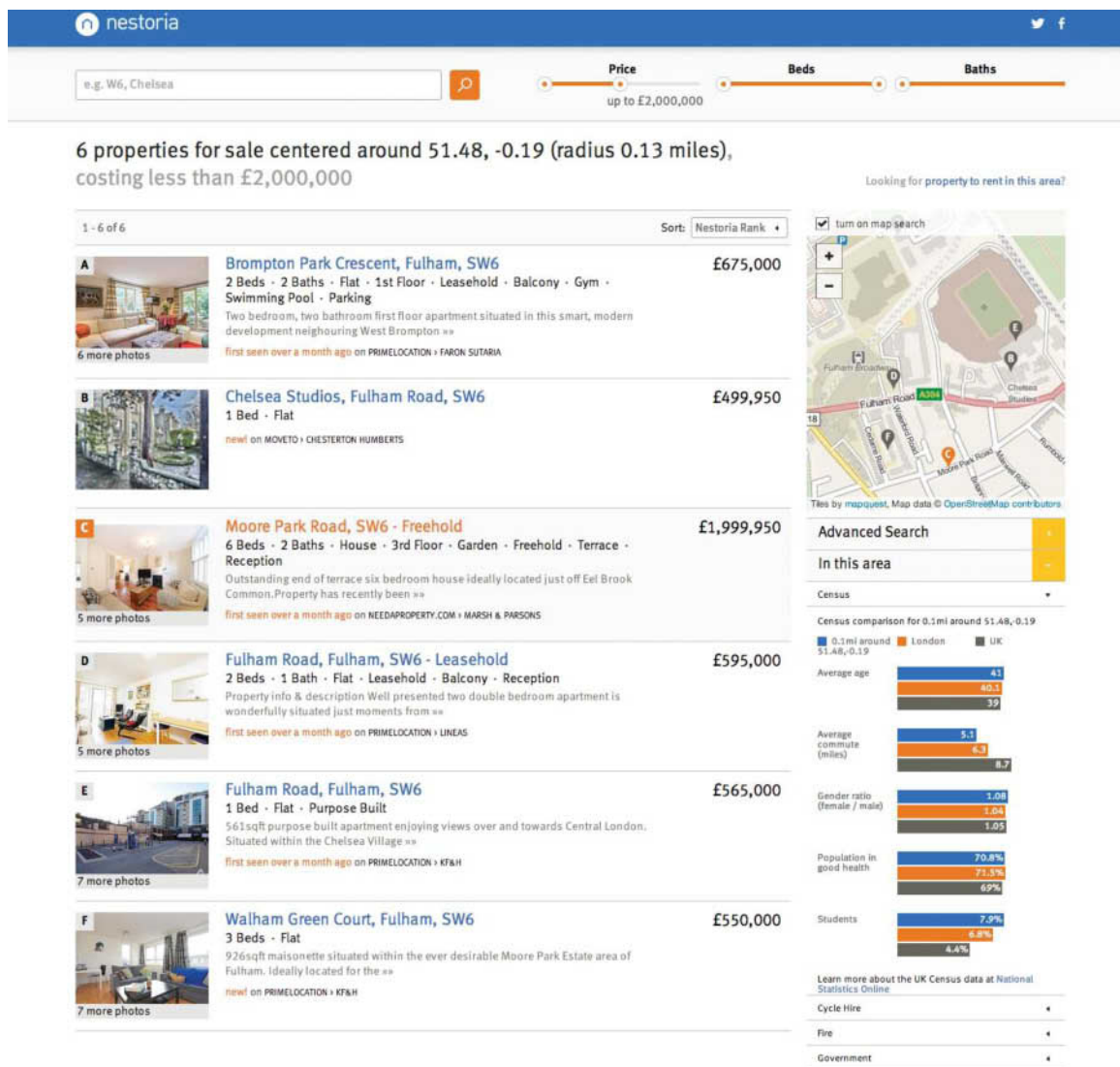
The Era of Openness and Pervasive Use	
2004	OpenStreetMap founded by Steve Coast to create citizen-enabled mapping for the UK. The OpenStreetMap Foundation is subsequently established to encourage the creation, development, and distribution of free geospatial data (notably through “mapping parties”) to provide geospatial data for anybody to use and share.
2004	Biggest GI system user in the world, National Imagery and Mapping Agency (NIMA), renamed National Geospatial-Intelligence Agency to signify emphasis on geo-intelligence.
2005	Launch of Google Earth service, the first major virtual 3-D globe with 150 million downloads in first 12 months.
2006	Launch of Amazon Web Services (AWS) as a “cloud,” a distributed utility computing service.
2007	Purchase by Nokia mobile phone company of NAVTEQ street data provider for \$8.1 billion.
2007	Launch of first touch-screen iPhone, including GPS-enabled mobile mapping with touch-screen functions.
2008	TeleAtlas street data provider purchased by TomTom for \$2.9 billion.
2009	Cost of RFID tags falls to the point that the “Internet of Things” becomes a widespread reality.
2009	Quantum GIS (QGIS; initially developed by Gary Sherman) launched. Developed through the Open Source Geospatial Foundation (OSGeo) and making extensive use of open-source plugins, this user-friendly multiple-operating-system software is available under public license.
2010	ESRI Inc. launches www.arcgis.com (tried in 2009 as www.arcgisonline.com), a Cloud-based GI platform offering a range of GI services.
2010	Advent of the UK Open Government Licence, allowing public bodies to publish a range of previously copyright material for public use. These are compatible with Creative Commons licences and are contributing to the wider dissemination of geographically referenced Open Data by governments around the world (see Section 17.4).
2011	Google launches indoor maps (floor plans) to enhance Google Maps service.

analysis; Standard, which offers more advanced tools for manipulating spatial databases; and Advanced, which has further additional facilities for data manipulation, editing, and analysis. Other vendors specialize in certain niche markets, such as the utility industry, or military and intelligence applications. The software industry employs several thousand programmers, software designers, systems analysts, application specialists, and sales staff, with backgrounds that include computer science, geography, and many other disciplines. Given the all-pervasive nature of GI systems and their linkage to other forms of software, it no longer makes much sense to guesstimate the global value of the GI system industry, but where such estimates are attempted, it seems clear that the industry continues to grow rapidly—for example, estimates of GI system sales in China, where GeoStar (www.geostar.com.cn/Eng/) is widely used, suggest that the value of the industry has increased 50-fold over the last decade. GI systems are a dynamic and evolving field, and their future continues to offer many exciting developments, some of which we will discuss in the final chapter.

