appropriate spatial sampling schemes and creating improved representations, which tell us still more about the real world and how we might represent it.

Spatial data provide the foundations for operational and strategic applications of GI, and these foundations must be developed creatively, yet rigorously, if they are to support the spatial analysis superstructure that we wish to erect on them. This entails much more than technical competence with software. An understanding of the nature of geographic data allows us to use induction (reasoning from observations) and deduction (reasoning from principles and theory) alongside each other to develop effective spatial representations that are safe to use.

Questions for Further Study

1. Many jurisdictions tout the number of miles of shoreline in their community—for example, Ottawa County, Ohio, claims 107 miles of Lake Erie shoreline. What does this mean, and how could you make it more meaningful?
2. With reference to Figure 2.11, list the design considerations that should be incorporated into GI software to measure accessibility of (a) a neighborhood medical center to wheelchair-bound pedestrians, (b) a grocery store to high-income customers, and (c) all residential buildings in a small town from a single fire service station.
3. The apparatus of inference was developed by statisticians because they wanted to be able to reason from the results of experiments involving small samples to make conclusions about the results of much larger, hypothetical experiments—for example, in using samples to test the effects of drugs. Summarize the problems inherent in using this apparatus for geographic data, in your own words.
4. "Can geometry deliver what the Greek root of its name [geo-] seemed to promise—truthful measurement, not only of cultivated fields along the Nile River but also of untamed Earth?" Discuss this challenge posed by Mandelbrot, offered in his 2012 autobiography, *The Fractalist: Memoir of a Scientific Maverick* (New York: Pantheon Books, p. xii).

Further Reading

Batty, M. and Longley, P. A. 1994. *Fractal Cities: A Geometry of Form and Function*. London: Academic Press. Available for download (free of charge) from www.fractalcities.org.

De Smith, M., Goodchild, M. F., and Longley, P. A. 2009. *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*. Chapter 5: Data Exploration and Spatial Statistics. Leicester: Troubador. See also www.spatialanalysisonline.com.

Goodchild, M. F. 2004. The validity and usefulness of laws in geographic information science and geography. *Annals of the Association of American Geographers* 94(2): 300–304.

Haklay, M. 2010. *Interacting with Geospatial Technologies*. Chichester: Wiley.

Harvey, F. J., Tobler, W., Chrisman, N. R., Frank, A. U., Sui, D., and Goodchild, M. F. 2012. *Are There Fundamental Principles in Geographic Information Science?* www.createspace.com.

Mandelbrot, B. B. 1984. *The Fractal Geometry of Nature*. San Francisco: Freeman.

3 Representing Geography

This chapter introduces the concept of representation, or the construction of a digital model of some aspect of the Earth's surface. Representations have many uses, allowing us to learn, think, and reason about places and times that are outside our immediate experience. This is the basis of scientific research, planning, and many forms of day-to-day problem solving.

The geographic world is extremely complex, revealing more detail the closer one looks, seemingly ad infinitum. In order to build a representation of any part of it, it is necessary to make choices about what to represent, at what level of detail, and over what time period. The large number of possible choices creates many opportunities for designers of geographic information (GI) systems.

Generalization methods are used to remove detail that is unnecessary for an application, in order to reduce data volume and speed up operations.

LEARNING OBJECTIVES

After studying this chapter you will understand:

- What representation is and why it is important.
- The concepts of fields and objects and their fundamental significance.
- What raster and vector representations entail, and how these data structures affect many GI principles, techniques, and applications.
- The similarities and differences between online map services and paper maps.
- Why map generalization methods are important, and how they are based on the concept of representational scale.
- The art and science of representing real-world phenomena in GI databases.

3.1 Introduction

We live on the surface of the Earth and spend most of our lives in a relatively small fraction of that space. Of the approximately 500 million sq. km of surface, only one-third is land, and only a fraction of that is occupied by the cities and towns in which most of us live. The rest of the Earth, including the parts we never visit, the atmosphere, and the solid ground under our feet, remains unknown to us except through the information that is communicated to us through books, newspapers, television, the Web, or the spoken word. We live lives that are almost infinitesimal in comparison with the 4.5 billion years of Earth history, or the over 10 billion years since the universe began, and we know about the Earth as it was before we were born only through the evidence compiled by geologists, archaeologists, historians, and other specialists. Similarly, we know nothing about the world that is to come, where we have only predictions to guide us.

Because we can observe so little of the Earth directly, we rely on a host of methods for learning about its other parts, deciding where to go as tourists or shoppers, choosing where to live, running the operations of corporations, agencies, and governments, and many other activities. Almost all human activities at some time require knowledge (see Section 1.2) about parts of the Earth that are outside our direct experience because they occur either elsewhere in space or elsewhere in time.

Sometimes this knowledge is used as a *substitute* for directly sensed information, creating a *virtual* reality. Increasingly, it is used to *augment* what we can see, touch, hear, feel, and smell through the use of mobile information systems that can be carried around (see Section 12.4.3). Our knowledge of the

Earth is not created entirely freely, but must fit with the mental concepts we began to develop as young children—concepts such as containment (Paris is *in* France) or proximity (Dallas and Fort Worth are *close*). In digital representations, we formalize these concepts through *data models* (Chapter 7), the structures and rules that are programmed into a GI system to accommodate data. These concepts and data models together constitute our *ontologies*, the frameworks that we use for acquiring knowledge of the world.

Almost all human activities require knowledge about the Earth—past, present, or future.

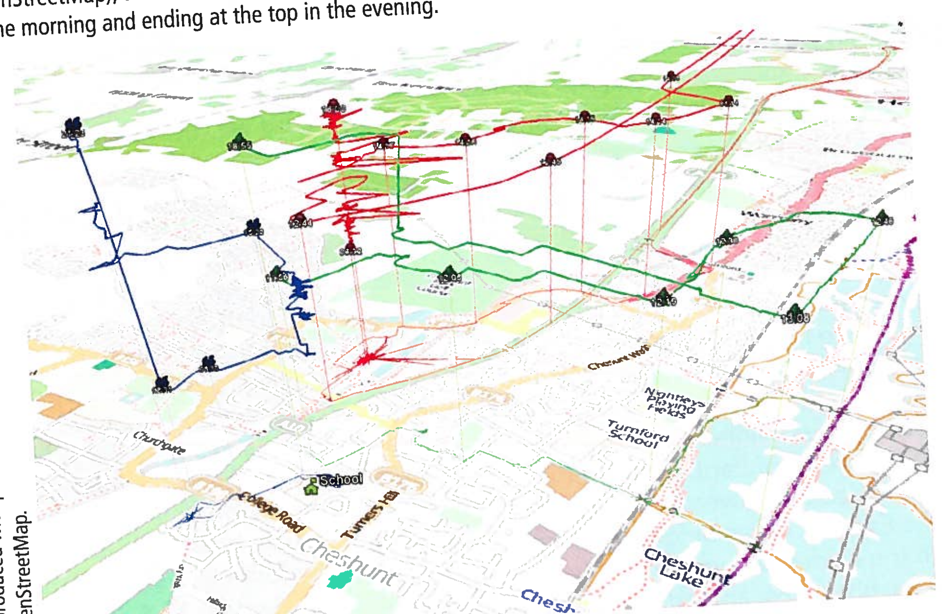
One such ontology, a way to structure knowledge of movement through time, is a three-dimensional diagram, in which the two horizontal axes denote a location on the Earth's surface, and the vertical axis denotes time. Figure 3.1 presents the *time space aquarium* that largely contains the activity space of three children living in Cheshunt, UK. The icons on the three trajectories identify the travel modes (walking, cycling, or automobile) that they use as they move through space (the base of the aquarium is rendered using OpenStreetMap: see Box 3.5) and time (the vertical dimension) on a single weekend day. The spatial and temporal granularity of the diagram is set in such a way that even quite small events are recorded, but the granularity could be reduced, for example, by taking only a single GPS signal on the hour or by recording only those changes in location that exceed a predetermined distance from the preceding point. When we view this diagram on the page of a book, much of the fine detail of the three activity patterns is lost: it is not possible to discern, for example, the precise trajectory of a child playing football on a field, or whether the cyclist dismounts for short sections of an uphill journey, or how long a car waits at traffic lights before continuing its journey. Indeed, some of these details cannot be recovered even if the GPS is set to record changes in location every 5 minutes. Closer perspectives could display more information, but such detail rarely adds much that is useful, and a vast storehouse would be required to capture the precise trajectories of many children throughout even a single day.

The real trajectories of the individuals shown in Figure 3.1 are complex, yet the figure is only a representation of them—a model on a piece of paper or computer screen, generated by a computer from a database. We use the terms *representation* and *model* because they imply a simplified relationship between the contents of the figure and the database and the real-world trajectories of the individuals.

Such representations or models serve many useful purposes and occur in many different forms. For example, representations occur

- in the human mind, when our senses capture information about our surroundings, such as the images captured by the eye, or the sounds captured by the ear, and memory preserves such representations for future use;
- in photographs, which are two-dimensional models of the light emitted or reflected by objects in the world into the lens of a camera;
- in spoken descriptions and written text, in which people describe some aspect of the world in language, in the form of travel accounts or diaries; or
- in the numbers that result when aspects of the world are measured, using such devices as thermometers, rulers, or speedometers.

Figure 3.1 Schematic representation of weekend activities of three children in Cheshunt, UK. The horizontal dimensions represent geographic space (rendered using OpenStreetMap), and the vertical dimension represents time of day. Each person's track plots as a three-dimensional line, beginning at the base in the morning and ending at the top in the evening.



Reproduced with permission: Yi Gong; base image courtesy OpenStreetMap.

By building representations, we humans can assemble far more knowledge about our planet than we ever could as individuals. We can build representations that serve such purposes as planning, resource management and conservation, travel, or the day-to-day operations of a parcel delivery service.

Representations help us assemble far more knowledge about the Earth than is possible on our own.

Representations are reinforced by the rules and laws that we humans have learned to apply to the unobserved world around us. When we encounter a fallen log in a forest, we are willing to assert that it once stood upright and once grew from a small shoot, even though no one may actually have observed or reported either of these stages. We predict the future occurrence of eclipses based on the laws we have discovered about the motions of the Solar System.

3.2 Digital Representation

This book is about one particular form of representation that has become very important in our society—representation in digital form. Today, almost all communication between people through such media as the telephone, Web pages, microblogs, music, television, newspapers and magazines, or e-mail is at some time in its life in digital form. Information technology based on digital representation is moving into all aspects of our lives, from science to commerce to daily existence. The smartphone is the most pervasive digital information-processing device, and it is estimated that as of June 2013 there were 2.1 billion mobile Internet users (29.5% of the global population). Computers are also the mainstay of most office work, and digital technology has pervaded many devices that we use every day, from the microwave oven to the automobile.

