Spatial data provide the foundation for making decisions with geographic information systems (GIS). Without real-world data, GIS analysis is reduced to theory and conjecture. Wilson (2001) may have summed it up best: “After more than thirty years, we’re still confronted by the same major challenge that GIS professionals have always faced: You must have good data. And good data are expensive and difficult to create” (p. 54).

Today, spatial data are available in more formats, obtainable by more methods, produced by more agencies and organizations, at more scales, in more themes, and at a higher resolution than ever before. Much of the data are in the public domain and literally at the fingertips of nearly every GIS user, ready to be streamed or downloaded and used in the decision-making process. Why then is there a need for this book? Despite the acknowledged importance of having access to good data, the increasing availability of public domain spatial data makes it tempting to plunge right in, download, and start using data within a GIS environment without knowing a great deal about the data:

How do you know if the data are suitable for your project?
How do you know if what you are planning to do with the data is appropriate?
Where did the data come from, who collected them, how were they collected, at what scale were they collected, and are they accurate enough for your project?

It is also important to understand that not all publicly available spatial data are in the public domain. Just because a dataset is available to download from a website, that does not guarantee it is free from copyright and licensing restrictions, or possibly subject to a fee.

Failure to appreciate both the benefits and limitations of using public domain spatial data sources will inevitably lead to, at best, inappropriate decisions with unforeseen economic, legal, or environmental consequences. At worst, lives and property could be at risk as a result of flawed analysis.

Equally as important as gaining knowledge about data and practicing skills in using them is a thorough understanding of the related issues and policies at the local, regional, and global levels. The hardware, software, operating systems, the Internet, and tools to manipulate data are continually changing, but changing just as rapidly are the environments in which all these components operate. This includes the organizations that produce and use spatial data, political systems, national and international standards, spatial data policies and laws, and finally user expectations. How can you, the GIS data user, keep track of these changes to fully appreciate the implications for your projects?

This book has five objectives:
1. To help you understand the sources, scales, accuracy, benefits, and limitations of spatial data.
2. To improve your awareness of the organizations and initiatives that produce spatial data, including public agencies (local, state, federal, tribal, and international), private companies, universities, nonprofit organizations, and multi-organization collaborative initiatives.
3. To enable you to become familiar with the history, current issues, and policies—both domestic and international—that affect the production of and access to public domain spatial data.
4. Through a series of associated exercises, to help you develop a working knowledge of, and practical skills in finding, accessing, downloading, and formatting public domain raster and vector data.
5. To augment your spatial analysis skills to effectively address issues and solve real-world, interdisciplinary problems by using public domain data with GIS software (Esri’s ArcGIS, both the online and the desktop platforms).
This book provides some practical guidance that will help you make more informed decisions about the choice of data sources for your projects. It also provides a reference for the types, formats, and variety of spatial data that are available and some background information on the often complex organizations that collect and manage those data. The accompanying exercises include siting a business, assessing flood potential, evaluating land use changes, and analyzing the historical impact of hurricanes.

This book also aims to foster critical and spatial thinking. Critical thinking is a combination of skills and techniques that start with asking questions, acquiring and reviewing evidence, and lead to deducing sound and substantiated conclusions. In the context of GIS, critical thinkers need to be skeptical of data and the problem-solving strategies that are employed. They need to resist the temptation to accept the easiest-to-use datasets or the first ones they find, and instead identify only those most appropriate to solve their particular problem. Critical thinking also requires understanding motives and biases, both in ourselves and in the sources of information we use.

Spatial thinking means adopting a geographic perspective that considers scale, temporal and place-based relationships, patterns, linkages, and trends. Geography is not a body of content to be memorized. Facts and figures are important, but all successful GIS users, regardless of their fields of interest, need to employ geographic skills. These skills are honed through spatial analysis and the tools we used to interrogate the data such as summarizing tabular data, computing mean centers, intersecting, and buffering. At each stage in that analysis, step back from the functions and ask why each procedure is being performed and what effect each has on the data and on your final results. This ensures the geographic perspective—a way of looking at and understanding the world—is embedded throughout the process. The chapters and the exercises in this book are designed for you to develop your critical and spatial thinking skills.

