8 Data Collection

LEARNING OBJECTIVES

Data collection is one of the most time-consuming and expensive, yet important, tasks associated with geographic information (GI). There are many diverse sources of GI, and many methods are available to enter them into a GI system. The two main methods of data collection are data capture and data transfer. It is useful to distinguish between primary (direct measurement) and secondary (derivation from other sources) data capture for both raster and vector data types. Data transfer involves importing digital data from other sources such as geoportals. Planning and executing an effective data collection plan touches on many practical issues. This chapter reviews the main methods of data capture and transfer and introduces key practical management issues.

After studying this chapter you will be able to:

- Describe data collection workflows.
- Understand the primary techniques of data capture in remote sensing and surveying.
- Discuss secondary techniques of data capture, including scanning, digitizing, vectorization, photogrammetry, and COGO (a contraction of the term *coordinate geometry*) feature construction.
- Understand the principles of data transfer, where to search for digital GI, and GI formats.
- Analyze practical issues associated with managing data capture projects.

8.1 Introduction

GI systems can contain a wide variety of geographic data types originating from many diverse sources. Indeed, one of the key defining characteristics of GI systems is their ability to integrate data about places and spaces from many different sources. Data collection activities for the purposes of organizing the material in this chapter are split into data capture (direct data input) and data transfer (input of data from other systems). From the perspective of creating geographic databases, it is convenient to classify both raster and vector geographic data as either primary or secondary (Table 8.1). Primary data sources are those collected in digital format specifically for use in a GI project.

Typical examples of primary GI sources include raster SPOT and Quickbird Earth satellite images and vector building-survey measurements captured using a total survey station (see Section 8.2 for discussion of these terms). Secondary sources are digital and analog datasets that may have been originally captured for another purpose and need to be converted into a suitable digital format for use in a GI project. Typical secondary sources include raster-scanned color aerial photographs of urban areas, or United States Geological Survey (USGS) or Institut Géographique National (IGN, France) paper maps that can be scanned and vectorized. There is a new trend for citizens to collect geographic data in the field directly into a GI system. Some issues surrounding volunteered geographic information are discussed in Section 2.5.

This classification scheme provides a useful organizing framework for this chapter, and more important, it highlights the number of processing transformations that a dataset goes through, and therefore the opportunities for errors to be introduced. However, the distinctions between primary and secondary and between raster and vector are not always easy to determine. For example, are digital satellite

Table 8.1 Classification of geographic data for data collection purposes with examples of each type.

	Raster	Vector
Primary	Digital satellite remote-sensing images	GPS measurements
	Digital aerial photographs	Field survey measurements
Secondary	Scanned maps or photographs	Topographic maps
	Digital elevation models from topographic map contours	Toponymy (place-name) databases

remote-sensing data obtained on a DVD primary or secondary? Clearly, the commercial satellite sensor feeds do not run straight into GI databases, but to ground stations where the data are preprocessed and then written onto digital media. Here they are considered primary because the data have usually undergone only minimal transformation since being collected by the satellite sensors and because the characteristics of the data make them suitable for direct use in GI projects.

Primary geographic data sources are captured by direct measurement specifically for use in GI systems. Secondary sources are reused from earlier studies or obtained from other systems.

Both primary and secondary geographic data may be obtained in either digital or analog format (see Section 3.7 for a definition of analog), although digital data are becoming much more common. Analog data must always be digitized before being added to a geographic database. Analog-to-digital transformation may involve scanning paper maps or photographs, optical character recognition (OCR) of text describing geographic object properties, or vectorization of selected features from an image. Depending on the format and characteristics of the digital data, considerable reformatting and restructuring may be required prior to importing into a GI system. Each of these transformations alters the original data and will introduce further uncertainty into the data (see Chapter 5 for a discussion of uncertainty).

This chapter describes the data sources, techniques, and workflows involved in data collection. The processes of data collection are also variously referred to as data capture, data automation, data conversion, data transfer, data translation, and digitizing. Although there are subtle differences between these terms, they essentially describe the same thing, namely, adding geographic data to a database. Here data capture refers to direct entry, and data transfer is the importing of existing digital data across a network connection (Internet, WAN, or LAN) or from physical media such as DVD or portable hard disk. This chapter focuses on the techniques of data collection; of equal, or perhaps more, importance to a real-world GI system implementation are project management, cost, legal, and organization issues. These are covered briefly in Section 8.7 as a prelude to more detailed treatment in Chapters 16 through 18.