One important characteristic of digital technology is that the representation itself is rarely if ever seen by the user because only a few technical experts ever see the individual elements of a digital representation. What we see instead are *views* designed to present the contents of the representation in a form that is meaningful to us. The term *digital* derives from *digits*, or the fingers, and our system of counting based on the 10 digits of the human hand. But although the counting system has 10 symbols (0 through 9), the representation system in digital computers uses only two (0 and 1). In a sense, then, the term *digital* is a misnomer for a system that represents all information using some combination of the two symbols 0 and 1, and the more exact term *binary* is more appropriate. In this book we follow the convention of using *digital* to refer to electronic technology based on binary representations.

Computers represent phenomena as binary digits. Every item of useful information about the Earth's surface is ultimately reduced by a GI database to some combination of 0s and 1s.

Over the years many standards have been developed for converting information into digital form. Box 3.1 shows the standards that are commonly used to store GI, whether they consist of whole or decimal numbers or text. There are many competing coding standards for images and photographs (GIF, JPEG, TIFF, etc.) and for movies (e.g., MPEG) and sound (e.g., MIDI, MP3). Much of this book is about the coding systems used to represent geographic data, especially Chapter 7, and as you might guess, this turns out to be quite complicated.

Digital technology is successful for many reasons, not the least of which is that all kinds of information share a common basic format (0s and 1s) and can be handled in ways that are largely independent of their actual meaning (see Box 3.1). The Internet, for example, operates on the basis of packets of information, consisting of strings of 0s and 1s, which are sent through the network based on the information contained in the packet's header. The network needs to know only what the header means and how to read the instructions it contains regarding the packet's destination. The rest of the contents are no more than a collection of bits, representing anything from an e-mail message to a short burst of music or highly secret information on its way from one military installation to another and are almost never examined or interpreted during transmission. This allows one digital communications network to serve every need, from electronic commerce to social networking sites, and it allows manufacturers to build processing and storage technology for vast numbers of users who have very different applications in mind. Compare this to earlier ways of communicating, which required printing presses and delivery trucks for one application (newspapers) and networks of copper wires for another (telephone).

Technical Box 3.1

The Binary Counting System

The binary counting system uses only two symbols, 0 and 1, to represent numerical information. A group of eight binary digits is known as a *byte*, and volume of storage is normally measured in bytes rather than bits (see Table 1.1). There are only two options for a single digit, but there are four possible combinations for two digits (00, 01, 10, and 11), eight possible combinations for three digits (000, 001, 010, 011, 100, 101, 110, 111), and 256 combinations for a full byte. Digits in the binary system (known as binary digits, or *bits*) behave like digits in the decimal system but use powers of two. The rightmost digit denotes units, the next digit to the left denotes twos, the next to the left denotes fours, and so on. For example, the binary number 11001 denotes one unit, no twos, no fours, one eight, and one sixteen and is equivalent to 25 in the normal (decimal) counting system. We call this the *integer* digital representation of 25 because it represents 25 as a whole number and is readily amenable to arithmetic operations. Whole numbers are commonly stored in GI databases using either *short* (2-byte or 16-bit) or *long* (4-byte or 32-bit) options. Short integers can range from –32767 to +32767, and long integers from –2147483647 to +2147483647.

The 8-bit ASCII (American Standard Code for Information Interchange) system assigns codes to each symbol of text, including letters, numbers, and common symbols. The number 2 is assigned ASCII code 50 (00110010 in binary), and the number 5 is 53 (00110101), so if 25 were coded as two characters using 8-bit ASCII its digital representation would be 16 bits long (0011001000110101). The characters 2=2 would be coded as 50, 61, 50 (001100100011110100110010). ASCII is used for coding text, which consists of mixtures of letters, numbers, and punctuation symbols.

Numbers with decimal places are coded using *real* or *floating-point* representations. A number such as 123.456 (three decimal places and six significant digits) is first transformed by powers of 10 so that the decimal point is in a standard position, such as the beginning (e.g., 0.123456 x 10^3). The fractional part (0.123456) and the power of 10 (3) are then stored in separate sections of a block of either 4 bytes (32 bits, *single* precision) or 8 bytes (64 bits, *double* precision). This gives enough precision to store roughly 7 significant digits in single precision, or 14 in double precision.

Integer, ASCII, and real conventions are adequate for most data, but in some cases it is desirable to associate images or sounds with places in GI databases, rather than text or numbers. To allow for this, GI system designers have included a BLOB option (standing for binary large object), which simply allocates a sufficient number of bits to store the image or sound, without specifying what those bits might mean.

Digital representations of geography hold enormous advantages over previous types—paper maps, written reports from explorers, or spoken accounts. We can use the same cheap digital devices—the components of smartphones, PCs, the Internet, or mass storage devices—to handle every type of information, independent of its meaning. Digital data are easy to copy, they can be transmitted at close to the speed of light, they can be stored at high density in very small spaces, and they are less subject to the physical deterioration that affects paper and other physical media.

Perhaps more important, data in digital form are easy to transform, process, and analyze. GI systems allow us to do things with digital representations that we were never able to do with paper maps: to measure accurately and quickly, to overlay and combine, and to change scale, zoom, and pan without respect to map sheet boundaries. The vast array of possibilities for processing that digital representation opens up is reviewed in Chapters 13 through 15 and is also covered in the applications that appear throughout this book.

Digital representation has many uses because of its simplicity and low cost.

3.3 Representation of What and for Whom?

Thus far we have seen how humans are able to build representations of the world around them, but we have not yet discussed why representations are useful, and why humans have become so ingenious at creating and sharing them. The emphasis here and throughout the book is on one type of representation, termed *geographic*, and defined as a representation of some part of the Earth's surface or near-surface, at scales ranging from the architectural to the global.

Geographic representation is concerned with the Earth's surface or near-surface at scales from the architectural to the global.

Geographic representations are among the most ancient, having their roots in the needs of very early societies. The tasks of hunting and gathering can be much more efficient if hunters are able to communicate the details of their successes to other members of their group—the locations of edible roots or game, for example. Maps must have originated in the sketches early people made in the dirt of campgrounds or on cave walls, long before language became sufficiently sophisticated to convey equivalent information through speech. We know that the peoples of the Pacific built representations of the locations of islands, winds, and currents from simple materials to guide each other, and that social insects such as bees use very simple forms of representation to communicate the locations of food resources.

Hand-drawn maps and speech are effective media for communication between members of a small group, but much wider communication became possible with the invention of the printing press in the Fifteenth Century. Now large numbers of copies of a representation could be made and distributed, and for the first time it became possible to imagine that something could be known by every human being—that knowledge could be the common property of humanity. Only one major restriction affected what could be distributed using this new mechanism: the representation had to be flat. If one were willing to accept that constraint, however, paper proved to be enormously effective; it was cheap, light and thus easily transported, and durable. Only fire and water proved to be disastrous for paper, and human history is replete with instances of the loss of vital information through fire or flood, from the burning of the Alexandria Library in the Seventh Century that destroyed much of the accumulated knowledge of classical times to the major conflagrations of London in 1666, San Francisco in 1906, or Tokyo in 1945, and the flooding of the Arno that devastated Florence in 1966.

One of the most important periods for geographic representation began in the early Fifteenth Century in Portugal and had parallels in China. Henry the Navigator (Box 3.2) is often credited with originating the Age of Discovery, the period of European history that led to the accumulation of large amounts of information about other parts of the world through sea voyages and land explorations. Maps became the medium for sharing information about new discoveries and for administering vast colonial empires, and their value was quickly recognized. Although detailed representations now exist of all parts of the world, including Antarctica, in a sense the spirit of the Age of Discovery continues in the explorations of the oceans, caves, and outer space, and in the constant process of remapping that is needed to keep up with frequent changes in the human and natural worlds.

It was the creation, dissemination, and sharing of accurate representations that distinguished the Age of Discovery from all previous periods in human history (and it would be unfair to ignore its distinctive negative consequences, notably the spread of European diseases and the growth of the slave trade). Information about other parts of the world was assembled in the form of maps and journals, reproduced in large numbers using the recently invented printing press, and distributed on paper. Even the modest costs associated with buying copies were eventually addressed through the development of free public lending libraries in the Nineteenth Century, which gave access to virtually everyone. Today, we benefit from what is now a long-standing tradition of free and open access to much of humanity's accumulated store of knowledge about the geographic world, in the form of paper-based representations, through the institution of libraries and the copyright doctrine that gives people rights to material for personal use (see Chapter 17 for a discussion of laws affecting ownership and access). The Internet is the present-day delivery channel that provides distributed access to geographic information through online virtual Earths (specifically those of Google, Apple, and Microsoft) and other GI services. In several countries, the Open Data (Section 17.4) initiatives of recent years have done much to provide individuals and organizations with the digital materials to produce their own representations, and wide dissemination of open software and GPS-enabled technologies has empowered communities to create mapping representations for general or specific community purposes (see Box 3.5).

In the Age of Discovery, maps became extremely valuable representations of the state of geographic knowledge.

Not surprisingly, representation also lies at the heart of our ability to solve problems using the digital tools that are available in GI systems. Any application of GI requires clear attention to questions of *what* should be represented and *how*. There is a multitude of possible ways of representing the geographic world in digital form, none of which is perfect and none of which is ideal for all applications.

The key GI representation issues are what to represent and how to represent it.

One of the most important criteria for the usefulness of a representation is its *accuracy*. Because the geographic world is seemingly of infinite complexity, choices are always to be made in building

Biographical Box 3.2

Prince Henry the Navigator and Admiral Zheng He

Prince Henry of Portugal (Figure 3.2), who died in 1460, was known as Henry the Navigator because of his keen interest in exploration. In 1433 Prince Henry sent a ship from Portugal to explore the west coast of Africa in an attempt to find a sea route to the Spice Islands. This ship was the first from Europe to travel south of Cape Bojador (latitude 26 degrees 20 minutes North). To make this and other voyages Prince Henry assembled a team of mapmakers, sea captains, geographers, ship builders, and many other skilled craftsmen. Prince Henry showed the way for Vasco de Gama and other famous Fifteenth-Century explorers. His management skills could be applied in much the same way in today's GI projects.

Figure 3.2 Prince Henry the Navigator, originator of the Age of Discovery in the Fifteenth Century and promoter of a systematic approach to the acquisition, compilation, and dissemination of geographic knowledge.

