reliable, accurate, and *verifiable* (by independent means), and only when it is *up-to-date, timely*, and *complete* (in terms of its attributes, and its spatial and temporal coverage). Information must be *intelligible*, that is, *comprehensible* by its recipients, *consistent* with other sources of information, *convenient* (easy to handle), and *adequately protected*. As such, much of the public domain spatial data we will be using in this book are really information, because the data have already been processed into a form that is useful to GIS users. However, gaps in the documentation about the information, or indeed the information itself, may prevent it from being used to its full potential, as we shall see later on. Whether that information should be used or not, and the pros and cons of doing so, is one of the major themes of this book.

The primary function of an information system is to change data into information using one of the following four processes:

Conversion: The conversion process transforms data from one format to another, from one unit of measurement to another, and/or from one feature classification to another.

Organization: Organization processes arrange data according to database management rules and procedures so that they can be stored and accessed effectively.

Structuring: The structuring process formats or reformats data so that they can be acceptable to a particular software application or information system.

Modeling: Modeling processes include visualization and analysis that improve the user's knowledge base through the ability to emulate real-world processes and document workflows. This allows procedures and techniques to be replicated and the outcome predicted and assists the decision-making process.

The concepts of organization and structure are crucial to the functioning of information systems: Without organization and structure, it would be impossible to turn data into information.

Spatial data

The term *spatial data* is often used synonymously with *geographic data* and *geospatial data*. Spatial data are considered by some to be wider in scope than geographic data by including anything about space that doesn't necessarily have geographic coordinates, such as an MRI (magnetic resonance imaging) scan or CAD (computer-aided design) drawings of a building. For the purpose of this book, the terms are used interchangeably. The book focuses on data that are pertinent to features above, on, or below the surface of the earth—in other words, those having real-world coordinates. Such data are collected, manipulated, and used to make decisions in association with some aspect of geography, such as the context, distribution, or spatial relationships at a given location. Consider the following two seemingly diverse issues: crime in a neighborhood and soil erosion in a county. Each occurs at different scales and at different rates. Each affects, and is affected by, people in different ways, yet manifests itself as geographic patterns and phenomena which may be represented by a single layer or multiple layers of spatial data.

Spatial data are geographically referenced; they are identified and located by coordinates. These coordinates, in turn, may be part of a global, national, or regional referencing system but all include a point of origin, axes, and measurement units. Spatial data include information about where they are located and their extent. Spatial data are stored as points, lines, areas, or pixels, and include a symbol, or set of symbols to represent the data. Most spatial data exist within a topological framework, which determines how data items in the same layer or in different layers are spatially related to one another. These elements can be thought of collectively as the *G* part of GIS. The elements represent real-world features or phenomena, such as ocean currents, soil types, or fault lines. Spatial data also contain a descriptive element that informs the users what the data represent. The descriptive element can be thought of as the *I* part of GIS, or *information*. This could be the name of the ocean current, the pH of the soil type, or the strike and dip of the fault line. The relationship between individual data items reflects the *S* or *systems* part of GIS. If specific soils are queried or selected on the map, the resulting selection is reflected in the table containing the attributes. Conversely, if the soils types are selected in the attribute table, the corresponding feature selection is displayed on the map. The graphical element is commonly referred to as spatial data, while the descriptive element is commonly referred to as nonspatial or attribute data. Taken together, these elements may be used to represent spatial phenomena, issues, and objects as mappable items.

Spatial information is obtained by processing geographic data, the aim of which is to improve the user's knowledge and understanding of the geography of features and resources. This will help promote a better understanding of the consequences of human activities associated with those features and resources by developing spatial intelligence for problem solving and decision making.

Some will argue that spatial data are special data, but for others that designation simply reinforces the notion that spatial data are difficult to use and remain the preserve of the specialist and technically minded. Capturing, managing, and maintaining spatial data does involve some specialized skills, but most people who work with spatial data are not so concerned about how the data are captured, stored, and updated. Rather, they just care that the data are available when they need it, that the source and any use restrictions are clearly documented, the data are easy to use, and help complete their projects.

Using spatial data and information to solve problems is nothing new. Some of the earliest known maps date back to prehistoric times and can be found in the cave paintings and engravings of the earliest civilizations. Although its interpretation as a map is debated by some, the markings on a fragment of mammoth tusk, excavated in Mezhirichi, Ukraine, in 1966, are thought to represent some dwellings and the location of fishing nets on the bank of a river (Harley and Woodward 1987). Examples from ancient Egypt include the Turin Papyrus Map, believed to date from 1160 BC, which was prepared to assist in recording the location of rock suitable for creating sculptures. The map was annotated with the location of gold deposits, the destinations of some of the wadi routes, and distances between quarries and mines, among other things.