Table 8.2 shows a breakdown of costs (in $1,000s) for two typical client–server GI system implementations: one with 10 seats (systems) and the other with 100. The hardware costs include desktop clients and servers only (i.e., not network infrastructure). The data costs assume the purchase of a landbase (e.g., streets, parcels, and landmarks) and digitizing assets such as pipes and fittings (water utility), conductors, and devices (electrical utility), or land and property parcels (local government). Staff costs assume that all core staff will be full time, but that users will be part time.

In the early days when geographic data were very scarce, data collection was the main project task, and typically it consumed the majority of the available resources. Even today, with the widespread availability of maps on the Internet, data collection still remains a time-consuming, tedious, and expensive process in many serious GI projects. Typically, it accounts for 15 to 50% of the total cost of a project (Table 8.2). Data capture costs can in fact be much more significant because in many organizations (especially those that are government funded) staff

Table 8.2 Breakdown of costs (in $1,000s) for two typical client–server GI systems as estimated by the authors.

	10 seats		100 seats	
	$	%	$	%
Hardware	30	3.4	250	8.8
Software	25	2.8	150	5.3
Data	400	44.7	450	15.8
Staff	440	48.2	2,000	70.2
Total	885	100.0	2,850	100.0

costs are often assumed to be fixed and are not used in budget accounting. Furthermore, as the majority of data capture effort and expense tends to fall at the start of projects, data capture costs often receive greater scrutiny from senior managers. If staff costs are excluded from a GI project budget, then in cash expenditure terms, data collection can be as much as 60 to 85% of costs.

Data capture costs can account for up to 85% of the cost of a GIS.

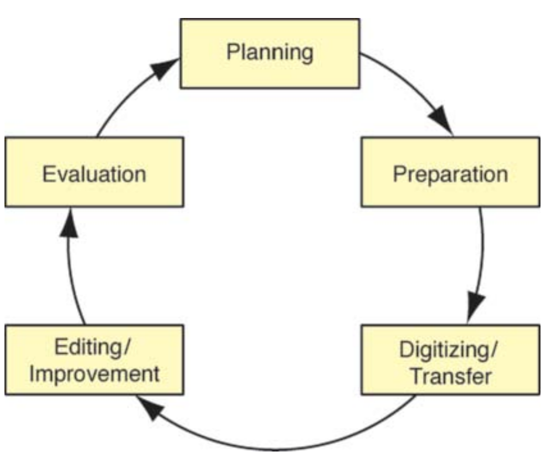
After an organization has completed basic data collection tasks the focus of a GI project moves on to data maintenance. Over the multiyear lifetime of a GI project, data maintenance can turn out to be a far more complex and expensive activity than initial data collection. This is because of the high volume of update transactions in many systems (for example, changes in land parcel ownership, maintenance work orders on a highway transport network, or logging military operational activities) and the need to manage multiuser access to operational databases. For more information about geographic data maintenance see Chapter 9.

8.1.1 Data Collection Workflow

In all but the simplest of projects, data collection involves a series of sequential stages (Figure 8.1). The workflow commences with planning, followed by preparation, digitizing/transfer (here taken to mean a range of primary and secondary techniques such as table digitizing, survey entry, scanning, and photogrammetry), editing and improvement, and finally, evaluation.

Planning is obviously important to any project, and data collection is no exception. It includes establishing user requirements, garnering resources (staff, hardware, and software), and developing a project plan. Preparation is especially important in data collection projects. It involves many tasks such as obtaining data, redrafting poor-quality map sources, editing scanned map images, and removing noise (unwanted data such as speckles on a scanned map image). It may also involve setting up appropriate hardware and software systems to accept data. Digitizing and transfer are the stages where the majority of the effort will be expended. It is naive to think that data capture is really just digitizing, when in fact it involves very much more as discussed later in this chapter. Editing and improvement follows digitizing/transfer. This covers many techniques designed to validate data, as well as to correct errors and to improve quality. Evaluation, as the name suggests, is the process of identifying project successes and failures; these may be qualitative or quantitative. Because all large data projects involve multiple stages, this workflow is iterative, with earlier phases (especially a first, pilot, phase) helping to improve subsequent parts of the overall project.

