Geographic Information: Science, Systems, and Society

LEARNING OBJECTIVES

This chapter sets the conceptual framework for and summarizes the content of the book by addressing several major questions:

- What exactly is geographic information (GI), and why is it important? What is special about it?
- What new technological developments are changing the world of GI?
- How do GI systems affect the lives of average citizens?
- What kinds of decisions make use of geographic information?
- What is a geographic information system (GI system), and how would you recognize one?
- What is geographic information science (GI science), and why is it important to GI systems?
- How do scientists and governments use GI systems, and why do they find them helpful?
- How do companies make money from GI systems?

After studying this chapter you will:

- Know definitions of many of the terms used throughout the book.
- Be familiar with a brief history of GI science and GI systems.
- Recognize the sometimes invisible roles of GI systems in everyday life, business, and government.
- Understand the significance of GI science and how it relates to GI systems.
- Understand the many impacts that GI systems and its underpinning science are having on society and the need to study those impacts.

1.1 Introduction: What Are GI Science and Systems, and Why Do They Matter?

Almost everything that happens, happens somewhere. We humans confine our activities largely to the surface and near-surface of the Earth. We travel over it and through the lower levels of its atmosphere, and we go through tunnels dug just below the surface. We dig ditches and bury pipelines and cables, construct mines to get at mineral deposits, and drill wells to access oil and gas. We reside on the Earth and interact with others through work, leisure, and family pursuits. Keeping track of all this activity is important, and knowing where it occurs can be the most convenient basis for tracking. Knowing where something happens is of critical importance if we want to go there ourselves or send someone there, to find more information about the same place, or to inform people who live nearby. In addition, geography shapes the range of options that we have to address things that happen, and once they are made, decisions have geographic consequences. For example, deciding the route of a new high-speed railroad may be shaped by topographic and environmental considerations, and the chosen route will create geographic winners and losers in terms of access. Therefore geographic location is an important component of activities, policies, strategies, and plans.

> **Almost everything that happens, happens somewhere. Knowing where something happens can be critically important.**

The focus of this book is on geographic information, that is, information that records *where* as well as *what* and perhaps also *when*. We use the abbreviation *GI* throughout the book. GI systems were originally conceived as something separate from the world they represent—a special kind of information system, often located on a user's desk, dedicated to performing special kinds of operations related to location. But today such information pervades the Internet, can be accessed by our smartphones and other personal devices, and is fundamental to the services provided by governments, corporations, and even individuals. Locations are routinely attached to health records, to Twitter feeds and photographs uploaded to Flickr, and to the movements of mobile phone users and vehicles. In a sense, then, the whole digital world has become one vast, interconnected GI system. This book builds on what users of this system already know—that use of GI services is integral to many of our interactions through the Internet. Later chapters will describe, for example, how storage and management of more and more data entail use of the Cloud, how Big Data and Open Data have become ubiquitous (but not necessarily useful), and how Web-based GI systems have become a fact of life.

Underlying these changes are certain fundamentals, however, and these have a way of persisting despite advances in technology. We describe them with the term *GI science*, which we define as the general knowledge and important discoveries that have made GI systems possible. GI science provides the structure for this book because as educators we believe that knowledge of principles and fundamentals—knowledge that will still be valid many years from now—is more important than knowledge of the technical details of today's versions of GI technology. We use the acronym *GISS*—geographic information science and systems—at various points in this book to acknowledge the interdependence between the underpinning science and the technology of problem solving.

At the outset, we also observe that GI science is also fundamentally concerned with solving applied problems in a world where business practices, or the realpolitik of government decision making, are important considerations. We also discuss the practices of science and social science that, although governed by clearly defined scientific principles, are imperfectly coupled in some fast-developing areas of citizen science.

1.1.1 The Importance of Location

Because location is so important, it is an issue in many of the problems society must solve. Some of these problems are so routine that we almost fail to notice them—the daily question of which route to take to and from work, for example. Others are quite extraordinary and require rapid, concerted, and coordinated responses by a wide range of individuals and organizations—such as responding to the major emergencies created by hurricanes or earthquakes (see Box 1.1). Virtually all aspects of human life involve location. Environmental and social scientists recognize the importance of recording location when collecting data; major information companies such as Google recognize the importance of providing mapping and driving directions and prioritizing searches based on the user's location; and citizens are increasingly familiar with services that map the current positions of their friends. Here are some examples of major decisions that have a strong geographic element and require GI:

- Health-care managers decide where to locate new clinics and hospitals.
- Online shopping companies decide the routes and schedules of their vehicles, often on a daily basis.
- Transportation authorities select routes for new highways and anticipate their impacts.
- Retailers assess the performance of their outlets and recommend how to expand or rationalize store networks.
- Forestry companies determine how best to manage forests, where to cut trees, where to locate roads, and where to plant new trees.
- National park authorities schedule recreational path creation, maintenance, and improvement (Figure 1.1).
- Governments decide how to allocate funds for building sea defenses.
- Travelers and tourists give and receive driving directions, select hotels in unfamiliar cities, and find their way around theme parks (Figure 1.2).
- Farmers employ new GI technology to make better decisions about the amounts of fertilizer and pesticides to apply to different parts of their fields.

If location and GI are important to the solution of so many problems, what distinguishes those problems from each other? Here are three bases for classifying problems. First, there is the question of *scale*, or level of geographic detail. The architectural design of a building involves GI, but only at a very detailed or local scale. The information needed to service the building is also local—the size and shape of the

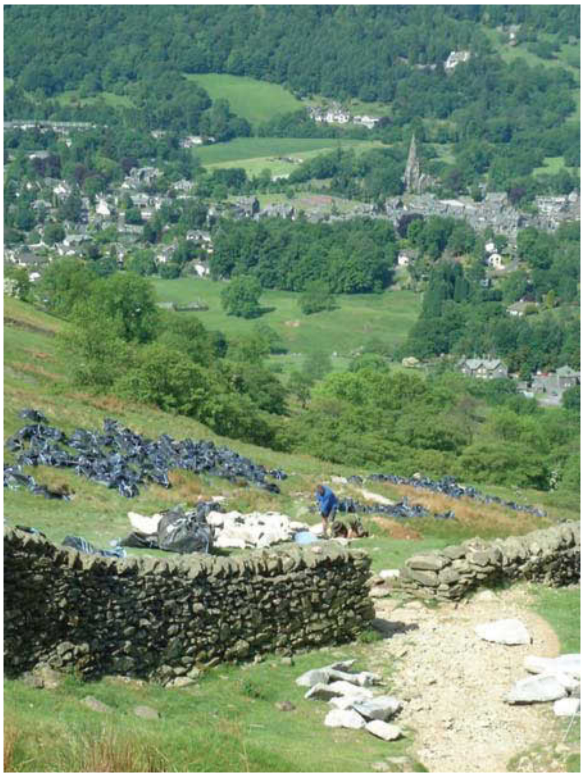


Figure 1.1 Maintaining and improving footpaths in national parks is a geographic problem.

parcel, the vertical and subterranean extent of the building, the slope of the land, and its accessibility using normal and emergency infrastructure. At the other end of the scale range, the global diffusion of epidemics and the propagation of tsunamis across the Pacific Ocean (Box 1.1) are phenomena at a much broader and coarser scale.