Today a single GI system vendor offers many different products for distinct applications.

1.5.2.2 The Data Industry

The acquisition, creation, maintenance, dissemination, and sale of GI data also account for a huge volume of economic activity. Traditionally, a large proportion of GI data have been produced centrally by national mapping agencies, such as the U.S. Geological Survey (USGS) or Great Britain’s Ordnance Survey. In response to U.S. federal government policy, USGS supplies Open Data, defined as being priced at no more than the cost of reproduction of data. Elsewhere, some or all of available national mapping agency data are charged for to recoup the costs of their collection, although this is changing under various government Open Data initiatives. In Great Britain, for example, many Ordnance Survey mapping series have been made available free of charge since 2010; other, commercially valuable, data sets are not made available, however, because the government is unwilling to bear the projected \$45 million cost of losing future revenue streams for highly detailed mapping products used by corporations such as utility companies. The innovation of free-to-view mapping services such as Google Maps (maps.google.com) and Microsoft Bing Maps and Virtual Earth, with their business model based on advertising revenues, is having profound implications for the provision of



Source: Nestoria

Figure 1.15 The Nestoria Website (www.nestoria.com) presents a consolidated view of available property for sale or rent, alongside Open Data about neighborhoods. It is funded by the realtors who supply the linked listings and who “pay by click” for users searching for property.

map data (see Chapter 18). Specialist mapping sites, such as those used in the real-estate industry, present a consolidated map view of all available properties for sale or rent, along with Open Data pertaining to neighborhood quality (Figure 1.15). Volunteer-driven, open-source approaches to online cartography such as OpenStreetMap (www.openstreetmap.org) are also revolutionizing online cartography with their novel approach to map production and have stimulated the growth of commercial firms such as CloudMade (cloudmade.com).

Open Data are revolutionizing the business model for GI applications.

Geographically referenced attribute data relating to customer transactions are collected by many

organizations in pursuit of their day-to-day operational activities. Subject to legal data protection requirements, such data may be reused in pursuit of the tactical or strategic objectives of the organization, and many retailers, energy suppliers, and financial-service providers employ teams of analysts to devise better ways of servicing customers and increasing market share. Still other companies (including cell phone companies) fulfill the role of *value-added data resellers* of commercial and public-sector data. Important applications for such services are opening up, including improving public-service delivery. The GeoWeb (see Chapter 10) is creating fertile environments in which a very wide range of public- and private-sector data sources can be combined, analyzed, and displayed.

Private companies are now also licensed to collect fine-resolution data using satellites and to sell them to customers—as, for example, with GeoEye’s (geofuse.geoeye.com) IKONOS satellite data (see Table 1.4). Other companies collect similar data from aircraft. Still other companies specialize in the production of good-quality data on street networks, a basic requirement of many delivery companies. TomTom (www.tomtom.com) is an example of this industry, employing over 1800 staff in producing, maintaining, and marketing good-quality street-network data worldwide.

1.5.2.3 GI Services

As developments in the information economy gather still further momentum, many organizations are becoming focused on delivering integrated business solutions rather than raw or value-added data. The Cloud makes possible easy user access to data from sites that may be remote from locations where more specialized analysis and interpretation functions are performed. In these circumstances, it is no longer incumbent on an organization to manage either its own data or those that it buys from value-added resellers. For example, Esri Inc. offers a geographic data management service, in which data are managed and maintained for a range of clients that are at liberty to analyze them in quite separate locations. This may in time lead to greater vertical integration of the software and data industry—for example, Esri Inc. has developed a business information solutions division and acquired its own geodemographic system (called Tapestry) to service a range of business needs. This merging of the software and data industries comes together in the business of providing *GI systems as a service*, in which the Cloud is used as a remote platform on which software and data reside and GI system operations are performed.

1.5.3 The Government Perspective

Most government administration and decision-making activities have spatial implications, and governments—including the military and security services—remain by some margin the biggest users of commercial GI systems. Governments play an important role in many of the issues discussed throughout this book—including the development of data standards and interoperability, the creation of Web portals to disseminate GI (e.g., catalog.data.gov/dataset and data.gov.uk/), and the development of apps that automate key workflows such as zoning, property tax collection, or planning. Governments (such as those of Brazil and Spain) are also involved in the creation of open GI software.

One of the most profound changes concerning government and GI in recent years is in the ways in which government information is shared with citizens and external organizations in many parts of the world (see Chapter 17). Public-sector information (PSI) is defined by the wide range of information that government bodies collect, produce, reproduce, and disseminate across various areas of activity to fulfill their public-task functions. In terms of the nomenclature of Table 1.1, PSI is most conveniently thought of as information that has not been subjected to processing or any other manipulation beyond that necessary for its first use by government. Given the multiplicity of government functions, information that has been processed for one function (such as estimating local levels of homelessness or overcrowding) may be further refined and reused in the process of devising different measures for other functions (such as the composite measures of barriers to housing and public services used in various deprivation, or hardship, indices).