Figure 8.1 Stages in data collection projects.

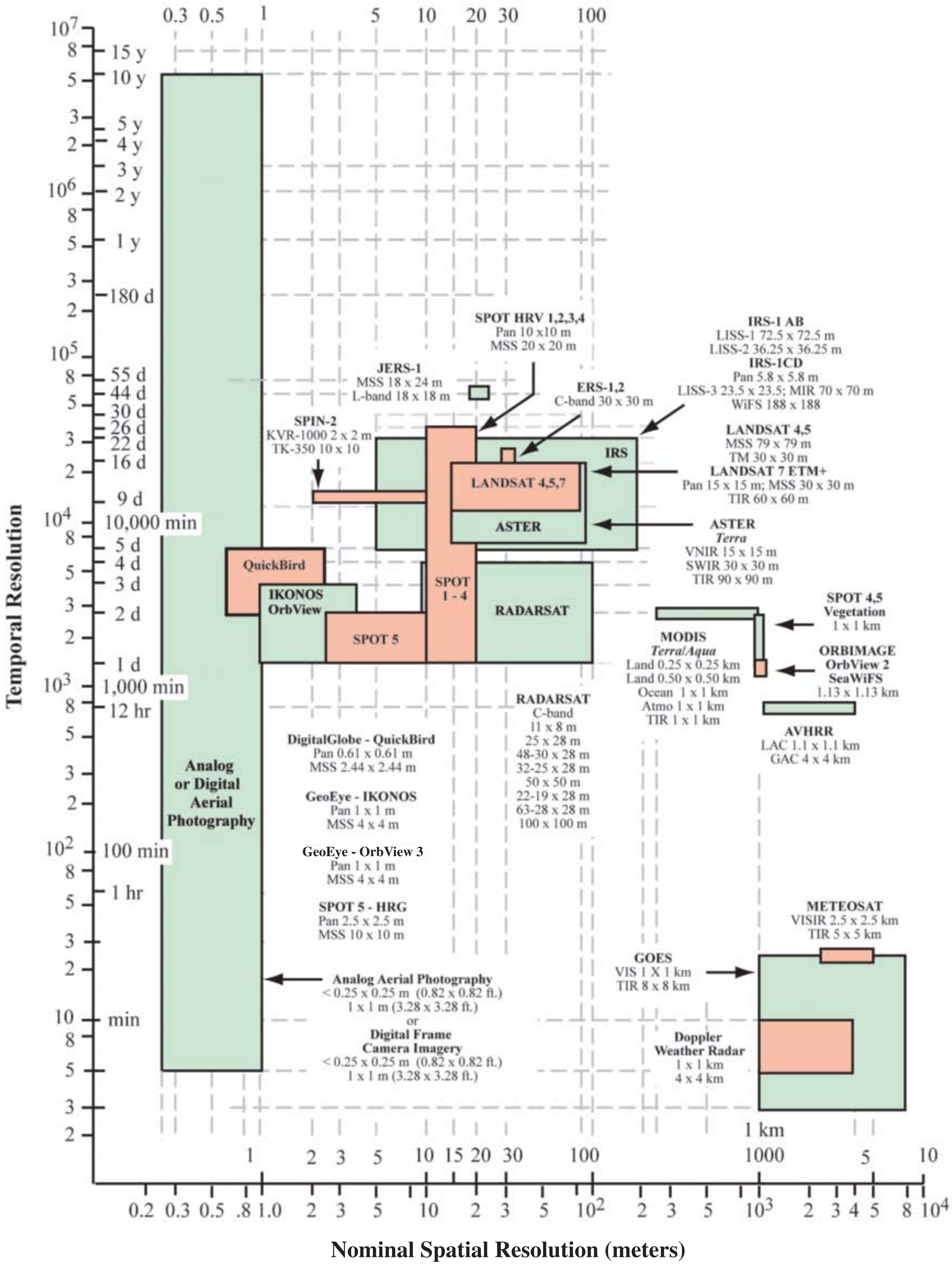


8.2 Primary Geographic Data Capture

Primary geographic capture involves the direct measurement of objects. Digital data measurements may be input directly into the GI database or may reside in a temporary file prior to input. Although the direct method is preferable as it minimizes the amount of time and the possibility of errors, close coupling of data collection devices and GI databases is not always possible. Both raster and vector primary data capture methods are available.