The target audience for this book is anyone who is working with, or may be considering, public domain spatial data as part of GIS-based project. The level of GIS skill required to work through the exercises is intermediate, a working familiarity with ArcGIS for Desktop and perhaps some experience in file handling and data management. The exercises are not solely concerned with improving your analytical GIS skills. They are designed to encourage you to think about what types of data you need to solve a particular problem, the steps required to get the data into the correct format for your GIS, the quality of the data, and what other datasets you may need to address the problem. Other issues raised in the chapters are applied to real-world problems.

The text is written for university-level and other adult learners, including students, professors and researchers in universities, GIS practitioners in government agencies, nonprofit organizations, and private enterprise. We want you to develop the necessary
skills for finding, understanding, downloading, using, and ultimately solving problems with public domain spatial data. Throughout the course of the book, we will review the policies, issues, history, and current developments in the provision and use of spatial data in GIS-based analysis. Although many websites providing access to or information about spatial data are referenced in the book, these are not the only resources available. By necessity, we have chosen a limited number of examples to illustrate particular points.

Each chapter of the book focuses on a major theme, such as data types and formats, spatial data infrastructures, and planning GIS projects. Each chapter also includes a discussion of issues associated with public domain spatial data, such as whether organizations should charge fees or offer their data for free. Also included are additional references for those who wish to investigate issues on their own. The exercises associated with each chapter invite you to apply what you have learned to a real problem using public domain spatial data and ArcGIS software.

Given the range of topics and issues to be covered, perhaps the logical place to start is by introducing the concepts of data and information, including definitions of data, spatial data, and finally public domain spatial data.

Data versus information

Although many people use data and information interchangeably, these two terms actually convey very different concepts. Data are facts or figures which have been gathered systematically for one or more specific purposes. Data can exist in the form of linguistic expressions such as name or age, symbolic expressions such as traffic signs, mathematical expressions such as e=mc², or signals such as electromagnetic waves. Information, on the other hand, is data which have been processed into a form that is meaningful to a recipient and has perceived value in current or prospective decision making.

Although data are essential precursors for information, not all data make useful information. For example, data not properly collected, organized, or documented may be a hindrance rather than an asset to someone seeking to extract information. How useful data are also varies with each individual and application; data that generate useful information for one person’s application may not be useful for another application. Information is only useful to its recipients when it is relevant for its intended purposes and with the appropriate level of detail. Information is also only useful when it is
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 you can 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.
The public domain

Numerous definitions for the term *public domain* exist. Although common themes in most of these definitions include the lack of copyright, the freedom from intellectual property rights (IPR), and no requirement to pay royalties, there is no universally accepted definition of public domain.

The US Copyright Office (2011) states: “The public domain is not a place. A work of authorship is in the ‘public domain’ if it is no longer under copyright protection or if it failed to meet the requirements for copyright protection. Works in the public domain may be used freely without the permission of the former copyright owner.”

In the United Kingdom, the Intellectual Property Office (2011) defines public domain as “The body of works not or no longer protected by Intellectual Property rights which are available for the public to use without seeking permission or paying royalties.”

Creative Commons (2011), a nonprofit organization established in 2001 to make “it easier for people to share and build upon the work of others, consistent with the rules of copyright,” defines the public domain as “When a work is in the public domain, it is free for use by anyone for any purpose without restriction under copyright law. Public domain is the purest form of open/free, since no one owns or controls the material in any way.”

The United Nations Educational, Scientific, and Cultural Organization (UNESCO) (2011) defines the term as “Public domain information is publicly accessible information, the use of which does not infringe any legal right, or breach any other communal right (such as indigenous rights) or any obligation of confidentiality.”