It is usual to think of PSI as having high fixed costs of assembly or collection and much smaller variable costs of copying and dissemination to users across the Web or through other media. It is also increasingly the case that the effort of rendering PSI into machine-readable or linked form can be absorbed into the fixed costs of information provision (see the star ratings schema discussed in Section 17.4.1). The small variable costs of disseminating some PSI have usually been absorbed in the United States, but this practice has been by no means universal, or even common, in other parts of the world. It is only in recent years that many governments worldwide have committed to greater availability of Open Data—which can be defined as data that can be used, reused, and redistributed freely by anyone, subject only at most to the requirement to attribute and share alike. Open Data are not necessarily free at the point of delivery, but any charge is usually no more than the cost of reproduction.

Open Data can be used, reused, and redistributed free of charge, subject only to Creative Commons or Open-Government licensing

The most obvious motivation for government supply of Open Data has been to improve transparency and accountability of government decision making, as well as to improve economic efficiency and to stimulate growth. It is often left to the private sector to spot opportunities for leveraging value and to lobby for new Open Data sources to be made available for the common good or commercial gain.

Open Data improve the transparency, accountability, and efficiency of decision making

The greater availability of PSI nevertheless raises a number of important issues for users of GI:

- Data are usually aggregated and cross tabulated to anonymize unit records, often using quite conservative procedures that mask important small-area variations.
- Government organizations are wary of the potential risk of *deanonymization*, that is, the combination of auxiliary data to link data to the individuals they characterize. GI systems provide the software environment in which such combination might take place
- Governments have traditionally emphasized data quality over timeliness, but are reluctant to extend quality-assurance procedures to rapidly delivered feeds of Open Data. There are likely to be issues with the provenance of Open Data as a consequence.
- Some of the core reference data produced by government provide the authoritative or definitive frameworks that are necessary to use other information—as, for example, with ZIP (post-) codes. Yet in many jurisdictions these are not open because they have considerable commercial value. There is considerable ongoing debate on whether government Open Data sources should build together into national (or international) information frameworks.

1.5.4 Computer-Science and Information-Science Perspectives

Computer scientists have been central to the development of GI technology, particularly in software development. Although sometimes mistakenly taken to be a branch of computer science, information science is a related but broad multidisciplinary area with a practical focus on the collection, classification, manipulation, storage, retrieval, analysis, movement, and dissemination of information. This brings with it concerns with database concepts, devising efficient algorithms for representing and accessing information, improving user interfaces, and developing computer architectures that are appropriate to different organizational settings. Like GI science and systems, these perspectives share concerns with the interactions between people, organizations, and computerized information systems.

All of this raises the question of whether GI is a fundamental part of computer science, or whether it merely presents a class of applications. The concepts underpinning spatiotemporal databases and computational geometry are core areas of computer science and information science that are directly relevant to GI technology, as are specific concerns with efficient computer processing, indexing schemes, database design, and computational geometry. Yet in each

case, the treatment of such issues is more focused on the technical aspects of GI technology than the context to which it is applied.

Context is important from a geographic perspective, defined by its concern with the human and environmental properties of the Earth's surface and near surface (Section 1.1), at scales from the architectural to the global. These concerns bring special focus to issues such as the nature of geographic data, their representation in GI systems, the ways in which they are georeferenced, and the uncertainties that arise when working with real-world places. They are not the priority of computer-science and information-science perspectives when these emphasize technology at the expense of context. Thus although GI science can be viewed as a branch of information science, the special nature of GI ensures that many of the principles of GI science have only a tenuous relationship to the broader principles of information science, and the same can be said about computer science.

In recent years much attention has been devoted to what is often termed *data science*, the issues arising from society's increasing dependence on data. It has been argued that science is increasingly data driven (sometimes called the *Fourth Paradigm* for research) and is based on the vast quantities of data that are now becoming available from ground-based and space-based sensors of various kinds, from social media, and from a host of other sources. From this perspective, some application domains produce such vast quantities of data that new and specialized techniques for storing, accessing, processing, and visualizing data are necessary to handle these enormous data problems. It has even been argued that such research neither needs nor produces theory, but instead mines data for patterns that may be useful in solving humanity's problems. We hear about the *exaflood*, the flood of information in quantities of exabytes (Table 1.2) and talk of science "drinking from a fire-hose." Data science studies the principles and techniques involved in managing these vast quantities of information, which include acquiring, sharing, documenting, managing, and archiving them.

Insofar as a substantial proportion of these data include information about location, it is clear that the issues of GI science have significant overlap with those of data science. Yet we maintain that the special characteristics of GI demand that many of these issues be treated separately and that education in the generic principles of data science is not necessarily sufficient qualification. Moreover, the kinds of infrastructure concerns that are often raised to prominence in data science, ranging from systematic documentation of data (metadata) to archiving and sharing, have a long history of development in GI science and may already be more advanced within this

specialized domain than in generic data science (see also Chapters 17 and 19).

1.5.5. The Geography Perspective

Geographers have long agonized over the content, coherence, and relevance of their discipline, and set against this background, it is perhaps surprising that they have nonetheless made time to investigate and innovate in the use of new research methods in their discipline. Throughout the period documented in Table 1.4, GI science and systems have developed and maintained a special relationship to the academic discipline of geography and other disciplines that deal with the Earth's surface, including geodesy, landscape architecture, planning, and surveying. Yet any special relationship is multifaceted, nuanced, and sometimes tense. This is even more so, given that the roots to much of geography lie in mapping for warfare, that the military remain heavy users of GI systems, and that commercial software applications are rarely open to full academic scrutiny in deference to business priorities.

The idiographic and nomothetic concerns of geography (Section 1.3) have recently been the focus of thinking about the representation of *place* and *place effects*. Place is a social construct of space by humans and is key to the way that they understand their surroundings. Thus, although space exists independent of the existence of people, it is human interactions and experiences that crystallize space into place through shared perceptions and recognition. Figure 1.16 shows one delineation of the place

known as “Paris” (France) through mapping the extent of attribution of this label to photographs uploaded to Flickr, along with the more limited extents of other more localized places identified by users.