Scale or level of geographic detail is an essential property of any project.

Second, problems can be distinguished on the basis of *intent*, or *purpose*. Some problems are strictly practical in nature—they must often be solved as quickly as possible and at minimum cost to achieve such practical objectives as saving lives in an emergency, avoiding fines by regulators, or responding to civil disorder. Others are better characterized as driven by human curiosity. When GI is used to verify the theory of continental drift, to map distributions of glacial deposits, or to analyze the historic movements of people in anthropological or biosocial research (see Box 1.2 and Figure 1.5), there is no sense of an immediate problem that needs to be solved. Rather, the intent is to advance human understanding of the world, which we often recognize as the intent of science.

Although science and practical problem solving can be thought of as distinct human activities, it is often argued that there is no longer any effective distinction between their methods. Many of the tools and methods used by a retail analyst seeking a site for a new store are essentially the same as those used by a scientist in a government agency to ensure the protection of an endangered species, or a transport planner trying to ameliorate peak-hour traffic congestion in a city. Each requires the most accurate measurement devices, employs terms whose meanings have been widely shared and agreed on, produces results that are replicable by others, and in general follows all the principles of science that have evolved over the past centuries. The knowledge-exchange activities carried out between research organizations and the government and business sectors can be used to apply many of the results of curiosity-driven science to the practical world of problem solving.

Figure 1.2 Navigating tourist destinations is a geographic problem.

The use of GI systems in support of science, routine application, and knowledge exchange reinforces the idea that science and practical problem solving are no longer distinct in their methods, as we will discuss later. As a consequence, GI systems are used widely in all kinds of organizations, from academic institutions to government agencies, not-for-profit organizations, and corporations. The use of similar tools and methods across so much of science and problem solving is part of a shift from the pursuit of curiosity within traditional academic disciplines to solution-centered, interdisciplinary teamwork.

Nevertheless, in this book we distinguish between uses of GI systems that focus on applications such as inventory or resource management, or so-called normative uses, and uses that advance science, or so-called positive uses (a rather confusing meaning of that term, unfortunately, but the one commonly used by philosophers of science—its use implies that science confirms theories by finding positive evidence in support of them and rejects theories when negative

Applications Box 1.1

The 2011 Tōhoku Earthquake and Tsunami

At 14.46 local time (05.56 GMT) on March 11, 2011, an undersea earthquake measuring 9.0 on the Richter scale occurred approximately 43 miles (70 kilometers) east of the Japanese coast of Tōhoku. This was the most powerful earthquake ever to have been scientifically documented in Japan, and the fifth most powerful earthquake in the world since modern record-keeping began in c. 1900. The earthquake moved Honshu (the main island of Japan) 2.4 m (8 ft) east and shifted the Earth on its axis by estimates of between 10 cm (4 in) and 25 cm (10 in). Of more immediate significance, the earthquake caused severe earth tremors on the main islands of Japan and triggered powerful tsunami waves that reached heights of up to 40.5 meters (133 ft) in Tōhoku Prefecture and traveled up to 10 km (6 mi) inland in Sendai.

Directly or indirectly, the earthquake led to at least 15,883 deaths and the partial or total collapse of over 380,000 buildings. It also caused extensive and severe structural damage in northeastern Japan (Figure 1.3B), including heavy damage to roads and railways, as well as fires in many areas and a dam collapse. In its immediate aftermath, 4.4 million households in northeastern Japan were left without electricity and 1.5 million without water. In the following days, the tsunami set in action events that led to cooling system failures, explosions, and major meltdowns at three reactors of the Fukushima Daiichi Nuclear Power Plant and the associated evacuation of hundreds of thousands of residents. The World Bank estimated the economic cost at US$235 billion, making it the costliest natural disaster in world history.

All of this happened to a very advanced economy in an earthquake-prone region, which was almost certainly the best prepared in the world for a natural disaster of this kind. GI systems had been used to assemble information on a full range of spatially distributed