Admiral Zheng was born in 1371 in what is now China's Yunnan Province. In a series of seven expeditions between 1405 and 1433, he explored the coasts of Thailand, Indonesia, India, Arabia, and East Africa, using massive fleets of up to 200 ships (Figure 3.3), mapping, trading, and settling Chinese along the way. His last two voyages were the most extensive, and speculation persists about the areas that he might have discovered (see the book *1421: The Year the Chinese Discovered America* by Gavin Menzies). Unfortunately, the Ming Emperor destroyed the records of these expeditions.

(A) (B)

Figure 3.3 (A) Admiral Zheng and (B) two of his ships.

any representation—what to include and what to leave out. When U.S. President Thomas Jefferson dispatched Meriwether Lewis to explore and report on the nature of the lands from the upper Missouri to the Pacific, he said Lewis possessed "a fidelity to the truth so scrupulous that whatever he should report would be as certain as if seen by ourselves." But Jefferson clearly did not expect Lewis to report everything he saw in complete detail: Lewis exercised a large amount of judgment about what to report and what to omit. (The related question of the accuracy of what is reported is taken up at length in Chapter 5.)

One more vital interest drives our need for representations of the geographic world, as well as the need for representations in many other human activities. When a pilot must train to fly a new type of aircraft, it is much cheaper and less risky for him or her to work with a flight simulator than with the real aircraft. Flight simulators can represent a much wider range of conditions than a pilot will normally experience in flying. Similarly, when decisions have to be made about the geographic world, it is effective to experiment first on models or representations, exploring different scenarios. Of course, this works only if the representation behaves as the real aircraft or world does, and a great deal of knowledge must be acquired about the world before an accurate representation can be built that permits such simulations. But the use of representations for training, exploring future scenarios, and for re-creating the past is now common in many fields, including surgery, chemistry, and engineering, and with GI technologies it is becoming increasingly common in dealing with the geographic world.

Many plans for the real world can be tried out first on models or representations.

3.4 The Fundamental Problem

Geographic data are built up from atomic elements or from facts about the geographic world. At its most primitive, an atom of geographic data (strictly, a datum) links a place, often a time, and some descriptive property. The first of these, place, is discussed at greater length in Chapter 4, and there are also many ways of specifying the second, time. We often use the term *attribute* to refer to the last of these three descriptive properties. For example, consider the statement "The temperature at local noon on December 2, 2014, at latitude 34 degrees 45 minutes north, longitude 120 degrees 0 minutes west, was 19 degrees Celsius." It ties location and time to the property or attribute of atmospheric temperature.

Geographic data link place, time, and attributes.

Other facts can be broken down into their primitive atoms. For example, the statement "Mount Everest is 8848 m high" can be derived from two atomic geographic facts, one giving the location of Mount Everest in latitude and longitude, and the other giving the elevation at that latitude and longitude. Note, however, that the statement would not be a geographic fact to a community that had no way of knowing where Mount Everest is located.

Many aspects of the Earth's surface are comparatively static and slow to change. Height above sea level changes slowly because of erosion and the movements of the Earth's crust, but these processes operate on scales of hundreds or thousands of years, and for most applications except geophysics we can safely omit time from the representation of elevation. In contrast, atmospheric temperature changes daily, and dramatic changes sometimes occur in minutes with the passage of a cold front or thunderstorm. Thus time is distinctly important, though such climatic variables as mean annual temperature can be represented as static over periods of a decade or two.

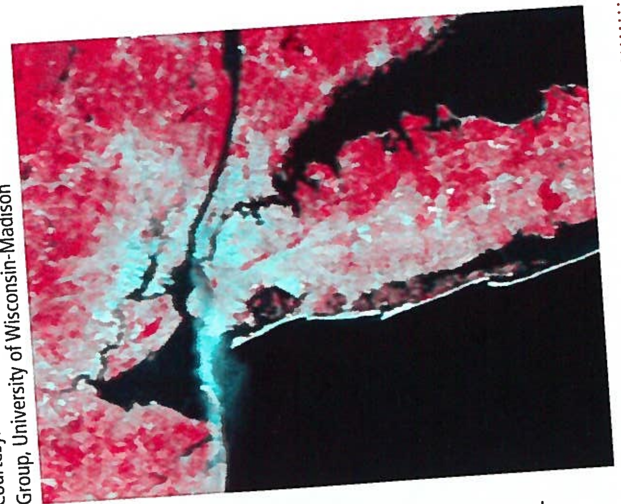
There is a vast range of attributes in geographic information. We have already seen that some attributes vary slowly and some rapidly. Some attributes are physical or environmental in nature, whereas others are social or economic. Some simply *identify* a place or an entity, distinguishing it from all other places or entities; examples include street addresses, social security numbers, or the parcel numbers used for recording land ownership. Other attributes measure something at a location and perhaps at a time (e.g., atmospheric temperature or elevation), whereas others classify into categories (e.g., the class of land use, differentiating between agriculture, industry, or residential land). The standard terms for the different types of attributes were discussed in Box 2.1.

But this idea of recording atoms of geographic information, combining location, time, and attribute, misses a fundamental problem, which is that the world is in effect infinitely complex, and the number of atoms required for a complete representation is similarly infinite. The closer we look at the world, the more detail it reveals—and it seems that this process extends ad infinitum. The shoreline of Maine appears complex on a map, but even more complex when examined in greater detail, and as more detail is revealed the shoreline appears to get longer and longer, and more and more convoluted (see Figure 2.14).

To characterize the world completely, we would have to specify the location of every person, every blade of grass, and every grain of sand—in fact, every subatomic particle, which is clearly an impossible task, because the Heisenberg Uncertainty Principle places limits on the ability to measure precise positions of subatomic particles. Thus in practice any representation must be partial—it must limit the level of detail provided, or ignore change through time, or ignore certain attributes, or simplify in some other way.

The world is infinitely complex, but computer systems are finite. Representations must somehow limit the amount of detail captured.

One very common way of limiting detail is to throw away or ignore information that applies only to small areas—in other words not look too closely. The image you see on a computer screen is composed of a million or so picture elements or *pixels*, and if the whole Earth were displayed at once, each pixel would cover an area roughly 10 km on a side, or about 100 sq km.



Courtesy: NOAA: Liam Gumley, MODIS Atmosphere Group, University of Wisconsin-Madison

Figure 3.4 An image from NASA's Terra satellite showing a large plume of smoke streaming southward from the remnants of the burning World Trade Towers in downtown Manhattan on September 11, 2001. The image was acquired using the Moderate-Resolution Imaging Spectroradiometer (MODIS), which has a spatial resolution of about 250 m. The red pixels in this scene show the location of vegetation. Light blue-white pixels show where there are concrete surfaces. The slightly darker blue pixels streaming southward toward the New Jersey coast are the smoke from the destroyed World Trade Center towers. The large black areas to the east are the waters of the Atlantic. The Hudson River runs northward just west of Manhattan Island.

At this level of detail the island of Manhattan occupies roughly 10 pixels, and virtually everything on it is a blur. We would say that such an image has a *spatial resolution* of about 10 km and know that anything much less than 10 km across is virtually invisible. Figure 3.4 shows Manhattan at a spatial resolution of 250 m, detailed enough to pick out the shape of the island and Central Park.

It is easy to see how this helps with the problem of too much information. The Earth's surface covers about 500 million sq km, so if this level of detail (a spatial resolution of 10 km) is sufficient for an application, a property of the surface such as elevation can be described with only 5 million pieces of information, instead of the 500 million it would take to describe elevation with a resolution of 1 km, and the 500 trillion (500,000,000,000,000) it would take to describe elevation with 1-m resolution.

Another strategy for limiting detail is to observe that many properties remain constant over large areas. For example, in describing the elevation of the Earth's surface we could take advantage of the fact that roughly two-thirds of the surface is covered by water, with its surface at sea level. Of the 5 million pieces of information needed to describe elevation at 10-km resolution, approximately 3.4 million will be recorded as zero, a colossal waste. If we could find an efficient way of identifying the area covered by water, then we would need only 1.6 million real pieces of information.

Humans have found many ingenious ways of describing the Earth's surface efficiently because the problem we are addressing is as old as representation itself and as important for paper-based representations as it is for binary representations in computers. But this ingenuity is itself the source of a substantial problem for GI: there are many ways of representing the Earth's surface, and users of GI thus face difficult and at times confusing choices. This chapter discusses some of those choices, and the issues are pursued further in subsequent chapters on uncertainty (Chapter 5) and data modeling (Chapter 7). Representation remains a major concern of GI science, and researchers are constantly looking for ways to extend GI representations to accommodate new types of information.

3.5 Discrete Objects and Continuous Fields

3.5.1 Discrete Objects

The level of detail as a fundamental choice in representation has already been mentioned. Another, perhaps even more fundamental, choice is between two conceptual schemes. There is good evidence that we as humans like to simplify the world around us by naming things and by seeing individual things as instances of broader categories. We prefer a world of black and white, of good guys and bad guys, to the real world of shades of gray.

The two fundamental ways of representing geography are discrete objects and continuous fields.

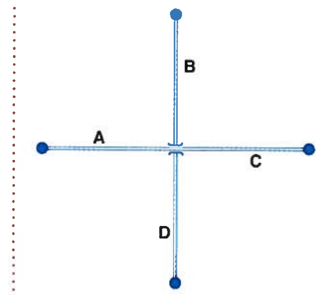
This preference is reflected in one way of viewing the geographic world, known as the *discrete object* view. In this view, the world is empty, except where it is occupied by objects with well-defined boundaries that are instances of generally recognized categories. Just as the desktop is littered with books, pencils, or computers, the geographic world is littered with cars, houses, lampposts, and other discrete objects. In a similar vein the landscape of Minnesota is littered with lakes and that of Scotland with mountains. One characteristic of the discrete object view is that objects can be counted, so license plates issued by the State of Minnesota carry the legend "10,000 lakes," and climbers know that there are exactly 284 mountains in Scotland over 3,000 ft (the so-called Munros, from Sir Hugh Munro, who originally listed 277 of them in 1891—the count was expanded to 284 in 1997).

The discrete object view represents the geographic world as objects with well-defined boundaries in otherwise empty space.

Biological organisms fit this model well, allowing us to count the number of residents in an area of a city or to describe the behavior of individual bears. Manufactured objects also fit the model, and we have little difficulty counting the number of cars produced in a year or the number of airplanes owned by an airline. But other phenomena are messier. It is not at all clear what constitutes a mountain, for example, or exactly how a mountain differs from a hill, or when a mountain with two peaks should be counted as two mountains.