Geography also tells us that widely scattered places (for example, in Washington, California, Florida, or Vermont) share important social or physical similarities that are manifest in different ways and to differing degrees. GI systems provide a way of representing these similarities, using standardized quantitative measures and, hence, the repetition of place effects across space. Geodemographic classifications provide one prominent example of the assignment of dispersed locations to place-relevant typologies, independent of their locational proximity to one another (see Figure 1.17).

In both of these examples, absolute location is not the principal focus, but it is the relative locations of places and the connectivity and relations between them that are important.

Recent years have seen the popularization of the term *neogeography* to describe developments in Web mapping technology and spatial data infrastructures that have greatly enhanced our abilities to assemble, share, and interact with geographic information online. Allied to this is the increased crowd sourcing by online communities of *volunteered geographic information* (VGI), discussed in Section 1.5.6 below. Neogeography is founded on the two-way, many-to-many interactions between users and Web sites that have emerged under Web 2.0, as embodied in projects such as Wikimapia (www.wikimapia.org) and OpenStreetMap

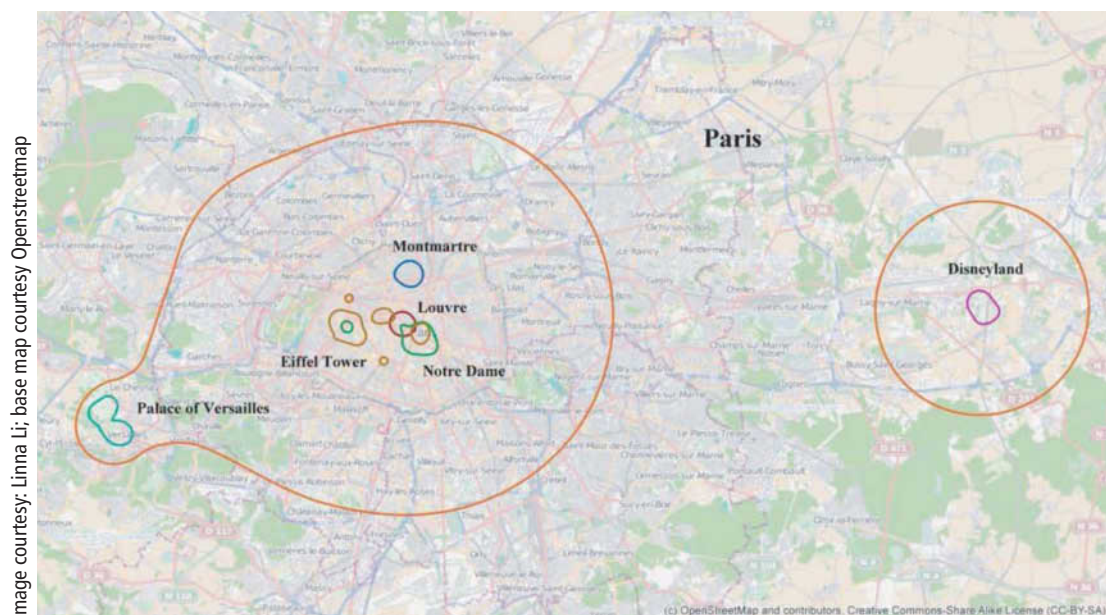


Figure 1.16 Delineation of the extent of Paris, France, and locations associated with it, based on the density of geotagged Flickr photos for each place name. Contours of the place surfaces are depicted by choosing a threshold value of point density visually (e.g., the threshold value for the contour of Louvre is 500 points per square kilometer).