The Linux Information Project (LINFO) (2011) defines it as “Public domain refers to the total absence of copyright protection for a creative work (such as a book, painting, photograph, movie, poem, article, piece of music, product design, or computer program).”

Wikipedia (2011) refers to public domain as “Works are in the public domain if they are not covered by intellectual property rights at all, if the intellectual property rights have expired, and/or if the intellectual property rights are forfeited.”

As a general rule, you should think of public domain spatial data as publicly accessible information about a spatial theme or phenomenon, the use of which does not infringe the legal rights of an individual or organization. Public domain spatial data encompasses all works or objects that may be used by anyone without any authorization.
Not only does public domain mean different things to different data users and data providers, the definition also varies from country to country. In fact, the concept of public domain simply does not exist in countries with no copyright laws, where the country owns all property including intellectual property, or where the country designates large amounts of information as state secrets or withheld in the interests of national security. As a further complication, copyright law is sometimes amended, for example when the US government extended copyright twenty years beyond the original terms stated by the Copyright Act of 1976. Such changes are controversial and subject to challenge, as in the case of *Eldred v. Ashcroft*, which sought to overturn the US government’s extension of the copyright period. As a result of the extension, a number of works were prevented from entering the public domain in 1998 and following years.

One area of difficulty applying national copyright laws to spatial data is that spatial data by their very nature often deal with phenomena that have no respect for human-made boundaries and borders. Which country's copyright laws apply in such international situations? Another difficulty in applying copyright to spatial data arises when datasets are derived from multiple sources. It is often difficult to trace which organization or individual originally gathered specific data, or to separate and identify discrete data items, and therefore to determine if those data should be copyrighted. Some spatial data also combine elements from old and new sources. Features such as contour lines from topographic maps may have been created in analog form on paper, vellum, or even copper plates decades or a century ago but may appear together now with new satellite imagery. What effect does the date of creation have on whether a work is considered to be in the public domain?

Although there are many similarities as to what the public domain means to different national bodies, there is a lack of consensus at the international level. If a work is made available in the public domain in one country, what effect does that have on its copyright status in another country? What you can do with a certain dataset in one country may not apply in another country, even though you are still working with the same data? The US Copyright Office posts the following advisory on its website: “Even if you conclude that a work is in the public domain in the United States, this does not necessarily mean that you are free to use it in other countries. Every nation has its own laws governing the length and scope of copyright protection, and these are applicable to uses of the work within that nation’s borders. Thus, the expiration or loss of copyright protection in the United States may still leave the work fully protected against unauthorized use in other countries.”
The GIS Guide to Public Domain Data

With respect to individual data providers, the terms and conditions for using those data are also varied. Companies such as Yahoo! and MapQuest have been providing web-based map services for a number of years. Although these data are generally available to the public, the data they provide access to are made available under proprietary licensing arrangements that impose certain restrictions on how the data may be used. Remember that there is a clear difference between what is publicly accessible versus what is in the public domain. Just because a dataset is downloadable does not place it in the public domain. Many photographs, for example, are available online today but are copyrighted, and some have a copyright symbol or watermark on the image itself. Spatial data are less likely to carry watermark or copyright symbols, but even if these are lacking, do not assume the data are in the public domain.

For Yahoo! Maps, the terms of use include:

“You agree to use the Data together with Yahoo! Maps solely for personal, non-commercial purposes for which you were licensed, and not for service bureau, time-sharing or other similar purposes. Accordingly, but subject to the restrictions set forth in the following paragraphs, you may copy this Data only as necessary for your personal use to (i) view it on your screen, (ii) print it, (iii) transfer a copy to a personal electronic device such as a personal digital assistant; (iv) save it, and (v) transfer a copy, in html form only, to a third party provided that you do not remove any copyright notices that appear and do not modify the Data in any way. You agree not to otherwise reproduce, copy, modify, decompile, disassemble or reverse engineer any portion of this Data, and may not otherwise transfer or distribute it in any form, for any purpose, except to the extent permitted by mandatory laws.”