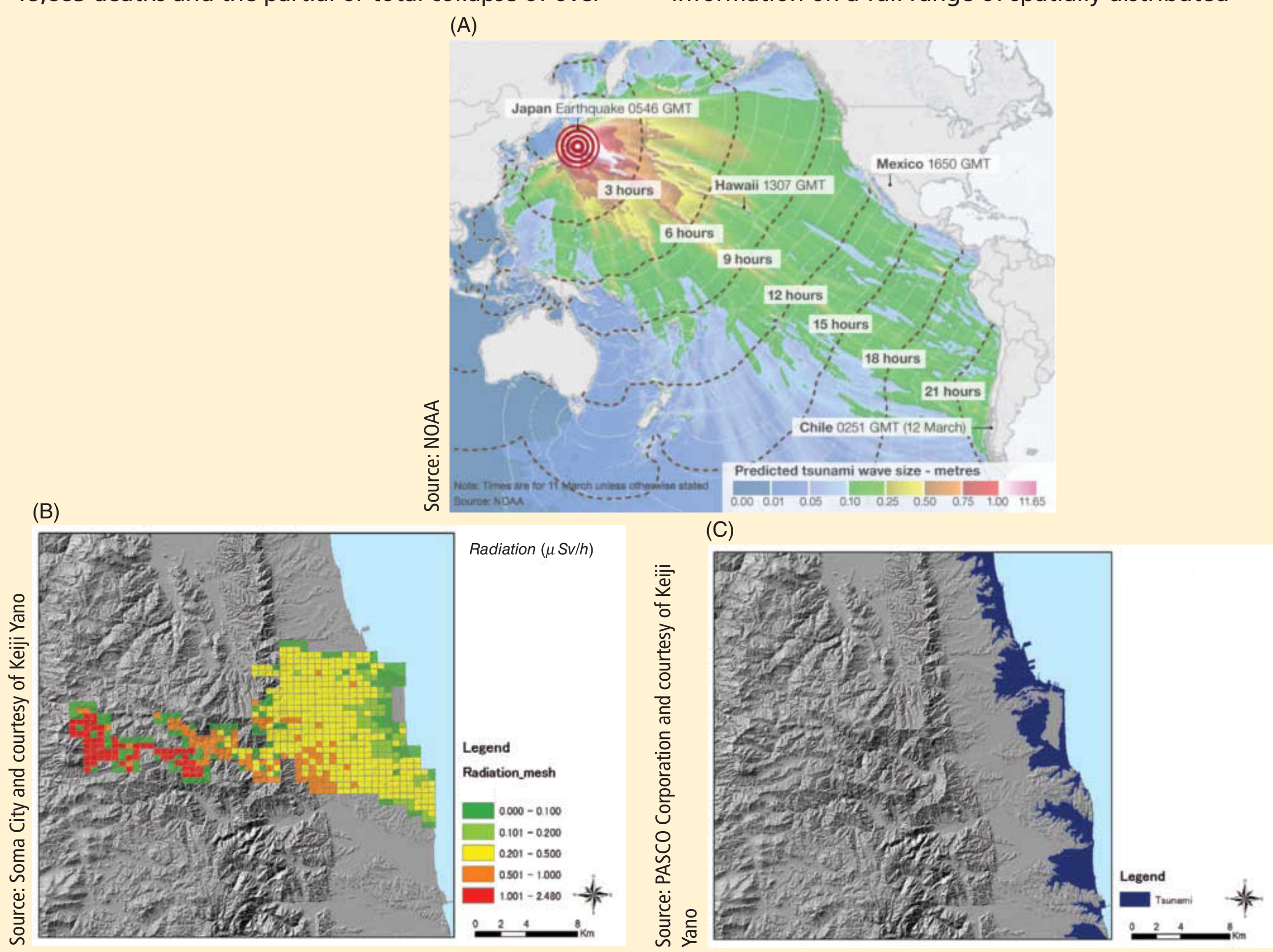


Figure 1.3 (A) The passage of the tsunami arising out of the Great East Japan (Tōhoku) earthquake of March 11, 2011. It had subsequent effects on Soma City in terms of (B) radiation (measured in μSv/h (micro Sievert per hour) and (C) tsunami inundation.

phenomena—including the human population, the built environment, and transportation infrastructure—in preparation for a major earthquake disaster and protection against many of its foreseeable consequences.

Yet the science of predicting the location, timing, and intensity of earthquakes has made little progress over the past century. A magnitude-9.0 earthquake is a very rare event and so did not fall within any disaster-management scenario prior to the event. For example, the Fukushima reactors had been built to withstand a magnitude-8.6 earthquake on the basis of historic occurrences plus a safety margin: but not an event of magnitude 9.0. However, even when major events are unforeseen, GI science and systems are integral to response and recovery in the short term (e.g., alerting populations to the imminent arrival of a tsunami, coordinating citizen reports of how localities have been affected, and organizing evacuation), the medium term (e.g., managing the disruption to industrial supply chains), and the long term (e.g., prioritizing repair and replacement of damaged transport infrastructure). All these actions take place in an organizational context. Early warning systems are very much an international effort. In terms of addressing effects after the event, the Tōhoku earthquake raised issues that were best addressed at the national level, whereas much of the implementation was best effected at local levels.

The three Ps of disaster management are prevention, preparedness, and protection. GI science and systems are integral to each of them.

evidence is found). Finding new locations for retailers, with its focus on design, is an example of a normative application of GI systems. But to predict how consumers will respond to new locations, it is necessary for retailers to analyze and model the actual patterns of behavior they exhibit. Therefore, the models they use will be grounded in observations of messy reality that have been tested in a positive manner.

Design is concerned with improving the world—with decisions that when implemented achieve certain desired objectives, such as constructing new housing subdivisions, developing conservation plans, or defining sales territories. In recent years the term *geodesign* has become a popular way of referring to design decisions at geographic scales, supported by GI systems. All of us would like to design improvements to the world, and GI systems are valuable tools for doing so. Although most work with GI systems is considerably more mundane, it is always good to bear its grander potential in mind. As we show in Section 14.4, geodesign combines two important functions of GI systems—the ability to capture new ideas through sketching (creating/editing new features) and the ability to evaluate them and assess their impacts. A user might sketch a design for a new development, for example, and ask the GI system to predict its impacts on transportation, groundwater, and air pollution.

With a single collection of tools, GI systems are able to bridge the gap between curiosity-driven science and practical problem solving

The third way in which problems can be distinguished is on the basis of their *time scale*, ranging in human terms from the dynastic (perhaps thousands of years; see Box 1.2) to the diurnal, but very much longer with respect to understanding geological or geomorphological change. At one end of the human time spectrum, some decisions are operational and are required for the smooth day-to-day functioning of an organization, such as how to control electricity inputs into grids that experience daily surges and troughs in usage. At slightly longer timescales, tactical decisions might include where to cut trees in next year's forest harvesting plan. Still other decisions are more infrequent and strategic in nature, such as those required to give an organization long-term direction, as when a retailer decides to expand or rationalize its store network (Figure 1.4). At the far end of the human time spectrum, Box 1.2 describes how the geographic

Figure 1.4 Many store location principles are generic across different retail markets, as with Tesco's investment in Ostrava, Czech Republic.

distributions of family names, past and present, can be used to indicate how settled (or otherwise) is the population of different places, and even the geography of the DNA of long-settled residents consequent on population movements in early human history (see Box 1.4).

Although humans like to classify time frames into hours, days, years, centuries, and epochs, the real world is somewhat more complex than this, and these distinctions may blur—what is theoretically and statistically a 1000-year flood in a river system influences strategic and tactical considerations, but may arrive a year after the previous one! Other problems that interest geophysicists, geologists, or evolutionary biologists may occur on timescales that are much longer than a human lifetime, but are still geographic in nature, such as predictions about the future physical environment of Japan or about the animal populations of Africa. GI databases are often *transactional* (see Section 9.9.1), meaning that they are constantly being updated as new information arrives, unlike paper maps, which stay the same once printed.

Applications are discussed to illustrate particular principles, techniques, analytic methods, and management practices (such as risk minimization) as these arise throughout the book.