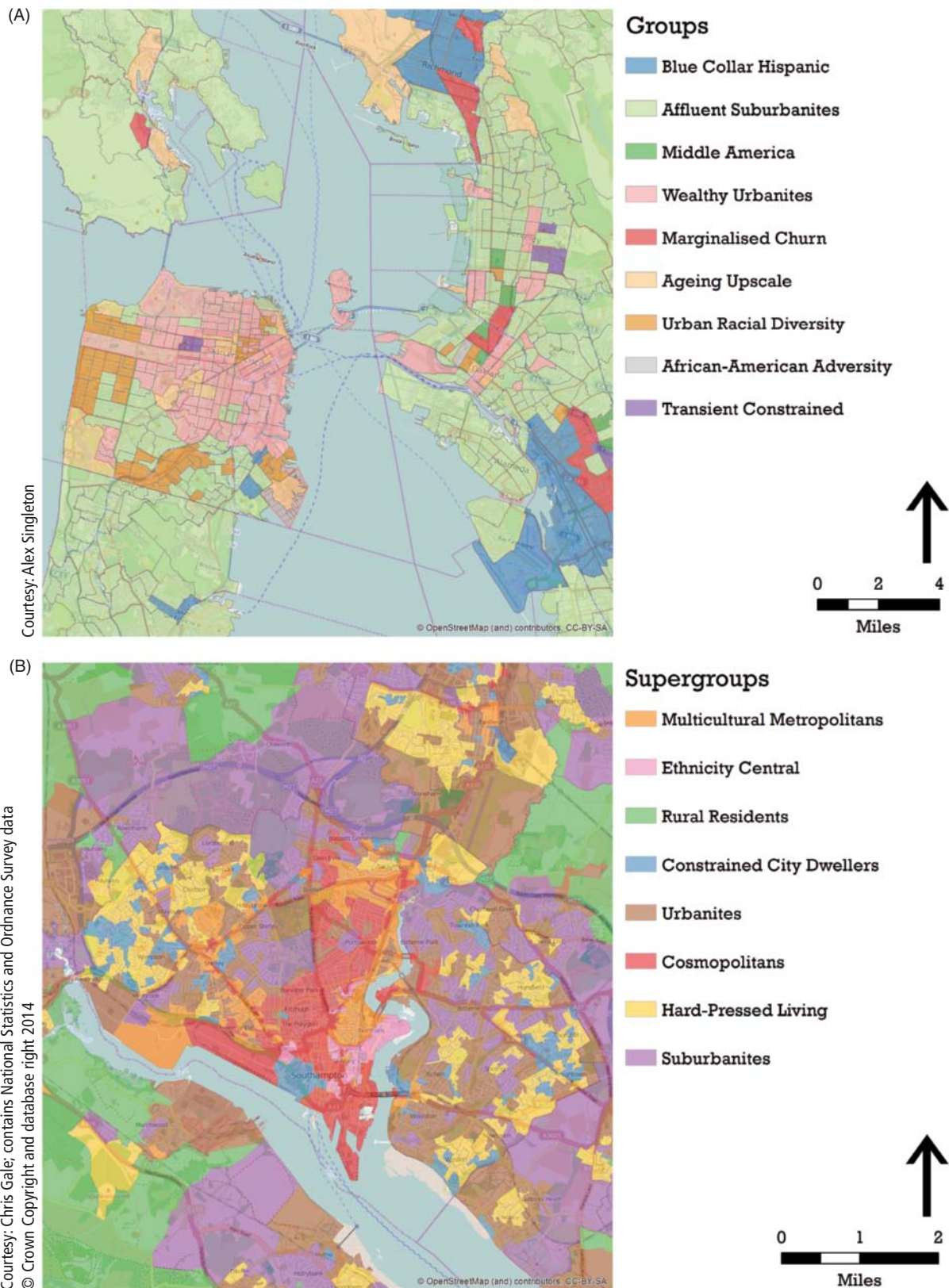


Figure 1.17 Geodemographic neighborhood classifications make it possible to quantify how similar one place is to another, subject to the availability of consistent data across a jurisdiction. (A) A classification based on 2010 U.S. Census data, showing part of San Francisco; and (B) the UK 2011 Output Area Classification, showing part of Southampton.

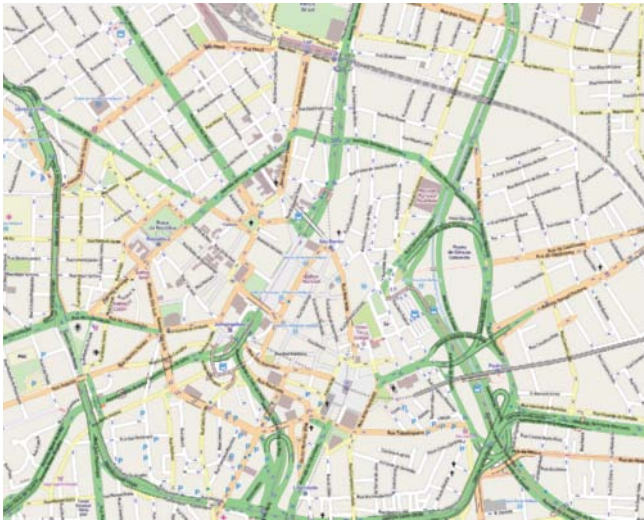


Figure 1.18 A crowd-sourced street map of part of São Paulo, Brazil.

(www.openstreetmap.org). Today, Wikimapia contains user-generated entries for more places than are available in any official list of place-names, whereas OpenStreetMap is well on the way to creating a free-to-use global map database through assimilation of digitized satellite photographs with GPS tracks supplied by volunteers (see Figure 1.18). This has converted many new users to the benefits of creating, sharing, and using geographic information, often through ad hoc collectives and interest groups. The creation, maintenance, and distribution of databases is no less than a “wikification of GI systems.” Neogeography brings GI systems and some uses of spatial data infrastructures to the masses, while also presenting new challenges to citizen privacy and confidentiality. The empowerment of many nonexpert GI system users also brings with it the new challenges of ensuring that tools are used efficiently, effectively, and safely (see Chapter 17) and reemphasizes that technology can never offer more than a partial solution to the effective deployment of GI systems.

A core motivation for GI science is to better understand why GI systems applications are successful in practice. In this way, it should be possible to develop still better research practices for the future. Yet the applied nature of these endeavors and their core concerns with data, technology, and procedures inevitably distances any application from the investigator’s direct experience. For some, this redefines the subject in ways that are inherently unacceptable in their reductionism; and no engagement with ever-more immersive technologies, richer data, or more sensitive analytical method can accommodate what are seen as failures inherent in empiricist methodology. In recent years protagonists of “critical GIS” have claimed

some success in illustrating how better technology, data, and methods make it possible to go beyond simplistic depiction of raw data. Yet for others, the deployment of computer systems inherently implies greater conviction than is warranted about the attributes and behaviors of the subjects of research. If this is the case, the “critical” prefix can never be more than a sop that implies, but does not deliver, acceptable levels of humility when engaging with the subjects of research.

Some geographers remain suspicious of the use of GI systems in geography.

Still broader issues arise when research is sponsored, directly or indirectly, by vested interests of commerce or the military. The latter was thrown into sharp focus in 2013 in Joel Wainwright’s *Geopiracy*—a critique of recent U.S. military funding of academic geography to fund expeditions to map the human terrain of other countries. The Bowman expeditions—named after Isaiah Bowman, an early Twentieth Century American geographer identified with both empiricism and empire building—attracted criticism from Mexican authorities for alleged nondisclosure of U.S. military funding. To some, this was an uncomfortable reminder of the role of mapping as a tool of warfare (see Box 11.1) and the historic role of military motivations in the development of geography. We discuss this further in Chapter 18.