8.2.1 Raster Data Capture

The most popular form of primary raster data capture is remote sensing. Broadly speaking, remote sensing is a technique used to derive information about the physical, chemical, and biological properties of objects without direct physical contact. Information is derived from measurements of the amount of electromagnetic radiation reflected, emitted, or scattered from objects. A variety of sensors, operating throughout the electromagnetic spectrum from visible to microwave wavelengths, are commonly employed to obtain measurements (see Section 3.6.1). Passive sensors rely on reflected solar radiation or emitted terrestrial radiation; active sensors (such as synthetic aperture radar) generate their own source of electromagnetic radiation. The platforms on which these instruments are mounted are similarly diverse. Although Earth-orbiting satellites and fixed-wing aircraft are by far the most common, helicopters, balloons, masts, booms, and even handheld devices are also employed (Figure 8.2). As used here,



Source: John Jensen

Figure 8.2 Spatial and temporal characteristics of commonly used Earth observation remote-sensing systems and their sensors.

the term *remote sensing* subsumes the fields of satellite remote sensing and aerial photography.

> **Remote sensing is the measurement of physical, chemical, and biological properties of objects without direct contact.**

From the perspective of GI, resolution is a key physical characteristic of remote-sensing systems. There are three aspects to resolution: spatial, spectral, and temporal. All sensors need to trade off spatial, spectral, and temporal properties because of storage, processing, and bandwidth considerations. For further discussion of the important topic of resolution, see also Sections 2.3, 5.3.2, and 15.1 and Box 2.3.

Three key aspects of resolution are spatial, spectral, and temporal.

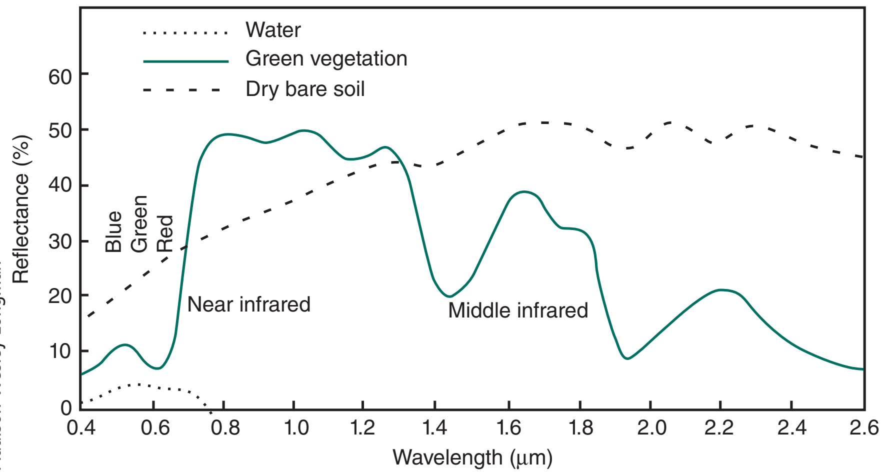
Spatial resolution refers to the size of object that can be resolved, and the most usual measure is the pixel size. Satellite remote-sensing systems typically provide data with pixel sizes in the range 0.4 m–1 km. The resolution of digital cameras used for capturing aerial photographs usually range from 0.01 m–5 m. Image (scene) sizes vary quite widely between sensors—typical ranges include 900 by 900 to 3,000 by 3,000 pixels. The total coverage of remote-sensing images is usually in the range 9 by 9 km to 200 by 200 km.

Spectral resolution refers to the parts of the electromagnetic spectrum that are measured. Because different objects emit and reflect different types and amounts of radiation, selecting which part of the electromagnetic spectrum to measure is critical for each application area. Figure 8.3 shows the spectral signatures of water, green vegetation, and dry soil. Remote-sensing systems may capture data in one part of the spectrum (referred to as a single band) or simultaneously from several parts (multiband or multispectral). The radiation values are usually normalized and resampled to give a range of integers from 0–255 for each band (part of the electromagnetic spectrum measured), for each pixel, in each image. Until recently, remote-sensing satellites typically measured a small number of bands, in the visible part of the spectrum. More recently, a number of hyperspectral systems have come into operation that measure very large numbers of bands across a much wider part of the spectrum. Temporal resolution, or repeat cycle, describes the frequency with which images are collected for the same area. There are essentially two types of commercial remote-sensing satellite: Earth-orbiting and geostationary. Earth-orbiting satellites collect information about different parts of the Earth surface at regular intervals. To maximize utility, typically orbits are polar, at a fixed altitude and speed, and are sun synchronous.