Similarly, for MapQuest, the terms of the license are as follows:

“MapQuest grants you a nonexclusive, non-transferable license to view and print the Materials solely for your own personal non-commercial use. You may

For Further Reading
If you are interested in learning more about copyright and public domain in the United States, Cornell University (2011) provides a useful report, updated annually, identifying “works” that are in the public domain.
Chapter 1  •  Spatial data and the public domain

You may share a map or directions with another individual for that individual’s personal non-commercial use using the email option on the webpage. You may not commercially exploit the Materials or the underlying data, including without limitation, you may not create derivative works of the Materials, use any data mining, robots, or similar data gathering and extraction tools on the Materials, frame any portion of the Materials, or reprint, copy, modify, translate, port, publish, sublicense, assign, transfer, sell, or otherwise distribute the Materials without the prior written consent of MapQuest. You shall not derive or attempt to derive the source code or structure of all or any portion of the Materials by reverse engineering, disassembly, decompilation or any other means. You shall not use the Materials to operate a service bureau or for any other use involving the processing of data of others.”

As such, Yahoo! Maps and MapQuest Maps are not in the public domain. Similarly, for those using Google Maps and Google Earth, the following restrictions on use terms are imposed:

“Unless you have received prior written authorization from Google (or, as applicable, from the provider of particular Content), you must not:

A. access or use the Products or any Content through any technology or means other than those provided in the Products, or through other explicitly authorized means Google may designate (such as through the Google Maps/Google Earth APIs);

B. copy, translate, modify, or make derivative works of the Content or any part thereof;

C. redistribute, sublicense, rent, publish, sell, assign, lease, market, transfer, or otherwise make the Products or Content available to third parties;

D. reverse engineer, decompile or otherwise attempt to extract the source code of the Service or any part thereof, unless this is expressly permitted or required by applicable law;

E. use the Products in a manner that gives you or any other person access to mass downloads or bulk feeds of any Content, including but not limited to numerical latitude or longitude coordinates, imagery, and visible map data;

F. delete, obscure, or in any manner alter any warning, notice (including but not limited to any copyright or other proprietary rights notice), or link that appears in the Products or the Content; or

G. use the Service or Content with any products, systems, or applications for or in connection with (i) real time navigation or route guidance, including but not limited to turn-by-turn route guidance that is synchronized to the position of a
user’s sensor-enabled device; or (ii) any systems or functions for automatic or autonomous control of vehicle behavior.”

Google Maps and Google Earth are frequently used for thousands of everyday tasks, even providing context mapping for the display of other data that may be in the public domain. However, as with Yahoo! Maps and MapQuest Maps, Google map products are not free from restrictions on use and are not in the public domain.

You should also be aware of situations where a public domain spatial data source becomes part of another data source. OneGeology, an international geological spatial data online resource, was established in 2007 to “create dynamic geological map data of the world, available to everyone via the web.” Most of the 116 contributors to the site are national geological mapping agencies, such as the USGS, the British Geological Survey, the Geological Survey of Pakistan, and the Geological Survey of Japan. The terms and conditions that govern access to data hosted on the site warn against copyright infringement if the data are used for commercial purposes. As such, this compilation of data from OneGeology is not in the public domain. However, some of the individual datasets that make up OneGeology may indeed be in the public domain. The OneGeology (2011) use agreement states “Where you intend to use the material commercially (e.g. in book, to sell as a map extract, etc.) you will need to pay the appropriate copyright holder which will be the owner of the geological information from the country concerned.” Although hosted by an international consortium and publicly available to anyone with Internet access, the data remain subject to the individual copyright restrictions of each national mapping and data collection agency. Regardless of them being a component of OneGeology, the USGS data are in the public domain because the data are produced by a US federal agency under section 105 of the Copyright Law of the United States. You should be aware of the distinction in copyright terms between compilations and the items that make up those compilations which may have different use constraints.