1.5.6 The Societal Perspective

The history of the development of GI systems has been one in which a technology of problem solving has evolved from specialized (and expensive) standalone configurations of hardware, software, and data to a background technology that is very much part of the information-technology mainstream. Today, when most of us casually use computers to pose “Where?” questions, we are rarely if ever cognizant that we are “doing GIS.” We rarely contemplate the costs that an organization, somewhere, incurs in providing the GI services that we consume—because the costs of using the geographically enabled search of online mapping systems, for example, are absorbed through the advertising and data-sharing revenues of corporations such as Google, Apple, or Microsoft, or the costs of real-time public-transit information systems are borne by the operator out of ticket revenues. Internet behemoths foist use of cookies on us as users that betray our locations and many characteristics of our online identities, and we may resent the sophistication of the resulting advertising that is targeted on us. It seems that opting out of default options of disclosure requires greater diligence and doggedness than in the pre-Internet era, and we discuss privacy and legal concerns in Chapter 17. But most of us accept this intrusiveness, most of the time,

The Top-Level Categories of the Geographic Information Science and Technology Body of Knowledge

Following are the 73 top-level categories of knowledge in the first edition (2006), and pointers (in brackets) to related chapters in this book where appropriate.

Knowledge Area: Analytical Methods (AM)

- Unit AM1 Academic and analytical origins (13)
- Unit AM2 Query operations and query languages (13 and 14)
- Unit AM3 Geometric measures (13 and 14)
- Unit AM4 Basic analytical operations (13 and 14)
- Unit AM5 Basic analytical methods (13 and 14)
- Unit AM6 Analysis of surfaces (14)
- Unit AM7 Spatial statistics (13)
- Unit AM8 Geostatistics (13)
- Unit AM9 Spatial regression and econometrics (13 and 14)
- Unit AM10 Data mining (13 and 14)
- Unit AM11 Network analysis (14)
- Unit AM12 Optimization and location-allocation modeling (14)

Knowledge Area: Conceptual Foundations (CF)

- Unit CF1 Philosophical foundations (1)
- Unit CF2 Cognitive and social foundations (1)
- Unit CF3 Domains of geographic information (1)
- Unit CF4 Elements of geographic information (2 and 3)
- Unit CF5 Relationships (2 and 3)
- Unit CF6 Imperfections in geographic information (5)

Knowledge Area: Cartography and Visualization (CV)

- Unit CV1 History and trends (11 and 12)
- Unit CV2 Data considerations (11 and 12)
- Unit CV3 Principles of map design (11)
- Unit CV4 Graphic representation techniques (11 and 12)
- Unit CV5 Map production (11)
- Unit CV6 Map use and evaluation (11)

Knowledge Area: Design Aspects (DA)

- Unit DA1 The scope of GI system design (6)
- Unit DA2 Project definition (16)
- Unit DA3 Resource planning (16)
- Unit DA4 Database design (9)
- Unit DA5 Analysis design (13 and 14)
- Unit DA6 Application design (1)
- Unit DA7 System implementation (16)

Knowledge Area: Data Modeling (DM)

- Unit DM1 Basic storage and retrieval structures (7)
- Unit DM2 Database management systems (6)
- Unit DM3 Tessellation data models (7)
- Unit DM4 Vector and object data models (7)
- Unit DM5 Modeling 3-D, temporal, and uncertain phenomena (7)

Knowledge Area: Data Manipulation (DN)

- Unit DN1 Representation transformation (7 and 8)
- Unit DN2 Generalization and aggregation (11 and 12)
- Unit DN3 Transaction management of geospatial data (8)

Knowledge Area: Geocomputation (GC)

- Unit GC1 Emergence of geocomputation (15)
- Unit GC2 Computational aspects and neurocomputing (15)
- Unit GC3 Cellular Automata (CA) models (15)
- Unit GC4 Heuristics (14)
- Unit GC5 Genetic algorithms (GA) (15)
- Unit GC6 Agent-based models (15)
- Unit GC7 Simulation modeling (15)
- Unit GC8 Uncertainty (5)
- Unit GC9 Fuzzy sets (5)

Knowledge Area: Geospatial Data (GD)

- Unit GD1 Earth geometry (4)
- Unit GD2 Land partitioning systems (4)



Unit GD3 Georeferencing systems (4)

Unit GD4 Datums (4)

Unit GD5 Map projections (4)

Unit GD6 Data quality (5)

Unit GD7 Land surveying and GPS (4)

Unit GD8 Digitizing (8)

Unit GD9 Field data collection (2, 8, 17)

Unit GD10 Aerial imaging and photogrammetry (8)

Unit GD11 Satellite and shipboard remote sensing (8)

Unit GD12 Metadata, standards, and infrastructures (10)

Knowledge Area: GIS&T and Society (GS)

Unit GS1 Legal aspects (18)

Unit GS2 Economic aspects (17)

Unit GS3 Use of geospatial information in the public sector (17)

Unit GS4 Geospatial information as property (17)

Unit GS5 Dissemination of geospatial information (17)

Unit GS6 Ethical aspects of geospatial information and technology (18)

Unit GS7 Critical GI systems (1)

Knowledge Area: Organizational and Institutional Aspects (OI)

Unit OI1 Origins of GIS&T (1)

Unit OI2 Managing GIS operations and infrastructure (16)

Unit OI3 Organizational structures and procedures (16)

Unit OI4 GIS&T workforce themes (1)

Unit OI5 Institutional and interinstitutional aspects (18)

Unit OI6 Coordinating organizations (national and international) (18)

either in ignorance or in implicit recognition of the convenience of the search engine and other services that such companies provide. Disclosure of online identity (of which location is an integral part) helps us to deal with the tyranny of small decisions in everyday life, even if GI service providers then try to use this to shape aspects of our identities as consumers in the interests of their advertising clients.

Choosing whether or not to use a particular search engine is an individual decision, but other GI applications are fundamentally underpinned by collective attitudes to creation and use. VGI is the collective harnessing of tools to create, assemble, and disseminate geographic data provided voluntarily by individuals—such as street locations (e.g., www.openstreetmap.org), geotagged photographs (e.g., www.flickr.com), or locationally referenced restaurant reviews. The motivations for supply of such information are multifaceted and may ebb and flow over time—recent years have seen much talk of “Wiki fatigue”—but all such projects are dependent on funding models for creating and maintaining GI systems—for example through individual donations or corporate sponsorship.