1.1.2 Spatial Is Special

The adjective *geographic* refers to the Earth's surface and near surface, at scales from the architectural to the global. This defines the subject matter of this book, but other terms have similar meaning. *Spatial* refers to any space, not only the space of the Earth's surface; this term is used frequently in the book, almost always with the same meaning as *geographic*. But many of the methods used in GI systems are also applicable to other nongeographic spaces, including the surfaces of other planets, the space of the cosmos, and the space of the human body that is captured by medical images. Techniques that are integral to GI systems have even been applied to the analysis of genome sequences on DNA. So the discussion of analysis

Applications Box 1.2

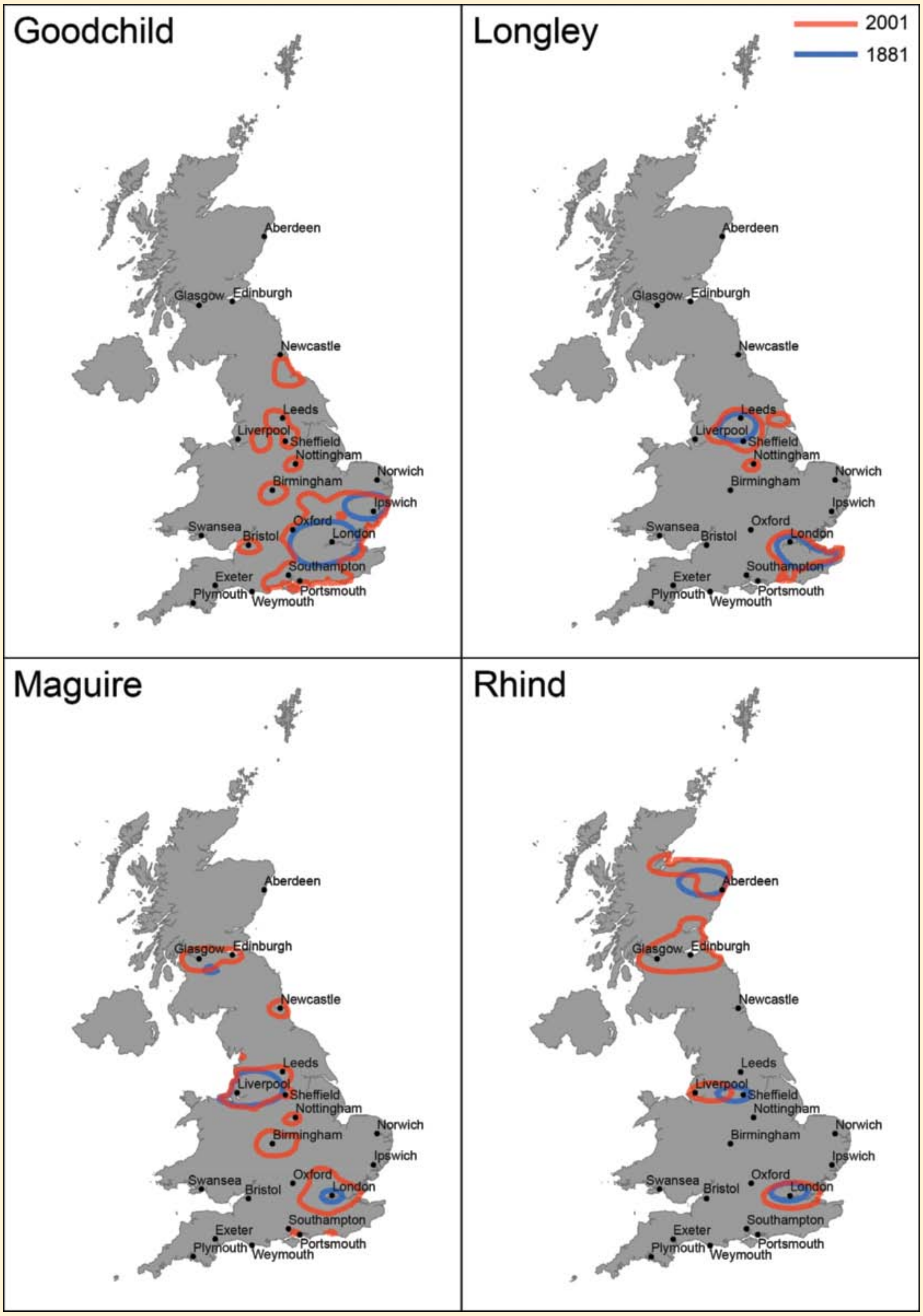
Researching Family Histories and Geo-Genealogy

As individuals, many of us are interested in *where*, in general terms, we came from at different points in recorded human history—for example, whether we are of Irish, Spanish, or Italian descent. More specific locational information can provide clues about the work and other lifestyle characteristics of our ancestors. Some of the best clues to our ancestry may come from our surnames (family names) because many surnames indicate geographic origins to greater or lesser degrees of precision (such clues are less important in some Eastern societies, where family histories are generally much better documented). Research at University College London uses GI systems to analyze historic and present-day lists of names to investigate the changing local and regional geographies of surnames across the world. Figure 1.5 illustrates how the bearers of four selected Anglo-Saxon names in Great Britain (the ancestors of the authors of this book) have mostly stayed put in those parts of the island where the names first came into common parlance at some point between the 12th and 14th centuries—although some have evidently migrated to urban centers.

It also turns out that the mix of names with similar geographic origins in any given area can provide a good indication of regional identity. Figure 1.6, derived from the PhD thesis of Jens Kandt, presents a regionalization of Great Britain on the basis of the present-day residences of bearers of different surnames. (This is essentially a geography of rural Britain. Note that the major urban areas have been excluded because they are characterized by mixes of names arising from urban–rural, interregional, and international migration over the last 200 or so years).

All of this is most obviously evident for Great Britain and many of the countries of Europe, where populations have remained settled close to the locations at which their names were first coined. But there is also evidence to suggest that the spatial patterning of names in former colonies, such as North America, Australia, and New Zealand, is far from random. Figure 1.7 illustrates this for the surname Singleton, which can be used to build evidence about the migration patterns of bearers of this name from their documented origins in northwest England.

Fundamentally, this is curiosity-driven research, driven by the desire among amateur genealogists to discover their roots. But the same techniques can be used to represent the nature and depth of affiliation that people feel toward the places in which they live. Moreover, the work of Sir Walter Bodmer and colleagues (Box 1.4) is highlighting probable links between surnames and genetics, rendering this curiosity-driven research relevant to the development of drug and lifestyle interventions.