One example of an online spatial data resource in the public domain is the Natural Earth website. Natural Earth data were created by cartographers who collaborated in a project supported by the North American Cartographic Information Society (NACIS). Natural Earth provides global small-scale raster and vector maps, based on three themes:

- Cultural, including countries, urban areas and administrative boundaries.
- Physical, including land, coastline, rivers and lakes.
- Raster, including shaded relief, bathymetry, and ocean.

The maps are at three different scales: 1:10,000,000; 1:50,000,000; and 1:110,000,000. The terms of use are as follows:
“All versions of Natural Earth raster + vector map data found on this website are in the public domain. You may use the maps in any manner, including modifying the content and design, electronic dissemination, and offset printing. The primary authors . . . and all other contributors renounce all financial claim to the maps and invite you to use them for personal, educational, and commercial purposes.”

Another example of an online spatial data resource is the North American Environmental Atlas, compiled by the Commission for Environmental Cooperation (CEC). The CEC was set up around the time the North American Free Trade Agreement (NAFTA) took effect, and includes the same countries that participate in NAFTA: Canada, the United States of America, and Mexico. The commission was established by a multinational agreement to address regional environmental concerns, help prevent potential conflicts between trade and the environment, and to promote the effective enforcement of environmental law. The commission created the 1:10,000,000-scale North American Environmental Atlas, a public domain dataset with an accompanying data viewer. Although data viewers provide a quick and easy way to inspect the data, they are of limited use for most GIS users as many do not support any analytical capabilities. However, the North American Environmental Atlas goes one step further in providing data files that can be downloaded and used by the GIS community. As the atlas contains data from a variety of sources, the terms of use restrictions provided on the CEC website accounts for these sources as follows:

“The creator of this data is outlined in the metadata and at the bottom left corner of each MXD or GeoPDF that is made available through the Commission for Environmental Cooperation (CEC). If the author of this data is the CEC: The user is permitted to use this data for non-commercial, non-sublicensable purposes. The user is allowed non-exclusive rights to represent this data as desired. If the author of the data is noted otherwise: Please refer to the license information of this individual dataset, or contact that organization or individual if you are unsure of your rights regarding redistribution, use, or license information of this particular dataset.”

The commission also provides disclaimers on the quality of the data in the atlas: “The CEC does not ensure that this is the most updated representation of this information, although updates will be conducted on an irregular basis. We make every effort to provide and maintain accurate, complete, usable, and timely information on our Web pages. However, some CEC data and information accessed through these pages may, of necessity, be preliminary in nature. These data and information are provided with the understanding that they are
not guaranteed to be complete. Users are cautioned to consider carefully the provisional nature of these data and information before using them for decisions that concern personal or public safety or the conduct of business that involves substantial monetary or operational consequences. Conclusions drawn from, or actions undertaken on the basis of, such data and information are the sole responsibility of the user.”

The commission’s license arrangements define how the data should be cited, documents the limitations of the data, and indicates who may use the data and for what purpose. The intention is for the user to not assume that every area feature is up-to-date. The data are in the public domain and are to be used as such. In common with many similar organizations, there are restrictions on the use of certain datasets for commercial purposes, just as there are restrictions on certain software packages for educational versus commercial use.

The few examples described here demonstrate that although a great deal of spatial data are publicly available, those data are not always in the public domain. Anyone wishing to make use of a publicly available spatial data resource should always read the label. That said, the amount of available public domain spatial data has expanded rapidly in recent years and has become an integral part of the evolving GIS community. However, no discussion of the public domain would be complete without also considering copyright and licensing, a subject we touched on briefly earlier but which requires some elaboration.