The French SPOT (Système Probatoire d'Observation de la Terre) 6 satellite launched in 2012, for example, orbits over the poles at an altitude of 694 km, sensing the same location on the Earth surface every 1 to 3 days. The SPOT 6 platform carries multiple sensors: a panchromatic sensor measuring radiation in the visible part of the electromagnetic spectrum at a spatial resolution of 1.5 m; and a multispectral sensor measuring blue, green, red, and near infrared radiation separately at a spatial resolution of 6 m. The SPOT 6 system is also able to provide stereo images from which digital elevation models and 3-D measurements can be obtained. SPOT 6 scenes cover areas from about 60 by 60 km to 120 by 120 km. A number of digital image products are available from the ground distribution center: orthoimages (rectified to the ground), stereo pairs for production of digital elevation models, and various image mosaics.

Much of the discussion so far has focused on commercial satellite remote-sensing systems. Of equal importance, especially in GI projects that focus on smaller areas in great detail, is aerial photography. Although the data products resulting from remote-sensing satellites and digital aerial photography systems are technically very similar (i.e., they are both images), there are some significant differences in the

Figure 8.3 Typical reflectance signatures for water, green vegetation, and dry soil.



Source: after Jones, C. 1997. *Geographic Information Systems and Computer Cartography*. Reading, MA: Addison-Wesley Longman way data are captured and can, therefore, be analyzed and interpreted. One notable difference is that older aerial photography systems use analog optical cameras and then later rasterize the photographs (e.g., scanning a film negative) to create an image. Both the quality of the optics of the camera and the mechanics of the scanning process affect the spatial and spectral characteristics of the resulting images. Modern aerial photography systems now use digital cameras with on-board position-fixing units to collect digital imagery (although the term *aerial photography* still persists when the digital cameras are carried on planes). Most aerial photographs are collected on an ad hoc basis using cameras mounted in airplanes flying at low altitudes (3,000–9,000 m) and are either panchromatic (black and white) or color, although multispectral cameras/sensors operating in the nonvisible parts of the electromagnetic spectrum are also used. Aerial photographs are very suitable for detailed surveying and mapping projects.

An important feature of satellite and aerial photography systems is that they can provide stereo imagery from overlapping pairs of images. These images are used to create a 3-D analog or digital model from which 3-D coordinates, contours, and digital elevation models can be created (see Section 8.3.2.3).

Satellite and aerial photograph data offer a number of advantages for GI projects. The consistency of the data and the availability of systematic global coverage make satellite data especially useful for large-area, very detailed projects (for example, mapping landforms and geology at the river catchment-area level) and for mapping inaccessible areas. The regular repeat cycles of commercial systems and the fact that they record radiation in many parts of the spectrum make such data especially suitable for assessing the condition of vegetation (for example, the moisture stress of wheat crops). Aerial photographs in particular are very useful for detailed surveying and mapping of, for example, urban areas and archaeological sites, especially those applications requiring 3-D data (see Chapter 11).

On the other hand, the spatial resolution of commercial satellites is too coarse for many detailed projects, and the data collection capability of many sensors is restricted by cloud cover. Some of this is changing, however, as the new generation of satellite sensors now provide data at 0.4 m spatial resolution and better, and radar data can be obtained that are not affected by cloud cover. The data volumes from both satellites and aerial cameras can be very large and create storage and processing problems for all but the most modern systems. The cost of data can also be prohibitive for a single project or organization.

Box 8.1 describes the work of Huadong Guo.