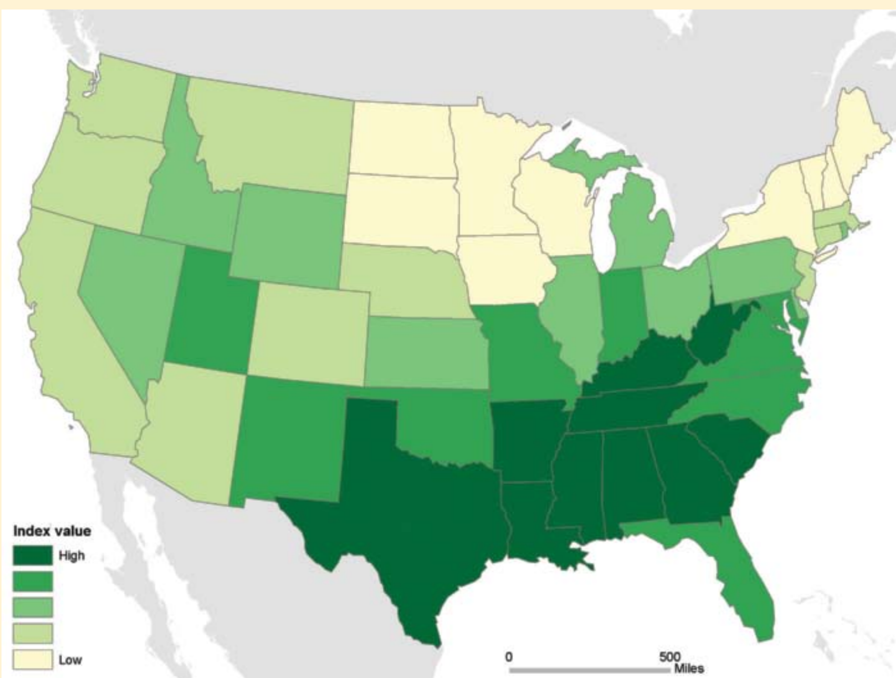


Courtesy: James Cheshire

Figure 1.5 The Great Britain Geography of the Longleys, Goodchilds, Maguires, and Rhinds. In each case the shorter (blue) line delineates the smallest possible area within which 95% of name bearers reside, based on 1881 Census of Population figures, and the outer (red) line encloses the smallest area that accommodates the same proportion of adult name bearers according to a recent address register.

Courtesy: Jens Kandt

Figure 1.6 A regionalization based on the coincidence of distinctive patterns of surnames, showing the southern part of Great Britain. Major urban areas do not fit into this regional pattern because their residents are drawn from a wide range of national and international origins.



Courtesy: Alex Singleton

Figure 1.7 The Singleton family name derives from a place in northwest England, and understandably the greatest concentration of this name today still occurs in this region. But why should the name be disproportionately concentrated in the south and west of the United States? Geographical analysis of the global pattern of family names can help us to hypothesize about the historic migrations of families, communities, and cultural groups.

Technical Box 1.3

Some Technical Reasons Why Geographic Information Is Special and Why GI Science and Systems Have Developed

- It is multidimensional, because *two* coordinates must be specified to define a location, whether they be *x* and *y* or latitude and longitude; and a third coordinate is needed when elevation is important.
- It is voluminous because a geographic database can easily reach a terabyte in size (see Table 1.2).
- It may be collected by citizens, governments, or other organizations, and it may prove useful to pool information from these diverse sources.
- It may be represented at different levels of spatial resolution, for example, by using a representation equivalent to a 1:1 million-scale map or a 1:24,000-scale one (see Section 3.7).
- It may be represented in different ways inside a computer (see Chapter 3), and how this is done can strongly influence the ease of analysis and the end results.
- It must often be projected onto a flat surface, for reasons identified in Section 4.8.
- It requires many special methods for its analysis (see Chapters 13 and 14).
- It may be transformed to present different views of the world, for example, to aid interpretation.

in this book is of *spatial* analysis (see Chapters 13 and 14), not geographic analysis, to emphasize this versatility.

Another term that has been growing in usage in recent years is *geospatial*—implying a subset of spatial applied specifically to the Earth's surface and near surface. In this book we have tended to avoid *geospatial*, preferring *geographic*, and we use *spatial* where we need to emphasize generality.

Although there are subtle distinctions between the terms *geographic(al), spatial,* and *geospatial,* for many practical purposes they can be used interchangeably.

People who encounter GI for the first time are sometimes driven to ask why geography is so important; why, they ask, is spatial special? After all, there is plenty of information around about geriatrics, for example, and in principle one could create a geriatric information system. So why has GI spawned an entire industry, if geriatric information has not done so to anything like the same extent? Why are there unlikely to be courses in universities specifically in geriatric information science and systems? Part of the answer should be clear already: almost all human activities and decisions involve a location component, and the location component is important. Another reason will become apparent in Chapter 2, where we will see that working with GI involves complex and difficult choices that are also largely unique. Other, more technical reasons will become clear in later chapters and are briefly summarized in Box 1.3.

1.2 Data, Information, Evidence, Knowledge, and Wisdom

Information systems help us to manage *what we know*, by making it easy to organize and store, access and retrieve, manipulate and synthesize, and apply to the solution of problems. We use a variety of terms to describe what we know, including the five that head this section and that are shown in Table 1.1. There are no universally agreed-on definitions of these terms. Nevertheless it is worth trying to come to grips with their various meanings because the differences between them can often be significant, and what follows draws on many sources and thus provides the basis for the use of these terms throughout the book. Data clearly refers to the most mundane kind of information and wisdom to the most substantive. *Data* consist of numbers, text, or symbols, which are in some sense neutral and almost context-free. Raw geographic facts, such as sensor measurements of temperature at a specific time and location, are examples of data. When data are transmitted, they are treated as a stream of bits; a crucial requirement is to preserve the integrity of the data set. The internal meaning of the data is irrelevant in such considerations. Data (the noun is the plural of datum) are assembled together in a

Table 1.1 A ranking of the support infrastructure for decision making.

Decision-making support infrastructure	Ease of sharing with everyone	GIS example
Wisdom ↑	*Impossible*	Policies developed and accepted by stakeholders
Knowledge ↑	*Difficult, especially tacit knowledge*	Personal knowledge about places and issues
Evidence ↑	*Often not easy*	Results of GIS analysis of many data sets or scenarios
Information ↑	*Easy*	Contents of a database assembled from raw facts
Data	*Easy*	Raw geographic facts

database (see Chapter 9), and the volumes of data that are required for some typical applications are shown in Table 1.2.