Geographic objects are identified by their dimensionality. We saw in Box 2.2 that GI objects are conceived as filling zero (points), one (lines), two (areas), and three (volumes) dimensions, and that fractal geometry allowed us to think of objects as filling intermediate amounts of space between these integer dimensions (Section 2.8). Of course, in reality, all objects that are perceptible to humans are three dimensional, and their representation in fewer dimensions can be at best an approximation. But the ability of GI systems to handle truly three-dimensional objects as volumes with associated surfaces remains limited. GI databases increasingly allow for a third (vertical) coordinate to be specified for all point locations. Buildings are sometimes represented by assigning height as an attribute, though if this option is used it is impossible to distinguish flat roofs from any other kind. Various strategies have been used for representing overpasses and underpasses in transportation networks because this information is vital for navigation but not normally present in strictly two-dimensional network representations. One common strategy is to represent turning options at every intersection, so that an overpass appears in the database as an intersection with no turns (Figure 3.5).

The discrete object view leads to a powerful way of representing geographic information about objects. Think of a class of objects of the same dimensionality—for example, all the brown bears (Figure 3.6) in the Kenai Peninsula of Alaska. We would naturally think of these objects as points. We might want to know the sex of each bear and its date of birth, if our interests were in monitoring the bear population. We might also have a collar on each bear that transmitted the bear's location at regular intervals. All this information could be expressed in a table, such as the one shown in Table 3.1, with each



From link	To link	Turn?
A	B	No
A	C	Yes
A	D	No
B	C	No
B	D	Yes
B	A	No
C	D	No
C	A	Yes
C	B	No
D	A	No
D	B	Yes
D	C	No

Figure 3.5 The problems of representing a three-dimensional world using a two-dimensional technology. The intersection of links A, B, C, and D is an overpass, so no turns are possible between such pairs as A and B.

Figure 3.6 Bears are easily conceived as discrete objects, maintaining their identities as objects through time and surrounded by empty space.

Table 3.1 Example of representation of geographic information as a table: the locations and attributes of each of four brown bears in the Kenai Peninsula of Alaska. Locations have been obtained from radio collars. Only one location is shown for each bear, at noon on July 31, 2014 (imaginary data).

Bear ID	Sex	Estimated year of birth	Date of collar installation	Location, noon on 31 July 2014
001	M	2008	02242009	−150.6432, 60.0567
002	F	2006	03312009	−149.9979, 59.9665
003	F	2013	04212009	−150.4639, 60.1245
004	F	2010	04212009	−150.4692, 60.1152

row corresponding to a different discrete object and each column to an attribute of the object. To reinforce a point made earlier, this is a very efficient way of capturing raw geographic information on brown bears.

But it is not perfect as a representation for all geographic phenomena. Imagine visiting the Earth from another planet and asking the humans what they chose as a representation for the infinitely complex and beautiful environment around them. The visitor would hardly be impressed to learn that they chose tables, especially when the phenomena represented were natural phenomena such as rivers, landscapes, or oceans. Nothing on the natural Earth looks remotely like a table. It is not at all clear how the properties of a river or the properties of an ocean should be represented as a table. So although the discrete object view works well for some kinds of phenomena, it misses the mark badly for others.

3.5.2 Continuous Fields

Although we might think of terrain as composed of discrete mountain peaks, valleys, ridges, slopes, and the like and might list them in tables and count them, there are unresolvable problems of definition for all these objects. Instead, it is much more useful to think of terrain as a continuous surface in which elevation can be defined rigorously at every point (see Box 3.3). Such continuous surfaces form the basis of the other common view of geographic phenomena, known as the *continuous field* view (not to be confused with other meanings of the word field). In this view the geographic world can be described by a number of *variables*, each measurable at any point on the Earth's surface and changing in value across the surface.

The continuous field view represents the real world as a finite number of variables, each one defined at every possible position.

Objects are distinguished by their dimensions and naturally fall into categories of points, lines, or areas. Continuous fields, on the other hand, can be distinguished by what varies and how smoothly. A continuous field of elevation, for example, varies much more smoothly in a landscape that has been worn down by glaciation or flattened by blowing sand than one recently created by cooling lava. Cliffs are places in continuous fields where elevation changes suddenly rather than smoothly. Population density is a kind of continuous field, defined everywhere as the number of people per unit area, though the definition breaks down if the field is examined so closely that the individual people become visible. Continuous fields can also be created from classifications of land into categories of land use or soil type. Such fields change suddenly at the boundaries between different classes. Other types of fields can be defined by continuous variation along lines rather than across space. Traffic density, for example, can be defined everywhere on a road network, and flow volume can be defined everywhere on a river. Figure 3.7 shows some examples of field-like phenomena.

Continuous fields can be distinguished by *what* is being measured at each point. As described in Box 2.1, the variable may be nominal, ordinal, interval, ratio, or cyclic. A *vector* field assigns two variables, magnitude and direction, at every point in space and is used to represent flow phenomena such as winds or currents; fields of only one variable are termed *scalar* fields.

Here is a simple example illustrating the difference between the discrete object and field conceptualizations. Suppose you were hired for the summer to count the number of lakes in Minnesota and were promised that your answer would appear on every license plate issued by the state. The task sounds simple, and you were happy to get the job. But on the first day you started to run into difficulty (Figure 3.8). What about small ponds—do they count as lakes? What about wide stretches of rivers? What about swamps that dry up in the summer? And is a lake with a narrow section connecting two wider parts one lake or two? Your biggest dilemma concerns the scale of mapping because the number of lakes shown on a map clearly depends on the map's level of detail; a more detailed map almost certainly will show more lakes.

Your task clearly reflects a discrete object view of the phenomenon. The action of counting implies that lakes are discrete, two-dimensional objects littering an otherwise empty geographic landscape. In a continuous field view, however, all points are either lake or nonlake. Moreover, we could refine the scale a little to take account of marginal cases; for example, we might define the scale shown in Table 3.2, which has five degrees of lakeness. The complexity of the view would depend on how closely we looked, of course, and so the scale of mapping would still be important. But all the problems of defining a lake as a discrete object would disappear (though there would still be problems in defining the levels of the scale). Instead of counting, our strategy would be to lay a grid over the map and assign each grid cell a score on the lakeness scale. The size of the grid cell would determine how accurately the result approximated the value we could theoretically obtain by visiting every one of the infinite number of points in the state. At the end, we would tabulate the resulting scores, counting the number of cells having each value of lakeness or averaging the lakeness score. We could even design a new and scientifically more reasonable license plate—"Minnesota, 12% lake" or "Minnesota, average lakeness 2.02." The vagaries of defining such objects in Finland (Figure 3.8), another lake-strewn territory, form

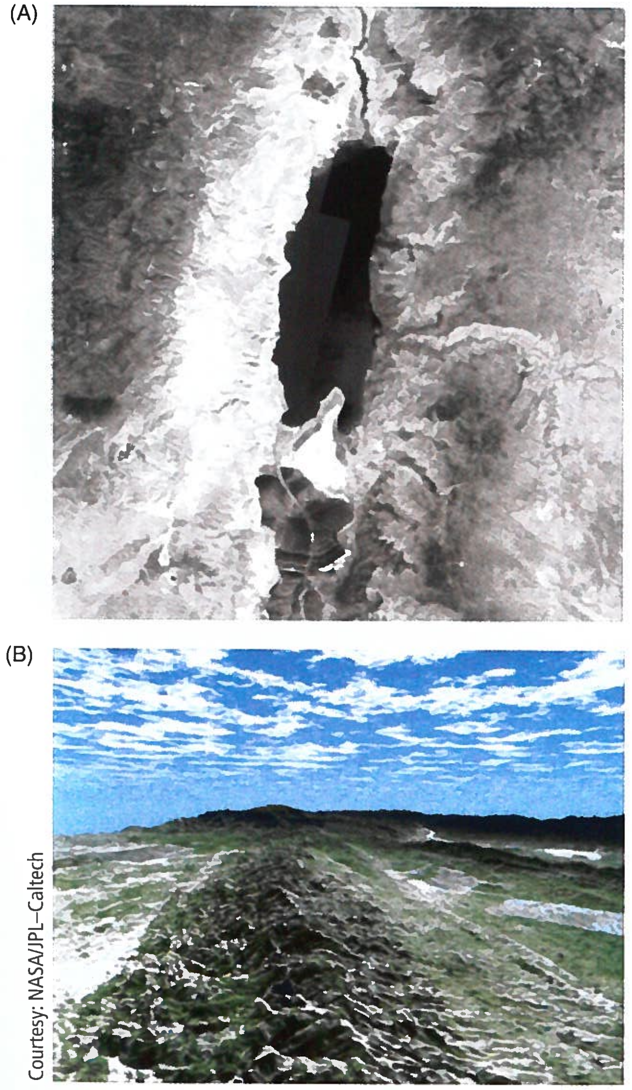
Technical Box 3.3

Dimensions

Areas are two-dimensional objects, and volumes are three dimensional, but GI users sometimes talk about 2.5-D. Almost without exception the elevation of the Earth's surface has a single value at any location (exceptions include overhanging cliffs or caves). So elevation is conveniently thought of as a continuous field, a variable with a value everywhere in two dimensions, and a full 3-D representation is only necessary in areas with an abundance of overhanging cliffs or caves, if these are important features. The idea of dealing with a three-dimensional phenomenon by treating it as a single-valued function of two horizontal variables gives rise to the term 2.5-D. Figure 3.7B shows an example, in this case an elevation surface.

(A)

(B)



Courtesy: NASA/JPL–Caltech

Figure 3.7 Examples of field-like phenomena. (A) Image of part of the Dead Sea in the Middle East. The lightness of the image at any point measures the amount of radiation captured by the satellite's imaging system. (B) A simulated image derived from the Shuttle Radar Topography Mission. The image shows the Carrizo Plain area of Southern California, with a simulated sky and with land cover obtained from other satellite sources.

Figure 3.8 Lakes (this one is in Helsinki, Finland) are difficult to conceptualize as discrete objects because it is often difficult to tell where a lake begins and ends, or to distinguish a wide river from a lake.

Table 3.2 A scale of lakeness suitable for defining lakes as a continuous field.

Lakeness	Definition
1	Location is always dry under all circumstances.
2	Location is sometimes flooded in spring.
3	Location supports marshy vegetation.
4	Water is always present to a depth of less than 1 m.
5	Water is always present to a depth of more than 1 m.

Biographical Box 3.4

Kirsi Virrantaus, Architect, Cartographer, and GI Scientist

Kirsi Virrantaus (Figure 3.9) is an architect by training, although she also studied land use planning, applied mathematics, and computer science in the course of her undergraduate education. This combination of practical skills in planning, design studies, and quantitative methods equipped her to undertake doctoral studies in cartography in the Department of Surveying of Helsinki University of Technology (HUT: now part of Aalto University), where she graduated in 1984. After spending some years working in the private and public sectors, she returned to academia as the first professor of cartography at HUT in 1988. There, she has developed a curriculum for cartography and GI science, initially for use in Helsinki but subsequently for wider deployment under the auspices of the International Cartographic Association (see icaci.org/files/documents/reference_docs/2009_ICA_ResearchAgenda.pdf).