Biographical Box 8.1

Huadong Guo

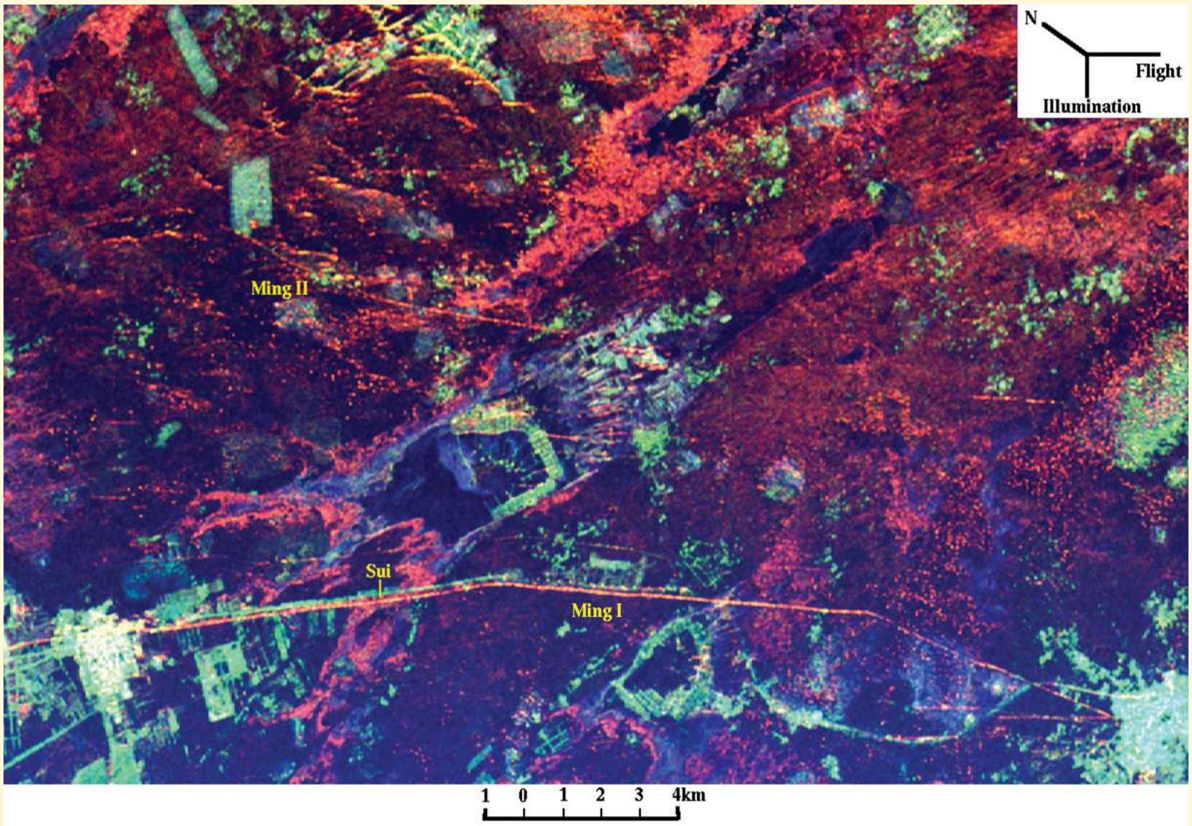
Huadong Guo (Figure 8.4) received his postgraduate degree in Cartography and Remote Sensing from the University of the Chinese Academy of Sciences (UCAS). He is a Professor and Director-General of the Chinese Academy of Sciences (CAS) Institute of Remote Sensing and Digital Earth (RADI) and a Member of CAS. He presently serves as President of the International Council for Science (ICSU) Committee on Data for Science and Technology (CODATA), Secretary-General of the International Society for Digital Earth (ISDE), and Editor-in-Chief of the *International Journal of Digital Earth* (IJDE), among other organizations. He has published more than 300 papers and 15 books and is the principal awardee of thirteen national and CAS prizes, one being "National Outstanding Expert," awarded by the State Council of China.

Courtesy: Huadong Guo

Figure 8.4 Huadong Guo.

Guo has over thirty years of experience in remote sensing, specializing in radar for Earth observation and remote-sensing applications (see, for example, Figure 8.5) and has been involved in research on Digital Earth since the end of the previous century. Through his research in radar remote sensing, Guo established a nonvegetation sand dune radar backscattering model, developed radar polarimetric theory, researched the depolarization phenomena of volcanic lava flow, and revealed the relationship between radar

Figure 8.5 Remote-sensing image of the Great Wall. SIR-C Image showing three segments of the Great Wall, of which two were built in the Ming dynasty (1474 A.D. and 1531 A.D.), and one was built in the Sui dynasty (585 A.D.) between the Ming Wall and the road.