The term *information* can be used either narrowly or broadly (and we use both in this book). In a narrow sense, information can be treated as devoid of meaning and therefore as essentially synonymous with data as defined in the previous paragraph. Others define information as *anything* that can be digitized, that is, represented in digital form (see Chapter 3), but also argue that information is differentiated from data by implying some degree of selection, organization, and preparation for particular purposes—information is data serving some *purpose* or data that have been given some degree of interpretation. Information is often costly to produce, but once digitized, it is cheap to reproduce and distribute. Geographic data sets, for example, may be very expensive to collect and assemble, but very cheap to copy and disseminate. One other characteristic of information is that it is easy to add value to it through processing and through merger with other information. GI systems are very useful for doing the latter because of the tools they provide for combining information from different sources.

GI systems do a better job of sharing data and information than knowledge, which is more difficult to detach from the knower.

Knowledge does not arise simply from having access to large amounts of information. It can be considered as information to which value has been added by interpretation based on a particular context, experience, and purpose. Put simply, the information available in a book or on the Internet or on a map becomes knowledge only when it has been read and understood, as when an experienced hiker chooses not to set off into unfamiliar terrain having read about it and taken stock of the weather forecast. How the information is interpreted and used will be different for different readers depending on their previous experience, expertise, and needs. It is important to distinguish two types of knowledge: *codified* and *tacit*. Knowledge is codifiable if it can be written down and transferred relatively easily to others. Tacit

Table 1.2 Potential GI database volumes in bytes for some typical applications (volumes estimated to the nearest order of magnitude). Strictly, bytes are counted in powers of 2—1 kilobyte is 1024 bytes, not 1000.

1 megabyte	1 000 000 (2^{20})	Single data set in a small project database
1 gigabyte	1 000 000 000 (2^{30})	Entire street network of a large city or small country
1 terabyte	1 000 000 000 000 (2^{40})	Elevation of entire Earth surface recorded at 30 m intervals
1 petabyte	1 000 000 000 000 000 (2^{50})	Satellite image of entire Earth surface at 1 m resolution
1 exabyte	1 000 000 000 000 000 000 (2^{60})	A possible 3-D representation of the entire Earth at 10 m resolution
1 zettabyte	1 000 000 000 000 000 000 000 (2^{70})	One-fifth of the capacity (in 2013) of U.S. National Security Agency Utah Data Center

knowledge is often slow to acquire and much more difficult to transfer. Examples include the knowledge built up during an apprenticeship, understanding of how a particular market works, or familiarity with using a particular technology or language. This difference in transferability means that codified and tacit knowledge need to be managed and rewarded quite differently. Because of its nature, tacit knowledge is often a source of competitive advantage.

Some have argued that knowledge and information are fundamentally different in at least three important respects:

- Knowledge entails a knower. Information exists independently, but knowledge is intimately related to people.
- Knowledge is harder to detach from the knower than information; shipping, receiving, transferring it between people, or quantifying it are all much more difficult than for information.
- Knowledge requires much more assimilation—we digest it rather than hold it. We may hold conflicting information, but we rarely hold conflicting knowledge.

Evidence is considered a halfway house between information and knowledge. It seems best to regard it as a multiplicity of information from different sources, related to specific problems, and with a consistency that has been validated. Major attempts have been made in medicine to extract evidence from a welter of sometimes contradictory sets of information, drawn from different geographic settings, in what is known as meta-analysis, or the comparative analysis of the results of many previous studies.

Wisdom is even more elusive to define than the other terms. Normally, it is used in the context of decisions made or advice given, which is disinterested, based on all the evidence and knowledge available. It is given with some understanding of the likely consequences of various actions and assessment of which is or are most beneficial. Almost invariably, knowledge is highly individualized rather than being easy to create and share within a group. Wisdom is in a sense the top level of a hierarchy of decision-making infrastructure.

1.3 GI Science and Systems

GI systems are computer-based tools for collecting, storing, processing, analyzing, and visualizing geographic information. They are tools that improve the efficiency and effectiveness of handling information about objects and events located in geographic space. They can be used to carry out many useful tasks, including storing vast amounts of GI in databases, conducting analytical operations in a fraction of the time they would take to do by hand, and automating the process of making useful maps. GI systems also process information, but there are limits to the kinds of procedures and practices that can be automated when turning data into useful information.

The question of whether and how such selectivity and preparation for purpose actually adds value, or whether the results add insight to interpretation in geographic applications, falls into the realm of GI science. This rapidly developing field is concerned with the concepts, principles, and methods that are put into practice using the tools and techniques of GI systems. It provides sound principles for the sample designs used to create data and the ways in which data can be turned into information that is representative of a study area. GI science also provides a framework within which new evidence, knowledge, and ultimately wisdom about the Earth can be created, in ways that are efficient, effective, and safe to use.

Like all sciences, an essential requirement of GI science is a method for discovering new knowledge. The GI scientific method must support:

- Transparency of assumptions and methods so that other GI scientists can determine how previous knowledge has been discovered and how they might themselves add to the existing body of knowledge
- Best attempts to attain objectivity through a detached and independent perspective that avoids or accommodates bias (unintended or otherwise)
- The ability of any other qualified scientist to reproduce the results of an analysis
- Methods of validation using the results of the analysis (internal validation) or other information sources (external validation)
- Generalization from partial representations that are developed for analytical purposes to the wider objective reality that they purport to represent

How, then, are problems solved using a scientific method, and are geographic problems solved in ways different from other kinds of problems? We humans have accumulated a vast storehouse of knowledge about the world, including information both on how it *looks*—that is, its *forms*—and how it *works*—that is, its dynamic *processes*. Some of those processes are natural and built into the design of the planet, such as the processes of tectonic movement that lead to earthquakes and the processes of atmospheric circulation that lead to hurricanes. Others are human in origin, reflecting the increasing influence that we have on ecosystems,

Figure 1.8 Social processes, such as carbon dioxide emissions, modify the Earth's environment independent of location.

through the burning of fossil fuels, the felling of forests, and the cultivation of crops (Figure 1.8). Still others are imposed by us, in the form of laws, regulations, and practices: for example, zoning regulations affect the ways in which specific parcels of land can be used.

Knowledge about how the world works is more valuable than knowledge about how it looks. This is because knowledge about how it works can be used to predict.