**FOR FURTHER READING**

See James Boyle’s *The Public Domain: Enclosing the Commons of the Mind* (Caravan Books 2008). Boyle is a professor at Duke University School of Law and founder of the Center for the Study of the Public Domain. This book is available online under a Creative Commons Attribution-Noncommercial-Share Alike License.
Copyright

What is copyright? In its simplest form, copyright is a form of protection provided by laws to authors who create original works. These works can be art, music, drama, literature, and—most important to GIS users—data. Copyright laws grant the authors of these works certain exclusive rights to do such things as use the works, reproduce the works, prepare derivative works, and distribute copies. Copyright laws also give authors the right to authorize who may copy and use the works they created, and when they may do so. Registration of the copyright brings the author additional legal rights (Patterson and Lindberg 1991).

Section 105 of the Copyright Law of the United States, titled “Subject matter of copyright: United States Government works,” has important implications for GIS users. It states that “Copyright protection under this title is not available for any work of the United States Government.” As the works of the US government, including spatial data, cannot be copyrighted, the spatial data created by US federal agencies have been available to and used by the GIS community. This particular piece of legislation resulted in an entire industry built upon government spatial data, as we shall see in this book, through its impact on privacy and fee structures. The amount of spatial data generated by federal agencies in the United States dwarfed the amount of data from federal agencies in other countries where, with a few exceptions, data are subject to copyright and licensing restrictions. This situation is, however, changing and will be discussed in greater detail in chapter 4.

Another aspect we will explore is that section 105 only applies to US federal government works; state and local spatial data can be copyrighted and licensed. A landmark case in the application of copyright in the United States was Feist Publications Inc. v. Rural Telephone Service Co. The case centered on the licensing of telephone directories and an alleged copyright infringement. The court decided that the listings contained in the directory were not subject to copyright so no infringement had taken place. By implication, facts alone were not copyrightable; a “minimal degree” of creativity must be involved to make something copyrightable. As such, any private company that enhances, or adds value to, US government spatial data may indeed copyright those data.

In 1989, the United States became a signatory to the Berne Convention for the Protection of Artistic and Literary Works, the worldwide agreement on copyright protection dating back to 1886. The Berne Convention provisions, which are codified by every adopting country, set forth the principle that once a work is deemed completed by its author, the work is automatically copyrighted. This principle was established to
overcome loss of copyright protection due to the filing and registration technicalities that some early copyright laws (such as the 1909 US Copyright Act) invoked. As a general rule, everything is copyrighted unless there is explicit language to the contrary. That explicit language is an affirmative statement that the author is making the work available free of copyright and in the public domain. Works come into the public domain for two major reasons; the duration or term of copyright lapses or expires or the author specifically places the work in the public domain.

It is illegal for anyone to violate any of the rights provided by the copyright law to the owner of copyright. However, the rights are not unlimited. For example, the US Copyright Act of 1976 established some exemptions, including one that affects many GIS data users: the doctrine of fair use (section 107). Fair use is a difficult concept to define, but four factors in US law help determine if a particular use is fair. These factors include the purpose and character of the proposed use (such as commercial or educational), the nature of the copyrighted work, the amount used compared to the copyrighted work as a whole, and the effect of the use upon the potential market or value of the work. Another example of a limitation of copyright law occurs when royalties are paid on some works and conditions are agreed to by the purchaser, the purchaser may use the copyrighted work under those conditions.

Creative Commons licensing and the public domain

What does a public domain license really mean and how it is administered? As we discussed earlier, the Creative Commons organization was specifically established to promote and facilitate access to what they refer to as “creativity (cultural, educational, and scientific content) in ‘the commons.’” To support this initiative, Creative Commons provides a set of licensing options for individuals and organizations alike to grant the required level of access to creative work produced by those individuals or organizations. The level of access may range from full copyright protection and all rights reserved to certified as public domain and no rights reserved. A summary of the main licenses and conditions, starting with the least restrictive and ending with the most restrictive, is provided in the following:
**Attribution (CC BY):** Others may distribute, remix, tweak, build on the work, and use for commercial purposes as long as originator is given credit. http://creativecommons.org/licenses/by/3.0/