Kirsi's early work focused upon the use of then-emerging technologies for the management of geographic data and the attendant challenges of map representation using new visual interfaces. Over the years, her research has frequently entailed reuse of secondary data sources to develop innovative GI applications in ways that are efficient, effective, and safe to use. These applications have also entailed the deployment of advanced analysis methods, spatial statistics, and innovative geocomputational tools. Throughout, Kirsi's work has required a thorough understanding of the quality of geographic representations, and this pan-European experience has led to her involvement in the development of International Organization for Standardization (ISO) measures of geographic data quality. Such quality assurance issues have been particularly pertinent in her development of systems for crisis and emergency management, where issues of user interaction (Chapter 12) come to the fore. Kirsi's work has also identified many of the pitfalls of using incomplete and imprecise data (Chapter 5) and she and her research group have found fuzzy approaches (Section 5.2.3) to be particularly relevant in devising workable solutions.


Courtesy: Kirsi Virrantaus

Figure 3.9 Kirsi Virrantaus: Architect, Cartographer, and GI scientist.

the backdrop to some of the work of cartographer and GI scientist Kirsi Virrantaus (see Box 3.4).

The difference between objects and fields is also well illustrated by photographs. Paper images produced from old-fashioned photographic film are created by variation in the chemical state of the material in the photographic film. In early photography, minute particles of silver were released from molecules of silver nitrate when the unstable molecules were exposed to light, thus darkening the image in proportion to the amount of incident light. We think of the image as a field of continuous variation in color or darkness. But when we look at the image, the eye and brain begin to infer the presence of discrete objects, such as people, rivers, fields, cars, or houses, as they interpret the content of the image.

3.6 Rasters and Vectors

Continuous fields and discrete objects define two conceptual views of geographic phenomena, but they do not solve the problem of digital representation. A continuous field view still potentially contains an infinite amount of information if it defines the value of the variable at every point because there are an infinite number of points in any defined geographic area. Discrete objects can also require an infinite amount of information for full description. As we saw in Section 2.8, a coastline contains an infinite amount of information if it is mapped in infinite detail. Thus continuous fields and discrete objects are no more than conceptualizations, or ways in which we think about geographic phenomena; they are not designed to deal with the limitations of computers.

Two methods are used to reduce geographic phenomena to forms that can be coded in computer databases, and we call these methods raster and vector. In principle, both can be used to code both fields and discrete objects, but in practice a strong association exists between raster and fields, and between vector and discrete objects.

Raster and vector are two methods of representing geographic data in digital computers.

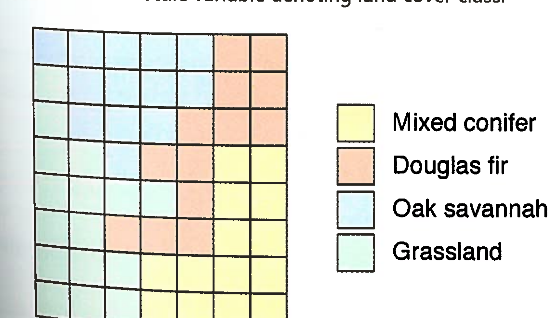
3.6.1 Raster Data

In a raster representation space is divided into an array of rectangular (usually square) cells (Figure 3.10). All geographic variation is then expressed by assigning properties or attributes to these cells. The cells are sometimes called pixels (short for *picture elements*).

Raster representations divide the world into arrays of cells and assign attributes to the cells.

One of the most common forms of raster data comes from remote-sensing satellites, which capture information in this form and send it to the ground to be distributed and analyzed. Data from the Landsat Thematic Mapper, for example, which are commonly used in GI applications, come in cells that are 30 m a side on the ground, or approximately 0.1 ha (hectare) in area. Similar data can be obtained from sensors mounted on aircraft. Imagery varies according to the spatial resolution (expressed as the length of a cell side as measured on the ground) and also according to the timetable of image capture by the sensor. Some satellites are in *geostationary* orbit over a fixed point on the Earth and capture images constantly. Others pass over a fixed point at regular intervals (e.g., every 12 days). Finally, sensors vary according to the part or parts of the spectrum that they sense. The visible parts of the spectrum are most important for remote sensing, but some invisible parts of the spectrum are particularly useful in detecting heat and the phenomena that produce heat, such as volcanic activities. Many sensors capture images in several areas of the spectrum, or *bands*, simultaneously because the relative amounts of radiation in different parts of the spectrum are often useful indicators of certain phenomena, such as green leaves, or water, on the Earth's surface. The AVIRIS (Airborne Visible InfraRed Imaging Spectrometer) captures no fewer than 224 different parts of the spectrum and is being used to detect particular minerals in the soil, among other applications. Remote sensing is a complex topic, and further details are available in Chapter 8.

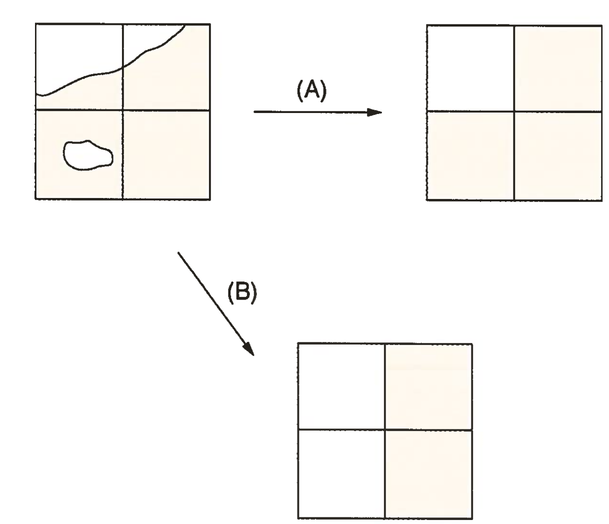
Figure 3.10 Raster representation. Each color represents a different value of a nominal-scale variable denoting land cover class.



Square cells fit together nicely on a flat table or a sheet of paper, but they will not fit together neatly on the curved surface of the Earth. So just as representations on paper require that the Earth be flattened, or projected, so too do rasters. (Because of the distortions associated with flattening, the cells in a raster can never be perfectly equal in shape or area on the Earth's surface.) Projections, or ways of flattening the Earth, are described in Section 4.8. Many of the terms that describe rasters suggest the laying of a tile floor on a flat surface—we talk of raster cells *tiling* an area, and a raster is said to be an instance of a *tessellation*, derived from the word for a mosaic. The mirrored ball hanging above a dance floor recalls the impossibility of covering a spherical object like the Earth perfectly with flat, square pieces.

When information is represented in raster form, all detail about variation within cells is lost, and instead the cell is given a single value. Suppose we wanted to represent the map of the counties of Texas as a raster. Each cell would be given a single value to identify a county, and we would have to decide on the rule to apply when a cell falls in more than one county. Often the rule is that the county with the *largest share* of the cell's area gets the cell. Sometimes the rule is based on the *central point* of the cell, and the county at that point is assigned to the whole cell. Figure 3.11 shows these two rules in operation. The largest-share rule is almost always preferred, but the central-point rule is sometimes used in the interests of faster computing and is often used in creating raster datasets of elevation.

Figure 3.11 Effect of a raster representation using (A) the largest-share rule and (B) the central-point rule.



3.6.2 Vector Data

In a vector representation, all lines are captured as points connected by precisely straight lines. (Some GI systems allow points to be connected by curves rather than straight lines, but in most cases curves have to be approximated by increasing the density of points.) An area is captured as a series of points or *vertices* connected by straight lines as shown in Figure 3.12. The straight edges between vertices explain why areas in vector representation are often called *polygons*, and in GI-speak the terms *polygon* and *area* are often used interchangeably. Lines are captured in the same way, and the term *polyline* has been coined to describe a curved line represented by a series of straight segments connecting vertices.

To capture an area object in vector form, we need only specify the locations of the points that form the vertices of a polygon. This seems simple and is also much more efficient than a raster representation, which would require us to list all the cells that form the area. These ideas are captured succinctly in the comment "Raster is vaster, but vector is correcter." To create a precise approximation to an area in raster, it would be necessary to resort to using very small cells, and the number of cells would rise proportionately (with every halving of the width and height of each cell resulting in a quadrupling of the number of cells). But things are not quite as simple as they seem. The apparent precision of vector is often not real because many geographic phenomena simply cannot be located with high accuracy (see Section 5.3). So although raster data may look less attractive, they may be more honest to the inherent quality of the data. Also, various methods exist for compressing raster data that can greatly reduce the capacity needed to store a given dataset (see Chapter 7). So the choice between raster and vector is often complex, as summarized in Table 3.3.

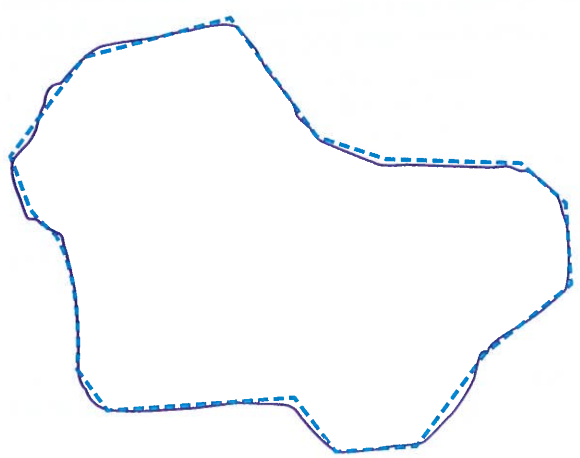


Figure 3.12 An area (purple line) and its approximation by a polygon (dashed blue line).

3.6.3 Representing Continuous Fields

Although discrete objects lend themselves naturally to representation as points, lines, or areas using vector methods, it is less obvious how the continuous variation of a field can be expressed in a digital representation. GI systems commonly implement six alternatives (Figure 3.13):

A. Capturing the value of the variable at each of a grid of regularly spaced sample points (for example, elevations at 30-m spacing in a DEM).

B. Capturing the value of the field variable at each of a set of irregularly spaced sample points (for example, variation in surface temperature captured at weather stations).

C. Capturing a single value of the variable for a regularly shaped cell (for example, values of reflected radiation in a remotely sensed scene).

D. Capturing a single value of the variable over an irregularly shaped area (for example, vegetation cover class or the name of a parcel's owner).