These two types of information differ markedly in their degree of generality. Form varies geographically, and the Earth's surface looks dramatically different in different places; compare the settled landscape of northern England with the deserts of the U.S. Southwest (Figure 1.9). But processes can be very general. The ways in which the burning of fossil fuels affects the atmosphere are essentially the same in China as in Europe, although the two landscapes look very different. Science has always valued such general knowledge over knowledge of the specific, and hence has valued process knowledge over knowledge of form. Geographers in particular have witnessed a long-standing debate, lasting centuries, between the competing needs of *idiographic* geography, which focuses on the description of form and emphasizes the unique characteristics of places, and *nomothetic* geography, which seeks to discover general processes. Both are essential, of course, because knowledge of general process is useful in solving specific problems only if it can be combined effectively with knowledge of form. For example, we can only assess the risk of roadside landslip if we know both how slope stability is generally affected by such factors as shallow subsurface characteristics and porosity and where slopes at risk are located (Figure 1.10).

One of the most important merits of a GI system as a problem-solving tool lies in its ability to combine the general with the specific, as in this example. A GI system designed to solve this problem would contain knowledge of local slopes, in the form of computerized maps, and the programs executed by the GI system would reflect general knowledge of how slopes affect the probability of mass movement under extreme weather conditions. The *software* of a GI system captures and implements general knowledge, whereas the *database* of a GI system represents specific information. In that sense, a GI system resolves the long-standing debate between nomothetic and idiographic camps by accommodating both.

GI systems solve the ancient problem of combining general scientific knowledge with specific information and give practical value to both.

Figure 1.9 The form of the Earth's surface shows enormous variability, for example, between (A) the deserts of the southwest United States and (B) the settled landscape of Northern England.

(A)

(B)

Figure 1.10 Predicting landslides requires general knowledge of processes and specific knowledge of the area—both can be brought together in a GI system.

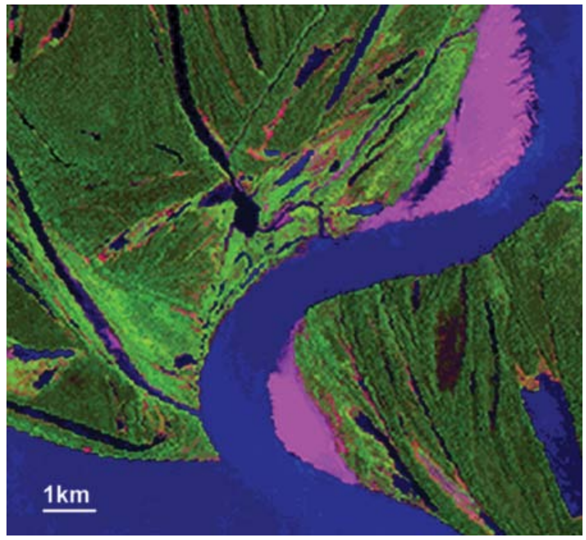


Figure 1.11 A classified Landsat image (at 30-meter resolution) of part of the Amazon region of Brazil.

This perspective is consistent with our understanding of *places* in the world as sites at which unique relations develop among people and the locations that they occupy and the accumulated effects of these relations over time. GI systems provide ways of generalizing about and between places, albeit in ways that acknowledge differences between them. Place-based methods in GI systems make it possible to think of geography as repetitive (where in the world is like this place?) while at the same time remaining sensitive to the unique context of unique places.

General knowledge about unique places comes in many forms. Classification is perhaps the simplest and most rudimentary and is widely used in problem solving. In many parts of the United States and other countries, efforts have been made to limit the development of wetlands in the interest of preserving them as natural habitats and avoiding excessive impact on water resources. To support these efforts, resources have been invested in mapping wetlands, largely from aerial photography and satellite imagery. These maps simply classify land, using established rules that define what is and what is not a wetland (Figure 1.11).

More sophisticated forms of knowledge include *rule sets*—for example, rules that determine what use can be made of wetlands, or what areas in a forest can be legally logged. The U.S. Forest Service has rules to define wilderness and to impose associated regulations regarding the use of wilderness, including prohibition on logging and road construction. Such rules can be captured in the data model of a GI database (see Chapter 7).

Much of the knowledge gathered by the activities of scientists suggests the term *law*. The work of Sir Isaac Newton established the Laws of Motion, according to which all matter behaves in ways that can be perfectly predicted. From Newton's Laws we are able to predict the motions of the planets almost perfectly, although Einstein later showed that certain observed deviations from the predictions of the Laws could be explained with his Theory of Relativity. Laws of this level of predictive quality are few and far between in the geographic world of the Earth's surface. The real world is the only laboratory that is available for understanding the effects of many factors on unique places in the social and environmental sciences, and considerable uncertainty is generated when we are unable to control for all conditions. These problems are compounded in the social realm, where the role of human agency makes it almost inevitable that any attempt to develop rigid laws will be frustrated by isolated exceptions. Thus, whereas market researchers use spatial interaction models, in conjunction with GI systems, to predict how many people will shop at each shopping center in a city, substantial errors will occur in the predictions—because people are in significant part autonomous agents. Nevertheless, the results are of great value in developing location strategies for retailing. The Universal Soil Loss Equation, used by soil scientists in conjunction with GI systems to predict soil erosion, is similar in its rather low predictive power, but again the results are sufficiently accurate to be very useful in the right circumstances. "Good" usually means "good enough for this specific application" in GI systems applications.

Solving problems involves several distinct components and stages. First, there must be an *objective*, or a goal that the problem solver wishes to achieve. Often this is a desire to maximize or minimize—find the solution of least cost, shortest distance, least time, greatest profit or make the most accurate prediction possible. These objectives are all expressed in *tangible* form; that is, they can be measured on some well-defined scale. Others are said to be *intangible* and involve objectives that are much harder, if not impossible, to measure. They include maximizing *quality of life* and *satisfaction* and minimizing *environmental impact*. Sometimes the only way to work with such intangible objectives is to involve human subjects, through surveys or focus groups, by asking them to express a preference among alternatives. A large body of knowledge has been acquired about such human-subjects research, and much of it has been employed in connection with the design of GI systems. For discussion of the use of such mixed objectives see Section 15.4. This topic is taken up again in Chapter 16 in the context of estimating the return on investment of GI systems.

Often a problem will have *multiple objectives*, each of which is measured in a different way. For example, a company providing a mobile snack service to construction sites will want to maximize the number of sites that can be visited during a daily operating schedule and will also want to maximize the expected returns by visiting the most lucrative sites. An agency charged with locating a corridor for a new power transmission line may decide to minimize cost, while at the same time seeking to minimize environmental impact. Such problems employ methods known as *multicriteria decision making* (MCDM).

Many geographic problems involve multiple goals and objectives, which often cannot be expressed in commensurate terms.