**Attribution-ShareAlike (CC BY-SA):** Others may distribute, remix, tweak, build on the work, and use for commercial purposes as long as originator is given credit and all new creations are licensed under identical terms. http://creativecommons.org/licenses/by-sa/3.0/

**Attribution-NoDerivs (CC BY-ND):** Allows for redistribution, commercial and noncommercial, as long as it is unchanged and in whole, and originator is given credit. http://creativecommons.org/licenses/by-nd/3.0/

**Attribution-NonCommercial (CC BY-NC):** Others remix, tweak, and build upon the work noncommercially, and although their new works must also acknowledge the originator and be noncommercial, they don’t have to license their derivative works on the same terms. http://creativecommons.org/licenses/by-nc/3.0/

**Attribution-NonCommercial-ShareAlike (CC BY-NC-SA):** Others may remix, tweak, and build upon the work noncommercially, as long as they credit the originator and license their new creations under the identical terms. Others can download and redistribute the work, translate, make remixes, and produce other work based on
the work. All new work will carry the same license, so any derivatives will also be noncommercial. http://creativecommons.org/licenses/by-nc-sa/3.0/

![Attribution-NonCommercial-NoDerivs (CC BY-NC-ND):](https://creativecommons.org/licenses/by-nc-nd/3.0/)

Others may download the work and share with others as long as they mention and link back to the originator. The work cannot be altered in any way or used commercially. http://creativecommons.org/licenses/by-nc-nd/3.0/ (Source: http://creativecommons.org.)

For those individuals and organizations that may wish to make their work available with no conditions, or to have it certified as in the public domain, Creative Commons introduced the concept of CC0: “CC0 which in effect removes all copyright from a piece of work or data with a No Rights Reserved license.” CC0 allows content originators and owners to waive copyright interests in their work and place them in the public domain worldwide.

Creative Commons has also recognized the need for global consensus in the interpretation and application of these licensing arrangements and is working to translate and adapt these agreements to international copyright jurisdictions.

---

**FOR FURTHER READING**

For a complete list of the countries that now have their licensing arrangements integrated into the Creative Commons licensing process, see the CC Affiliate Network on the Creative Commons website.
To assist data providers in creating an affirmative statement that the author is placing the work in the public domain, Creative Commons offers a Copyright-Only Dedication (based on US law) or Public Domain Certification. The wording reads as follows:

“The person or persons who have associated work with this document (the ‘Dedicator’ or ‘Certifier’) hereby either (a) certifies that, to the best of his knowledge, the work of authorship identified is in the public domain of the country from which the work is published, or (b) hereby dedicates whatever copyright the dedicatores holds in the work of authorship identified below (the ‘Work’) to the public domain. A certifier, moreover, dedicates any copyright interest he may have in the associated work, and for these purposes, is described as a ‘dedicator’ below. A certifier has taken reasonable steps to verify the copyright status of this work. Certifier recognizes that his good faith efforts may not shield him from liability if in fact the work certified is not in the public domain. Dedicator makes this dedication for the benefit of the public at large and to the detriment of the Dedicator’s heirs and successors. Dedicator intends this dedication to be an overt act of relinquishment in perpetuity of all present and future rights under copyright law, whether vested or contingent, in the Work. Dedicator understands that such relinquishment of all rights includes the relinquishment of all rights to enforce (by lawsuit or otherwise) those copyrights in the Work. Dedicator recognizes that, once placed in the public domain, the Work may be freely reproduced, distributed, transmitted, used, modified, built upon, or otherwise exploited by anyone for any purpose, commercial or non-commercial, and in any way, including by methods that have not yet been invented or conceived.”