E. Capturing the linear variation of the field variable over an irregularly shaped triangle (for example, elevation captured in a triangulated irregular network or TIN, see Section 7.2.3.4).

F. Capturing the isolines of a surface as digitized lines (for example, digitized contour lines representing surface elevation).

Table 3.3 Relative advantages of raster and vector representation.

Issue	Raster	Vector
Volume of data	Depends on cell size	Depends on density of vertices
Sources of data	Remote sensing, imagery	Social and environmental data
Applications	Resources, environmental	Social, economic, administrative
Software	Raster GI systems, image processing	Vector GI systems, automated cartography
Resolution	Fixed	Variable

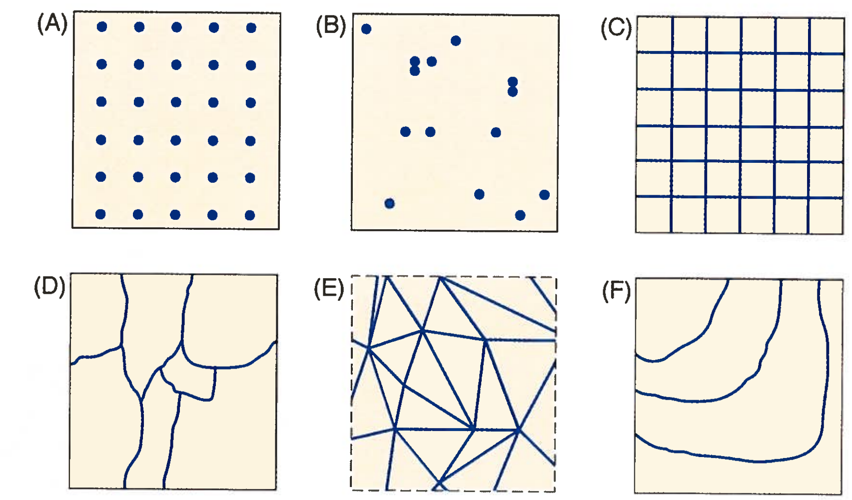


Figure 3.13 The six approximate representations of a field used in GI systems. (A) Regularly spaced sample points. (B) Irregularly spaced sample points. (C) Rectangular cells. (D) Irregularly shaped polygons. (E) Irregular network of triangles, with linear variation over each triangle (the Triangulated Irregular Network or TIN model; the bounding box is shown dashed in this case because the unshown portions of complete triangles extend outside it). (F) Polylines representing contours (see the discussion of isopleth maps in Box 2.3).

Each of these methods succeeds in compressing the potentially infinite amount of data in a continuous field to a finite amount, using one of the six options, two of which (A and C) are raster and four (B, D, E, and F) are vector. Of the vector methods one (B) uses points, two (D and E) use polygons, and one (F) uses lines to express the continuous spatial variation of the field in terms of a finite set of vector objects. But unlike the discrete object conceptualization, the objects used to represent a field are not real, but simply artifacts of the representation of something that is actually conceived as spatially continuous. The triangles of a TIN representation (E), for example, exist only in the digital representation and cannot be found on the ground, and neither can the lines of a contour representation (F).

Representation also affects the conception of spatial autocorrelation (Section 2.3). If a phenomenon of interest is conceived as a field, then spatial autocorrelation measures the smoothness of the field using data from the sample points, lines, or areas that represent the field. If the phenomenon is conceived as a set of discrete objects, then spatial autocorrelation measures how the attribute values are distributed among the objects, distinguishing between arrangements that are clustered, random, and locally contrasting.

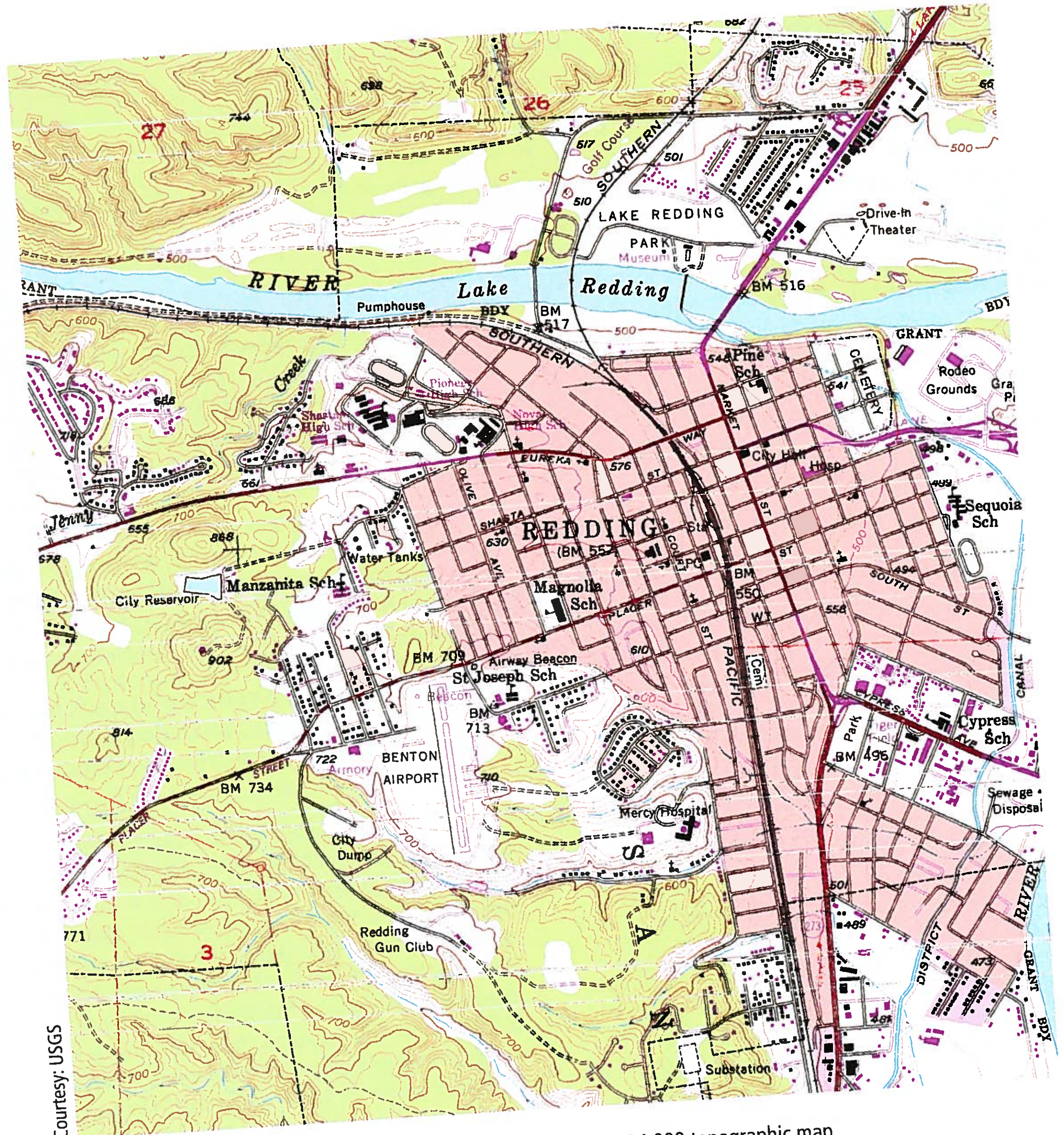
3.7 The Paper Map

The paper map has long been a powerful and an effective means of communicating geographic information. In contrast to digital data, which use coding schemes such as ASCII, it is an instance of an *analog* representation, or a physical model in which the real world is scaled; in the case of the paper map, part of the world is scaled to fit the size of the paper. A key property of a paper map is its *scale* or *representative fraction*, defined as the ratio of distance on the map to distance on the Earth's surface. For example, a map with a scale of 1:24,000 reduces everything on the Earth to one-24,000th of its real size. This is a bit misleading because the Earth's surface is curved and a paper map is flat, so scale cannot be exactly constant.

A paper map is a source of data for geographic databases, an analog product from a GI system, and an effective communication tool.

Maps have been so important, particularly prior to the development of digital technology, that many of the ideas associated with GI are actually inherited directly from paper maps. For example, scale is often cited as a property of a digital database, even though the definition of scale makes no sense for digital data—ratio of distance *in the computer* to distance on the ground; how can there be distances in a computer? What is meant is a little more complicated: when a scale is quoted for a digital database, it is usually the scale of the map that formed the source of the data. So if a database is said to be at a scale of 1:24,000, one can safely assume that it was created from a paper map at that scale and includes representations of the features that are found on maps at that scale. The many meanings of scale were discussed in Box 2.2, and we discuss the importance of scale to the concept of uncertainty in Chapter 5.

There is a close relationship between the contents of a map and the raster and vector representations discussed in the previous section. The U.S. Geological Survey, for example, distributes two digital versions of its topographic maps, one in raster form and one in vector form, and both attempt to capture the contents of the map as closely as possible. In the raster form, or *digital raster graphic* (DRG), the map is scanned at a very high density, using very small pixels, so that the raster looks very much like the original (Figure 3.14). The coding of each pixel simply records



Courtesy: USGS

Figure 3.14 Part of a digital raster graphic, a scan of a U.S. Geological Survey 1:24 000 topographic map.

the color of the map picked up by the scanner, and the dataset includes all the textual information surrounding the actual map.

In the vector form, or *digital line graph* (DLG), every geographic feature shown on the map is represented as a point, polyline, or polygon. The symbols used to represent point features on the map, such as the symbol for a windmill, are replaced in the digital data by points with associated attributes and must be regenerated when the data are displayed. Contours, which are shown on the map as lines of definite width, are replaced by polylines of no width and given attributes that record their elevations. In both cases, and especially in the vector case, there is a significant difference between the analog representation of the map and its digital equivalent.

So it is quite misleading to think of the contents of a digital representation as a map and to think of a GI database as a container of digital maps. Digital representations can include information that would be very difficult to show on maps. For example, they can represent the curved surface of the Earth, without the need for the distortions associated with flattening. They can represent changes, whereas maps must be static because it is very difficult to change their contents once they have been printed or drawn. Digital databases can represent all three spatial dimensions, including the vertical, whereas maps must always show two-dimensional views. So although the paper map is a useful metaphor for the contents of a geographic database, we must be careful not to let it limit our thinking about what is possible in the way of representation. This issue is pursued at greater length in Chapter 7, and map production is discussed in detail in Chapter 11.