1.4 The Technology of Problem Solving

Today it is a truism to reflect that geographic information is everywhere and that we access and divulge it from many different sources and in many different contexts. A system is usually thought of as a *bounded* set of components, and in a world in which geographic information is transmitted and shared across physical, public/private, political, and institutional *networks*, it hardly seems to make sense to think in terms of bounded systems at all. However, although geographic information may be pervasive and ubiquitous, the notion of a networked system remains useful in understanding the components of the tools that in turn underpin GI science (Figure 1.12).

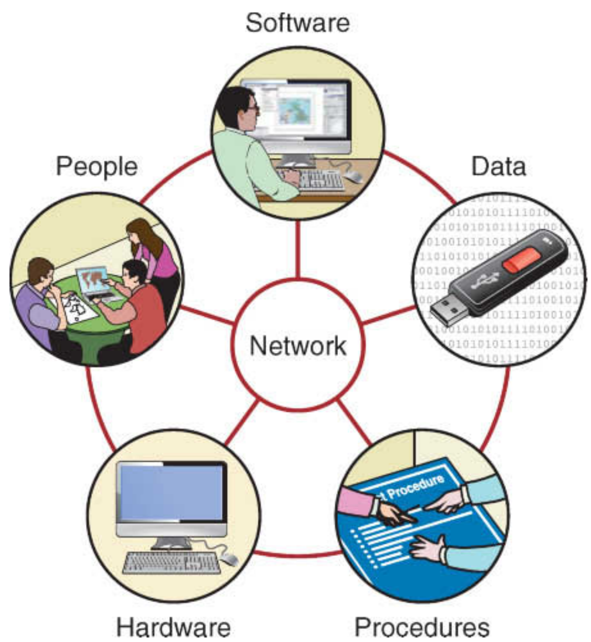


Figure 1.12 The six component parts of a GI system.

Today, almost all GI software products are designed as components of a network. *Cloud computing* (see Chapter 10) is a colloquial expression that is widely used in business to describe the supply of hosted services to industry and commerce, using computer infrastructure that is located remotely. Networks of large numbers of computers in different locations may be used for collection, storage, and analysis of data in real time. Cloud computing makes it possible to gain convenient, on-demand network access to a shared pool of computer hardware, software, data storage, and other services. Many of these components were previously colocated prior to the innovations of fast wide-area networks; powerful, inexpensive server computers; and high-performance virtualization of computer hardware.

In terms of hardware, the user's device is the *client*, connected through the network to a *server*, or a *server farm* in the Cloud, that is designed to handle many other user clients simultaneously. The client may be *thick*, if it performs a large part of the work locally, or *thin* if it does little more than link the user to the server (as with a mobile phone application, for example). In Cloud computing applications, most or all of the computation is performed remotely.

Uptake and use of the Internet to link computers has been remarkably quick, diffusion being considerably faster than almost all comparable innovations (for

Table 1.3 World Internet usage and penetration statistics as of June 30 2012. (Source: www.internetworldstats.com)

World Region	2012 Population	Internet Users (Dec. 2000)	Internet Users (June 2012)	Penetration (% Population)	Growth 2000–12	Users % of all Table
Africa	1,073,380,925	4,514,400	167,335,676	15.6%	3,607%	7.0
Asia	3,922,066,987	114,304,000	1,076,681,059	27.5%	842%	44.8
Europe (inc. EU)	820,918,446	105,096,093	518,512,109	63.2%	393%	21.5
Middle East	223,608,203	3,284,800	90,000,455	40.2%	2,640%	3.7
North America (excl. Canada)	348,280,154	108,096,800	273,785,413	78.6%	153%	11.4
Latin America/ Caribbean	593,688,638	18,068,919	254,915,745	42.9%	1,311%	10.6
Oceania/Australia	35,903,569	7,620,480	24,287,919	67.6%	219%	1.0
WORLD TOTAL	7,017,846,922	360,985,492	2,405,518,376	34.3%	566%	100.0

example, the radio, the telephone, and the television). RealTimeStatistics.org estimated that in 2013 some 2.4 billion of the world's 7 billion population were Internet users, although stark variations in Internet availability and usage remain—see Table 1.3.

Many of the early Internet applications of GI systems remain in use, in updated form, today. They range from using GI systems on the Internet to disseminate information on the location of businesses (e.g., www.yell.com), to consolidated lists of available goods and services, to direct revenue generation through subscription services, to helping members of the public to participate in important local, regional, and national debates. The Internet has become very popular as a vehicle for delivering business GI system applications for several reasons. It provides an established, widely used platform and accepted standards for interacting with information of many types. It also offers a cost-effective way of linking distributed users (for example, telecommuters and office workers, customers and suppliers, students and teachers). From the early days onward, the interactive and exploratory nature of navigating linked information became a great hit with users.

Internet-enabled devices became portable in the early 2000s (see Section 10.3) with the wide diffusion of location-aware smartphones and other handheld devices and the availability of wireless networks in public places such as airports and railway stations. The subsequent innovation of 3G and 4G mobile broadband now routinely allows portable and in-vehicle devices to deliver *location-based services* (see Section 10.3.2) to users on the move. Users receive real-time geographic services such as mapping, routing, traffic congestion, and geographic yellow pages. These services are usually funded directly or indirectly through advertising, with Google perhaps the most obvious exponent of understanding the importance of location in delivering targeted advertising.

We now turn to consider the other components of a GI system. First, the user's *hardware* is the device that the user interacts with directly in carrying out GI system operations, by typing, pointing, clicking, or speaking, and that returns information by displaying it on the device's screen or generating meaningful sounds. Traditionally, this device sat on an office desktop, but today's user has more options and much more freedom because GI system functions can also be delivered through smartphones, notebooks, and in-vehicle devices.

The second component is the software programs that represent the world by running locally in the user's machine or remotely in the Cloud. Increasing numbers of users manipulate geographic information using executable *open-source software* code that is often freely available for download across the Web. Users can execute this code and also modify it if they wish. Other *open software* is also available for use as linked executable files, although the computer code that was used to generate it is not made available by its authors and so cannot be modified by other users. Both of these types of software may be made available by their authors in the interests of solving particular problems, or they may be made available as part of larger linked software libraries, such as the R project for statistical computing and graphics (www.r-project.org/). Some open software libraries have a focus on geographic problem solving and as such are described as GI systems—with the Quantum GIS Project (www.qgis.org/) providing perhaps the best contemporary example. The international "Geo for All" initiative (www.geoforall.org/) seeks