One of the best examples from Roman times is the *Forma Urbis Romae*, a highly detailed ground plan of every building and monument in Rome. The plan was incised onto marble slabs that hung on a wall in the Templum Pacis in Rome. The exact purpose of the plan is unclear, but given its size and location, it is thought by some to be a decorative representation of the more utilitarian cadastral mapping of the time that may have had some role in urban planning. Also from Roman times, the Peutinger Map or *Tabula Peutingeriana*, a copy of a Roman map made in the thirteenth century, shows an elongated schematic of the road network in the Roman Empire. The map depicts Roman settlements and the roads connecting them, rivers, mountains, forests, and seas. In addition, the distances between the settlements are recorded in a variety of measurements.

During the Age of the Enlightenment, the use of maps greatly expanded as interest arose in faraway places. Maps stirred imaginations and inspired explorations of the unknown. Yet mapmaking was reserved for the select few who had access to the data from surveyors and explorers and also possessed the tools and the necessary skills to bring those data to life. Having access to accurate mapped information bestowed a great deal of power on those who were able to exploit it for their own gains.

No longer the preserve of the intelligentsia, the production of, and access to, spatial data is evolving in the twenty-first century at ever-increasing speeds. By 2010, more than one million servers existed around the world; every day, Google alone processed one billion search requests and 20 petabytes (1000 terabytes or 1,000,000 gigabytes) of user-generated data. As much of this data are in the form of custom-made maps, it is probably safe to say that more maps are now made each month than the sum total of maps made in all recorded human history. The amount of spatial data available to GIS practitioners is greater now than at any time in the fifty-year history of GIS.

In the early days of GIS, when comparatively little digital data were available, a standard piece of hardware next to nearly every GIS workstation was the digitizing tablet; the first task in most projects was to collect data. While GIS users today still generate some of their own data, the amount of data amassed from decades of investment by government agencies, academia, industry, and nonprofit organizations means that many users can simply tap into the vast reserves of data created by others for their own projects.

Not only has the amount of available data increased, but also the variety of spatial data has greatly improved. During the 1980s, geographic base files/dual independent map encoding (GBF/DIME) files from the US Census Bureau along with Digital Line Graphs and Landsat satellite imagery from the US Geological Survey (USGS) comprised the bulk of spatial data available in the United States. Federal agencies were responsible for the majority of data production, as only they had the necessary staff and technical resources required to convert paper maps to digital data and create new digital data from satellite imagery. Nowadays, spatial data are available for hundreds of themes and phenomena, literally from A to Z—agriculture to zebra mussels. In addition, the number of data formats has also expanded. Data are available in raster form, in vector form, as static documents, and through real-time dynamic services. At the same time, the ways to access and obtain data have never been greater. Data can be downloaded as files and stored on a local computer, data can be streamed from the web without having to store a copy locally, or data may be ordered on physical media.

The number of organizations providing spatial data increased steadily through the first decade of the 2000s. Higher educational institutions, nonprofit organizations, private companies, and local and regional governments joined ranks with the federal agencies as the top data providers. As the decade came to a close, individual GIS users contributed to national and international data collection initiatives on an unprecedented scale thanks to easy-to-use application programming interfaces (APIs), broadband access, and the networking and hardware infrastructure necessary to support this emerging online community. Known as citizen science or crowdsourcing, GIS users can now regularly provide their own volunteered geographic information. The new paradigm of citizen science brings renewed attention to familiar data issues:

Can user-generated data be trusted?

How can anyone using the data understand the contexts, scales, accuracy, and situations where these data should be used or, perhaps more importantly, should not be used?

What is the role of traditional data providers in this new era?

Despite the tremendous expansion in the amount, variety, formats, means of distribution, and number of data providers, challenges still remain today for those wishing to work with spatial data. Perhaps one of the greatest challenges now is locating the most appropriate spatial data for a particular project. It has become very easy to establish a data portal, resulting in the ever-increasing number and variety of portals we see today. However, not all portals are created equally. Some are well documented, while others are not. Some work only with certain web browsers, while others require web browser settings to be adjusted or browser plug-ins to be installed. Some cover multiple regions, such as the SERVIR Regional Visualization and Monitoring System website where some sections are populated with more data (SERVIR Mesoamerica), while others are less populated (SERVIR Africa). There is still no comprehensive and all-inclusive one-stop shop for finding spatial data.

Obtaining spatial data

There are several ways to obtain spatial data. You could purchase data from an organization or individual that has already collected (or will collect) them for you or download data for free. Alternatively, you could create the data yourself: gather the relevant attribute data along with spatial coordinates from a GPS-enabled device, geocode points from a set of address data, scan paper maps or imagery, or gather data via other means. Regardless of which method you choose, there are associated costs and benefits in terms of finances, time, data quality, copyright, and other issues, as we shall examine later.

This book focuses on downloading or streaming spatial data and, in particular, data that are available for free with subsequent usage not subject to copyright or licensing constraints. In other words, public domain data. To appreciate what public domain data can offer GIS analysts and end users, you need to understand what public domain means and what can you do with the data from those sources. It is also important to understand the impact public data have had on the use and application of GIS in the past and the impact they will have in the future.