3.8 Generalization

In Section 3.4 we saw how thinking about geographic information as a collection of atomic links—between a place, a time (not always, because many geographic facts are stated as if they were permanently true), and a property—led to an immediate problem because the potential number of such atomic facts is infinite. If seen in enough detail, the Earth's surface is unimaginably complex, and its effective description impossible. So instead, humans have devised numerous ways of simplifying their view of the world. Rather than making statements about each and every point, we describe entire areas, attributing uniform characteristics to them, even when areas are not strictly uniform; we identify features on the ground and describe their characteristics, again assuming them to be uniform; or we limit our descriptions to what exists at a finite number of sample points, hoping that these samples will be adequately representative of the whole (Section 2.4).

A geographic database cannot contain a perfect description; instead, its contents must be carefully selected to fit within the limited capacity of computer storage devices.

3.8.1 Generalization about Places

From this perspective some degree of generalization is almost inevitable in all geographic data. But cartographers often take a somewhat different approach, for which this observation is not necessarily true. Suppose we are tasked to prepare a map at a specific scale, say 1:25,000, using the standards laid down by a national mapping agency, such as the Institut Géographique National (IGN) of France. Every scale used by IGN has its associated rules of representation. For example, at a scale of 1:25,000 the rules specify that individual buildings will be shown only in specific circumstances, and similar rules apply to the 1:24,000 series of the U.S. Geological Survey (see Figure 3.14). These rules are known by various names, including *terrain nominal* in the case of IGN, which translates roughly but not very helpfully to "nominal ground," and is perhaps better translated as "specification." From this perspective a map that represents the world by following the rules of a specification precisely can be perfectly accurate *with respect to the specification*, even though it is not a perfect representation of the full detail on the ground.

A map's specification defines how real features on the ground are selected for inclusion on the map.

Consider the representation of vegetation cover using the rules of a specification. For example, the rules might state that at a scale of 1:100,000, a vegetation cover map should not show areas of vegetation that cover less than 1 ha (hectare). But small areas of vegetation almost certainly exist, so deleting them inevitably results in information loss. But under the principle discussed earlier, a map that adheres to this rule must be accurate, *even though it differs substantively from the truth as observed on the ground*.

The level of detail of a GI dataset is one of its most important properties, as it determines both the degree to which the dataset approximates the real world and the dataset's complexity. In the interests of compressing data, it is often necessary to remove detail, fitting them into a storage device of limited capacity, processing them faster, or creating less confusing visualizations that emphasize general trends. Consequently, many methods have been devised for generalization, and several of the more important are discussed in this section.

McMaster and Shea identify the following types of generalization rules:

- *Simplification*, for example, by weeding out points in the outline of a polygon to create a simpler shape
- *Smoothing*, or the replacement of sharp and complex forms by smoother ones
- *Collapse*, or the replacement of an area object by a combination of point and line objects
- *Aggregation*, or the replacement of a large number of distinct symbolized objects by a smaller number of new symbols
- *Amalgamation*, or the replacement of several area objects by a single area object
- *Merging*, or the replacement of several line objects by a smaller number of line objects
- *Refinement*, or the replacement of a complex pattern of objects by a selection that preserves the pattern's general form
- *Exaggeration*, or the relative enlargement of an object to preserve its characteristics when these would be lost if the object were shown to scale
- *Enhancement*, through the alteration of the physical sizes and shapes of symbols
- *Displacement*, or the moving of objects from their true positions to preserve their visibility and distinctiveness

The differences between these types of rules are much easier to understand visually, and Figure 3.15 is based upon McMaster's and Shea's

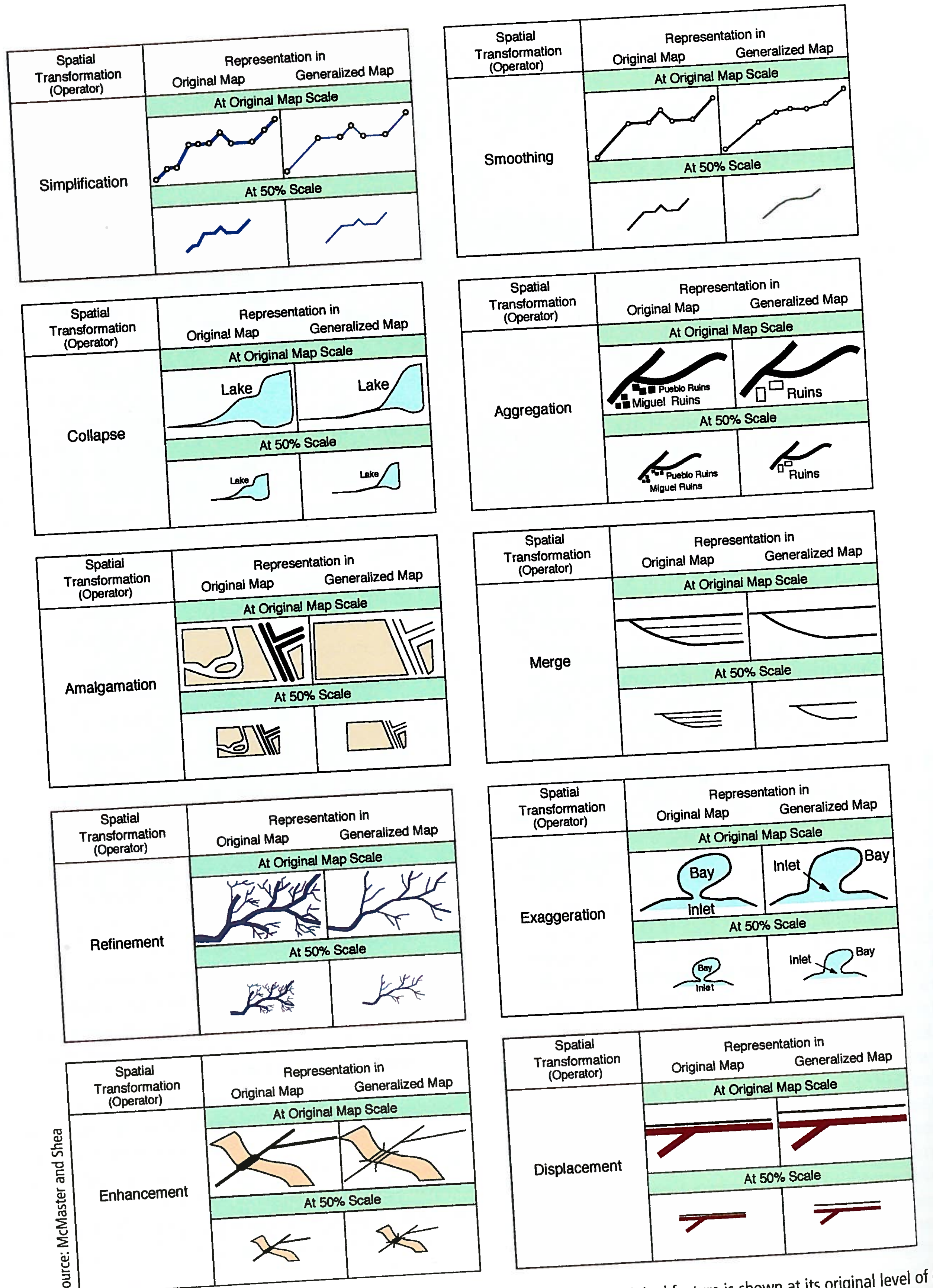


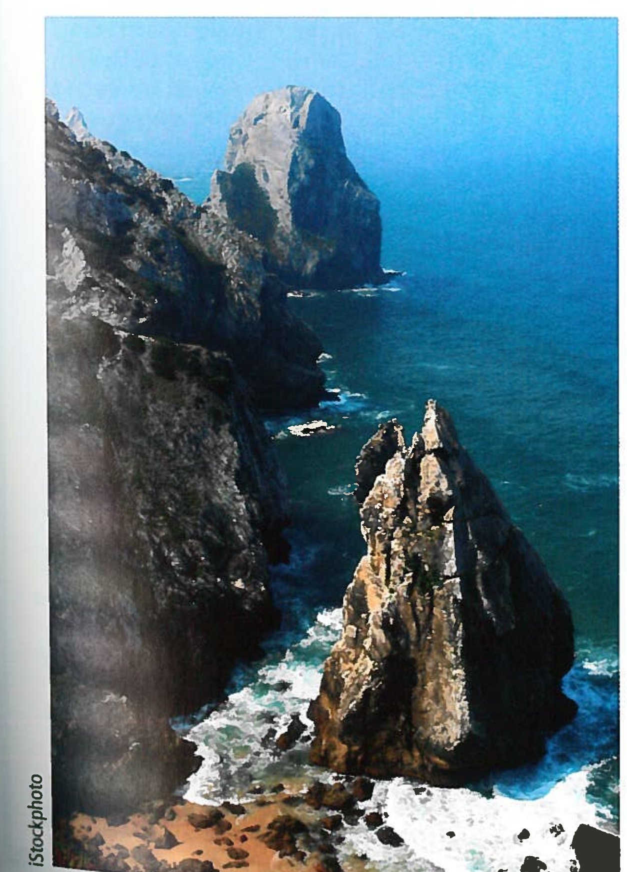
Figure 3.15 Illustrations from McMaster and Shea of their 10 forms of generalization. The original feature is shown at its original level of detail, and below it at 50% coarser scale. Each generalization technique resolves a specific problem of display at coarser scale and results in the acceptable version shown in the lower right.

illustrative drawings. In addition, the drawings describe two forms of generalization of attributes, as distinct from geometric forms of generalization. *Classification* generalization reclassifies the attributes of objects into a smaller number of classes, whereas *symbolization* generalization changes the assignment of symbols to objects. For example, it might replace an elaborate symbol including the words "Mixed Forest" with a color identifying that class.

One of the most common forms of generalization of GI is the process known as weeding, or the simplification of the representation of a line represented as a polyline. The process is an instance of McMaster and Shea's simplification. Standard methods exist in GI systems for doing this, and the most common by far is the method known as the Douglas–Poiker algorithm after its inventors, David Douglas and Tom Poiker. The operation of the Douglas–Poiker weeding algorithm upon features such as that shown in Figure 3.16 is shown in Figure 3.17.