Two examples of publicly available, but not public domain, spatial data sources provided under a Creative Commons licensing arrangement are OpenStreetMap (OSM) and European Environment Agency (EEA). OSM is a collaborative project that aims to develop and provide a free, editable, map of the world. XML exports of the data are available for use under Creative Commons Attribution-Share Alike 2.0 license. EEA collates and produces integrated environmental datasets for the European Union. Data are available to download from the EEA website under a Creative Commons Attribution license. We will return to the OSM project in chapter 8 when we cover crowdsourced data.

The Creative Commons approach to licensing is not without disadvantage; it was never developed with spatial data and databases in mind, nor with the attendant issues of combined data stores and generating derived data. Another valid approach may be the Open Database License proposed by the OSM community, which would see the
introduction of an Attribution Share Alike (CC BY-SA) license for databases. Although the Creative Commons approach is perhaps better suited to literary works rather than data, and Creative Commons does not recommend its license model for software, at present the Creative Commons approach appears to be emerging as the de facto licensing arrangement for spatial data.

The web and public domain data portals

Since its inception in the 1960s, GIS, like other aspects of information technology in the digital age, has undergone a rapid evolution. Over the same time span, data have also changed—their formats, resolution, themes and types, how they are stored, and how they are served. The main platform providing access to data—the web—is also changing rapidly, almost daily. Throughout the course of this book, you will be referred to websites and some of the URLs will have changed or are not available. If they are available, realize that some layouts and navigation tools change frequently. Be flexible, creative and, above all, patient when working with these websites. We have tried to use sites that are relatively stable and have provided the high-level URLs to each site’s home page, along with directions on how to navigate to the individual data depositories (specific URLs change frequently). However, you may still find some sites are not available from time to time. If a particular site is not available, search for other sites that may host the same, or similar, data. You could try a different web browser, check your security settings, try again later, or contact the site administrator if the problem persists. If the interface has changed, use the site’s search facilities to locate the resources you require. As an additional resource, a supplementary website to support this book will also be available, where recent updates and any major changes will be posted on a regular basis.

Summary and looking ahead

In chapter 1, we have introduced and defined public domain spatial data and discussed some of the licensing, access, and copyright issues associated with these data. By now
you should appreciate that although a vast amount of information is publicly available, not all of it is in the public domain. This is an important distinction to keep in mind as you read this book and, more importantly, as you search for and use spatial data in your projects.

Chapter 2 begins with defining spatial data models and then focuses on vector data, including transportation, soils, hydrography, land parcel, agriculture, biomes, hazards, names, and demography. The section on data quality discusses some of the issues in defining what quality means.

Chapter 3 focuses on raster data portals, sources, and quality; in particular, elevation, land cover, imagery, climate, and population. It also examines some of the privacy issues associated with spatial data.

Chapter 4 begins with a focus on the true costs of spatial data and reviews the ongoing debate as to whether public domain spatial data should be free or fees charged for their use. The chapter will also consider whether spatial data should be subject to copyright and government policies, both existing and emerging.

How do data portals, national and state, and metadata standards affect your use of public domain spatial data? Chapter 5 focuses on these issues and explores the national spatial data infrastructure initiatives in a selected number of countries, providing a discussion on metadata and the issues surrounding data sharing.

Moving to the international scale is the focus of chapter 6, which looks at some of the issues involved and a consideration of whether we are moving toward universal access to global data.

By chapter 7, you will have covered most of the main issues and had enough practice with the exercises to “put public domain spatial data to work.” In this chapter, you will examine some of your decisions as you locate and work with spatial data, and how to evaluate those decisions at each step of your project.

The last three chapters will focus on new initiatives that are already affecting GIS data and will have an increasing influence. Chapter 8 covers the dual role of data user and data provider. Chapter 9 discusses how the cloud computing revolution—something you have undoubtedly heard much about—is set to deliver the next generation of GIS data and services and what this will mean for data and GIS, with examples using ArcGIS online and elsewhere.

Finally, chapter 10 reflects on the future of public domain data, the emerging issues and initiatives, and how these will affect data quality, data availability, privacy, licensing, copyright, metadata, and user choice.
References