The funding imperative holds in all other societal deployments of GI. Governments pay the cost of ensuring the availability of Open Data and public-sector information (PSI). The costs of creating open software are often borne by research-grant-awarding

organizations or funding by universities using student-fee income (justified as enhancing the public-facing profile of the organization). We discuss these different funding models in greater detail in Chapter 18.

The diversity of motivations underpinning the creation, maintenance, and dissemination of GI and the software to analyze it leads to the following societal concerns (see also the contribution of Sarah Elwood, discussed in Box 10.4):

- The links between knowledge and power. The ways in which GI systems represent the Earth’s surface, and particularly human society, have been observed to privilege certain people, phenomena, and perspectives, at the expense of others. Minority views, and the views of individuals, can be submerged in this process, as can information that differs from the official or consensus view. GI systems often force knowledge into forms that are more likely to reflect the view of the majority, or the official view of government, and as a result *marginalize* the opinions of minorities or the less powerful. Countering this outlook is the view that greater access to PSI (see Section 17.4), especially over the Web, has enlivened debate and has gone some considerable way toward leveling the playing field in terms of data access. Moreover, the near-ubiquitous availability of mapping services and

virtual Earths with limited GI system functionality (e.g., maps.google.com, apple.com/ios/maps/, and www.microsoft.com/virtualearth) brings rudimentary mapping and analysis capabilities to almost anyone with an Internet connection.

- In principle it is possible to use GI systems for any purpose, and so in practice they may sometimes be used for purposes that may be ethically questionable or may invade individual privacy—such as surveillance and the gathering of military and industrial intelligence. The technology may appear neutral, but it is always used in a social context. Although most would agree that the consequences of misuse of GI systems are not cataclysmic, there are nevertheless some similarities with the debates over the atomic bomb in the 1940s and 1950s: the scientists who develop and promote the use of GI systems surely bear some responsibility for how they are eventually used to the net benefit of humankind. The idea that a tool can be inherently neutral, and its developers therefore immune from any ethical debates, is now strongly questioned.
- The very success of GI systems is a cause of concern. There are qualms about a field that appears to be led by technology and the marketplace, rather than by human need. There are fears that GI systems have become too successful in modeling socioeconomic distributions and that as a consequence GI systems have become a tool of the “surveillance society.” And there are fears that many users of services such as social media do not understand the uses to which their data may be put and the ways that they may be linked to other sources. Geography provides a powerful means of linking rich data sources pertaining to individual behavior.
- There remains an underrepresentation of applications of GI systems in *critical* research that focuses on the connections between human agency and particular social structures and contexts. Some detractors of GI systems have also suggested that such connections are not amenable to digital representation in whole or in part. Applications of “critical GIS” (see Section 1.5.5) have been developed that investigate, for example, the compatibility between GI systems and feminist epistemologies and politics; yet for some, the “critical” prefix does not address fundamental concerns with empiricism and the use of technologies to analyze the subjects of critical research.
- Some view the association of GI systems with the scientific and technical project as fundamentally flawed. More narrowly, there is a view that GI systems are inextricably bound to the philosophy and assumptions of the approach to science known

as *logical positivism* (see also the reference to “positive” in Section 1.1.1). As such, the argument goes, GI systems can never be more than a positivist tool and a normative instrument and cannot enrich other more critical perspectives in social science and even core areas of science itself. This is a criticism not just of GI systems, but also of the application of the scientific method to the analysis of social systems.

1.6 GI Science and Spatial Thinking

GI systems are useful tools, helping everyone from scientists to citizens to solve geographic problems. But like many other kinds of tools, such as computers themselves, their use raises questions that are sometimes frustrating and sometimes profound. For example, how does a GI system user know that the results obtained are accurate? What principles might help a GI system user to design better maps? How can location-based services be used to help users to navigate and understand human and natural environments? Some of these questions concern GI system design, and others are about the nature of GI and the methods used to analyze it. Taken together, we can think of them as questions that arise from the use of GI systems—that are stimulated by exposure to GI systems in their many guises. Many of them are addressed in detail in this book, and the book’s title emphasizes the enduring importance of the science of problem solving.

Science is a systematic enterprise that builds and organizes knowledge in the form of testable explanations and predictions about the world (and beyond). It also refers to a body of knowledge itself, of the type that can be rationally explained and reliably applied. The U.S. University Consortium for Geographic Information Science (www.ucgis.org) is an organization of roughly 70 research universities that engages in research agenda setting, lobbying for research funding, and related activities. In 2006 it developed and published *Geographic Information Science and Technology Body of Knowledge* (BoK) in collaboration with the Association of American Geographers (a revised second edition is under development). The BoK is intended to be a comprehensive summary of knowledge in the field, to be used to evaluate the content of courses and programs, to set standards for job descriptions, and many other uses.

Box 1.5 lists the top-level contents of the 2006 first edition, which is currently available from www.aag.org/galleries/publications-files/GIST_Body_of_Knowledge.pdf. Each of the top-level topics is elaborated in detail in the document.

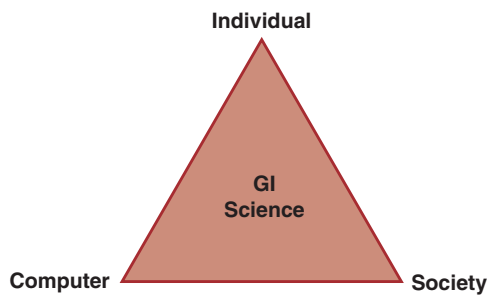


Figure 1.19 The remit of GI science, according to Project Varenus (www.ncgia.ucsb.edu/varenus/varenus.html).