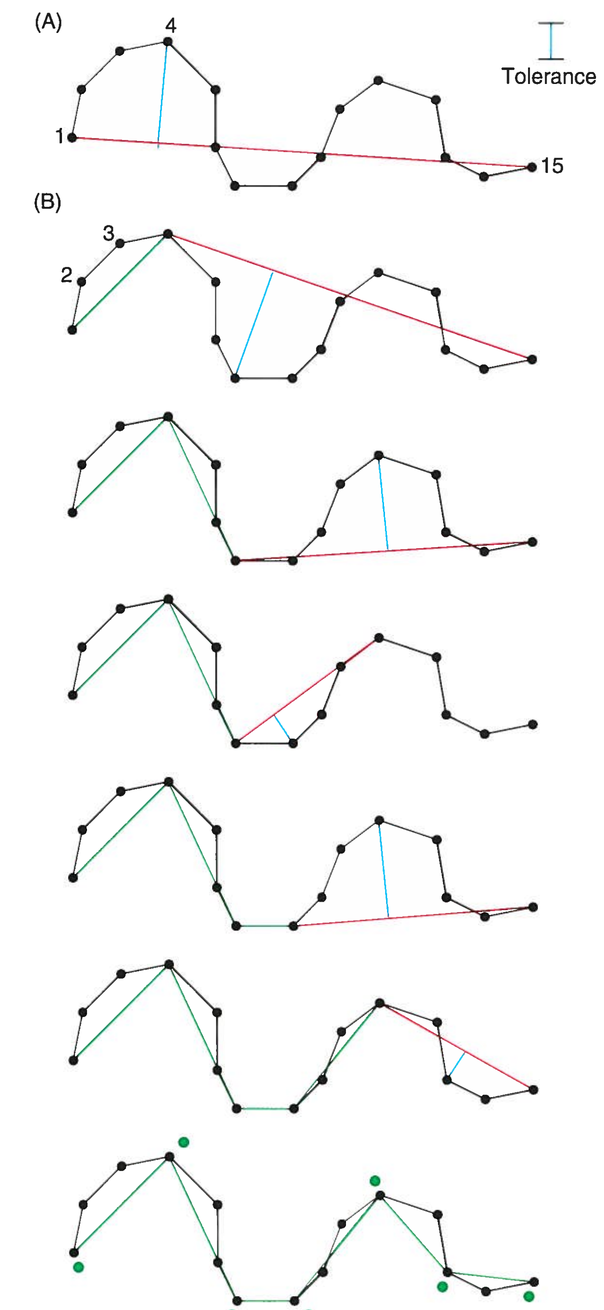
Weeding is the process of simplifying a line or an area by reducing the number of points in its representation.

Figure 3.16 The Douglas–Poiker algorithm is designed to simplify complex objects like this shoreline by reducing the number of points in its polyline representation.



Note that the algorithm relies entirely on the assumption that the line is represented as a polyline—in other words, as a series of straight-line segments. GI systems increasingly support other representations,

Figure 3.17 The Douglas–Poiker line simplification algorithm in action. The original polyline has 15 points. In (A) Points 1 and 15 are connected (red), and the furthest distance of any point from this connection is identified (blue). This distance to Point 4 exceeds the user-defined tolerance. In (B) Points 1 and 4 are connected (green). Points 2 and 3 are within the tolerance of this line. Points 4 and 15 are connected, and the process is repeated. In the final step 7 points remain (identified with green disks), including 1 and 15. No points are beyond the user-defined tolerance distance from the line.



including arcs of circles, arcs of ellipses, and Bézier curves, but there is little consensus to date on appropriate methods for weeding or generalizing them, or on methods of analysis that can be applied to them. The general procedures of generalization of map labels and features such as coastlines in Web mapping systems are explored in Box 3.5.

3.8.2 Generalization about Properties

We saw in Sections 1.3 and 1.4 how the pace of scientific research had accelerated over the past two decades, in no small part because the Internet now provides a globally linked network of data warehouses and clearinghouses. In theory, the provenance of each of these various sources is known and is documented using metadata ("data about data": see Section 10.2). Many GI representations bring together multiple properties of places as composite *indicators* of conditions at particular locations, in order to fulfill particular needs. Thus, for example, summary measures of social "deprivation" (often described as "hardship" in the United States) combine representations of conditions with respect to heath, employment, and housing, among other social phenomena. In a similar vein, multiple criteria might be combined in order to define "wilderness" in an environmental application (see Section 1.3). Problems arise if some

Applications Box 3.5

Online Map Generalization with OpenStreetMap, Mapnik, and Leaflet

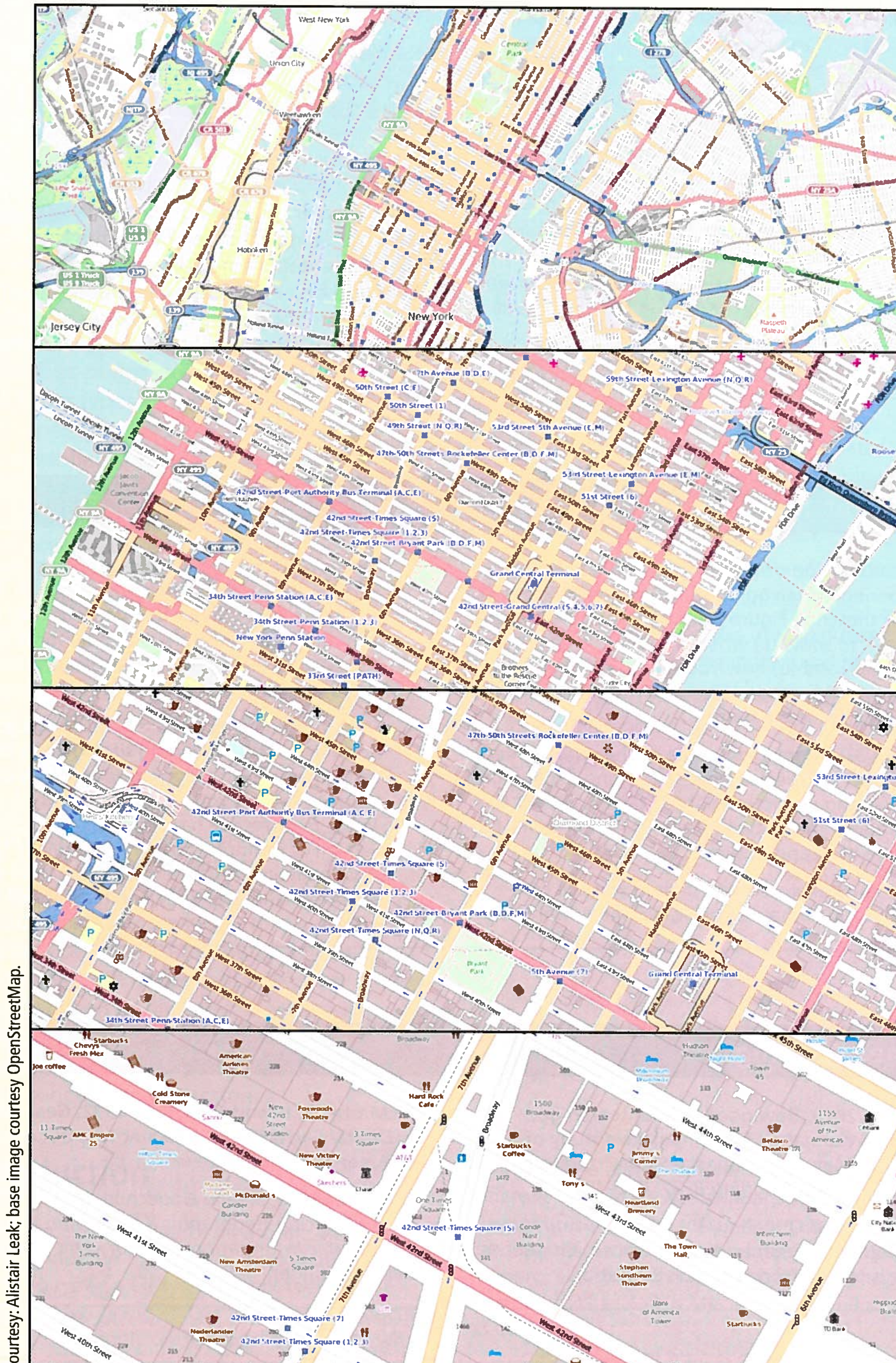
OpenStreetMap is a digital map of the world (www.openstreetmap.org), created using crowd-sourced volunteered geographic information (VGI: Section 2.5). Just as with commercial (advertising-based) products like Google (maps.google.com), Apple (www.apple.com/ios/maps/), and Bing (www.bing.com/maps/), when a user zooms through successive levels of detail, so different features are revealed.

In OpenStreetMap, this effect is achieved using Mapnik (mapnik.org), a free software tool for developing mapping applications. The user views a rasterized version of the vector data that were used to create the digital base map through a JavaScript-powered map interface called Leaflet (leafletjs.com/). The setting of Mapnik's scale-dependent cartographic rules enables the display of some features earlier and others later in the map style. Rasterized versions of the original vector map data are used because they load much quicker when users pan and zoom across the map base.

One exception to this rule is the representation of coastlines, which are derived from two different simplified coastline datasets rather than crowd-sourced data—one for coarsely granular levels with simplified boundaries, and the second for fine levels of granularity with more complex, detailed coastlines (see the discussion of scale in Box 2.3). Figure 3.18 illustrates several different zoom levels of OpenStreetMap for Times Square, New York—the noncommercial nature of this Open Data (Section 17.4) product means that the map image includes a mix of points of interest uploaded by volunteers and map content is not determined by commercial sponsorship.

Similar methods for serving rasterized versions of vector data were used in early versions of services provided by the commercial online map services. However, the ever-increasing use of handheld GPS-enabled devices such as smartphones has brought four issues to the fore. First, the output of software like Mapnik is a series of rasterized maps at fixed zoom levels; this means that map labels (such as points of interest or street names) are fixed in position and orientation, so when the map is rotated (for example, using the compass facility of a smartphone), the labels may appear upside down. Second, map labels may be obscured by overlay information such as the direction path rendered on top of the map. Third, the jumps between discrete prerendered representations that are each at fixed scale are not conducive to the pinching gestures that users have become accustomed to using on recent-generation devices. Fourth, although raster data can be served very efficiently and rapidly by the map server, slow mobile data networks may lack the bandwidth to serve up successive tiles fast enough for user requirements. (Users can avoid this fourth problem by loading a map of an area they will visit and then navigating around using GPS even without a mobile connection. Cached maps are much used in tourist maps products, e.g., offmaps.com.)

More recent versions of commercial online mapping services use vector maps, in which all coastlines, roads, labels, and other data are represented as mathematical lines rather than as fixed rasterized graphic images. (The rendering of the map image is still done by the server, so the client smartphone or other device still receives raster images.) This means that when a mobile user rotates a map, the text labels dynamically reorient in order to remain legible, and the text size of labels scales smoothly when a user zooms in or out. This is possible because the labels are rendered live as dynamic text, rather than as a graphic image that must be "repainted" for every zoom level.



Courtesy: Alistair Leak; base image courtesy OpenStreetMap.

Figure 3.18 Web mapping at successive levels of detail using OpenStreetMap, centered on Times Square, New York.