One disarmingly simple way of viewing this taxonomy is in the framework of the Varenus project (www.ncgia.ucsb.edu/varenus/varenus.html; Figure 1.19). Here, GI science is viewed as anchored by three concepts: the individual, the computer, and society. These form the vertices of a triangle, and GI science lies at its core. The various terms that are used to describe GI-science activity can be used to populate this triangle. Thus research about the individual is dominated by cognitive science, with its concern for understanding of spatial concepts, learning and reasoning about geographic data, and interaction with the computer. Central to such research is the desire to engineer more readily intelligible user interfaces, in the interest of improved human–computer interaction. Allied to this is the desire to improve geovisualization and render intelligible the results of spatiotemporal analysis. Research about the computer is dominated by issues of representation, the adaptation of new technologies, computation, and visualization. And finally, research about society addresses issues of impact and societal context.

1.7 GI Systems and Science in Society

GI science is often about using the software environment of GI systems to redefine, reshape, and resolve problems that are even older than the first GI systems. The need for methods of spatial analysis, for example, dates from the first maps, and many methods were developed long before the first GI system appeared on the scene in the mid-1960s. Another way to look at GI science is to see it as the body of knowledge that GI systems implement and exploit. Map projections (Chapter 4), for example, are part of GI science and are used and transformed through use of GI systems. A further area of great importance to GI systems is cognitive science, particularly the scientific

understanding of how people think about their geographic surroundings. If GI systems are to be easy to use, they must fit with human ideas about such topics as driving directions or how to construct useful and understandable maps.

Many roots to GI systems can be traced to the spatial analysis tradition in the discipline of geography.

Yet if there are enduring qualities to the body of knowledge that constitutes GI science, so there is also profound change in the setting in which it is applied. Astonishing improvements in our technical capacity to create and manipulate digital information have profoundly changed the ways in which GI systems are understood and used, both in business and in government. But perhaps the most profound changes of all have taken place in society at large. Just as a GI “system” is no longer a readily identifiable and clearly bounded “thing,” so the notion of a “GI system user” has become a similarly anomalous conception—for the simple reason that we are now all consumers of GI and applications that use it. Other changes characterize the environment in which GI science is applied, given our vastly improved abilities to identify and monitor the movements of objects and (consenting) individuals, the rise of citizen science, and the availability of vast arrays of data that make it possible to devise rich yet comparable depictions of unique places.

We wonder where this will all end. Resolving the “Where?” question places unending demands on society’s evolving information infrastructures and, in turn, begs many more questions concerning the sourcing, assembly, maintenance, and deployment of GI. How might the quality and reliability of volunteered geographic information be ascertained? How can users best understand the standards by which the information that they rely on has been created? How might user interfaces be improved to render GI intelligible to all relevant users? What are the obligations of governments to supply Open Data that are timely, comprehensive, and safe to use? What issues arise when commercial spatial data infrastructures (such as Google Earth) are reused for scientific or policy purposes? And what rights does the citizen have to privacy, or to be forgotten?

These are some of the key questions of GI science that we address in this book. They have profound implications for the ways in which we interact with one another and the ways in which we interact with places. For billions of casual users, GI systems have become an integral part of the information technology mainstream, yet this begs important questions about the roles, functioning, and status of pervasive GI technologies in society.

Although we continue to refer to “GI systems,” an emphasis throughout much of this book is on this new societal vision of the emerging globally connected “GI system.” As the software and procedures that are used to manipulate and analyze data continue to evolve, GISS provides focus to the way that better system design advances science and vice versa. Much of this book is about the more enduring principles of GI science, which are applied through society’s use of geographic information, using the standards and procedures that are integral to GI system design.

The title of the first three editions of this book, which appeared between 2001 and 2011, was

Geographic Information Systems and Science. Our revision of the title, as well as the content, of this fourth edition is designed to emphasize the enduring importance of scientific principles against a background of improvements in GI system design. Over this period, GI systems have evolved from isolated, monolithic software to society’s resource of choice when the question “Where?” arises in any of its many guises. We believe strongly that effective users of GI systems require some awareness of *all* aspects of geographic information, from the basic principles and techniques to concepts of management and familiarity with applications. We hope this book provides that kind of awareness.

Questions for Further Study

1. Examine the geographic data available for the area within 80 km (50 miles) of either where you live or where you study. Use it to produce a short (2500-word) illustrated profile of either the socio-economic or the physical environment. (See, for example, www.data.gov/ or data.gov.uk/data)
2. What are the distinguishing characteristics of the scientific method? Discuss the relevance of each to the history of GI systems.
3. Has society-wide use of GI systems transformed the research agenda for GI science? Give reasons for your answer.
4. Locate a subset of the issues identified in Box 1.5 in two triangular “GI science” diagrams like that shown in Figure 1.17—one for themes that predominantly relate to developments in science and one for themes that predominantly relate to developments in technology. Further details on each of the themes is available in the online *BoK*. If time is available, give short written reasons for your assignments. Compare the distribution of issues within each of your triangles to assess the relative importance of the individual, the computer, and society in the development of GI science and of GI technologies.

Further Reading

De Smith, M., Goodchild, M. F., and Longley, P. A. 2009. *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools* (3rd ed.). Leicester: Troubador and www.spatialanalysisonline.com.

DiBiase, D., deMers, M., Johnson, A., Kemp, K., Taylor Luck, A., Plewe, B. and Wentz, E. (eds.). 2006. *Geographic Information Science & Technology Body of Knowledge*. Washington, DC: Association of American Geographers. www.aag.org/galleries/publications-files/GIST_Body_of_Knowledge.pdf

Foresman, T. W. (ed.). 1998. *The History of Geographic Information Systems: Perspectives from the Pioneers*. Upper Saddle River, NJ: Prentice Hall.

Goodchild, M. F. 1992. Geographical information science. *International Journal of Geographical Information Systems* 6: 31–45.

Wainwright, J. 2013. *Geopiracy: Oaxaca, Militant Empiricism, and Geographical Thought*. New York: Palgrave Macmillan.