Courtesy: Huadong Guo

phase information and vegetation parameters. He is a forerunner in the use of full polarization radar and has achieved high-accuracy crop classification mapping.

With the unceasing development of science and technology, the role of data is becoming increasingly important. Scientists are surrounded by petabyte-level and even exabyte-level data in the realm of remote sensing. Guo has established an all-weather, day and night remote-sensing monitoring system for disaster reduction. The system was used in the aftermath of the Wenchuan and Yushu earthquakes in 2008 and 2010. Data were made available to all partners who needed the acquired data. In all, 4,851 GB of data were shared with 15 government agencies.

Since the advent of the Digital Earth concept, Guo has actively promoted the vision of Digital Earth. He led a research team to build up the first Digital Earth prototype system in China. This system comprises integrated, advanced technology both in hardware and software to conduct Earth system science at global, national, and regional scales.

Guo's research and contributions in radar remote-sensing theory, integrated technologies for mineral exploration, Earth observation for disaster mitigation, and Digital Earth have demonstrated the importance of remote sensing and its applications.

8.2.2 Vector Data Capture

Primary vector data capture is a major source of geographic data. The two main branches of vector data capture are ground surveying and GPS (which is covered in Section 4.9), although as more surveyors use GPS routinely, so the distinction between the two is becoming increasingly blurred. This is also leading to a fundamental shift from measurement-based to coordinate-based ground surveying (i.e., surveyors in the field work with geographic coordinates, rather than angle and distance measurements, which are harder to interpret).

8.2.2.1 Surveying

Ground surveying is based on the principle that the 3-D location of any point can be determined by measuring angles and distances from other known points.

Surveys begin from a benchmark point. If the coordinate system and location of this point are known, all subsequent points can be collected in this coordinate system. If they are unknown, then the survey will use a local or relative coordinate system (see Section 4.8).

Because all survey points are obtained from survey measurements, their known locations are always relative to other points. Any measurement errors need to be apportioned between multiple points in a survey. For example, when surveying a field boundary, if the last and first points are not identical in survey terms (within the tolerance employed in the survey), then errors need to be apportioned between all points that define the boundary (see Sections 5.3.2 and 5.4.2). As new measurements are obtained, these may change the locations of points.

Traditionally, surveyors used equipment such as transits and theodolites to measure angles, and tapes and chains to measure distances. Today these have been replaced by electro-optical devices called total stations that can measure both angles and distances to an accuracy of 1 mm (Figure 8.6). Total stations automatically log data, and the most sophisticated can create vector point, line, and area objects in the field, thus providing direct validation. The basic principles of surveying have changed very little in the past 100 years, although new technology has considerably improved accuracy and productivity. Two people are usually required to perform a survey, one to operate the total station and the other to hold a reflective prism that is placed at the object being measured. On some remote-controlled systems a single person can control both the total station and the prism.

Ground survey is a very time-consuming and expensive activity, but it is still the best way to obtain highly accurate point locations. Surveying is typically used for capturing buildings, land and property boundaries, manholes, and other objects that need to be located accurately. It is also used to obtain reference marks for use in other data capture projects. For example, detailed, fine-scale aerial photographs and satellite images are frequently georeferenced using points obtained from ground survey.

8.2.2.2 LiDAR

LiDAR (light detection and ranging, also known as airborne laser swath mapping, or ALSM) is a relatively recent technology that employs a scanning laser range finder to produce accurate topographic surveys of great detail. A LiDAR scanner is an active remote-sensing instrument; that is, it transmits electromagnetic radiation and measures the radiation that is scattered back to a receiver after interacting with the Earth's atmosphere or objects on the surface. LiDAR uses radiation in the ultraviolet, visible, or infrared region of the electromagnetic spectrum. The scanner is typically carried on a low-altitude aircraft that also has an inertial navigation system and a differential GPS to provide location. LiDAR scanners are capable of collecting extremely large quantities of very detailed information (i.e., scanning of the order of 30,000 points per second at an accuracy of around 15 cm). The data collected from a LiDAR scanner can be described as a point cloud, that is, a massive collection of independent points with (x, y, z) values. After initial data capture, extensive processing is usually required to remove tree canopies, buildings, and other unwanted features and to correct errors in order to provide a "bare Earth" point dataset. The points in a LiDAR dataset are often rasterized to create a digital

Figure 8.6 A tripod-mounted total station.

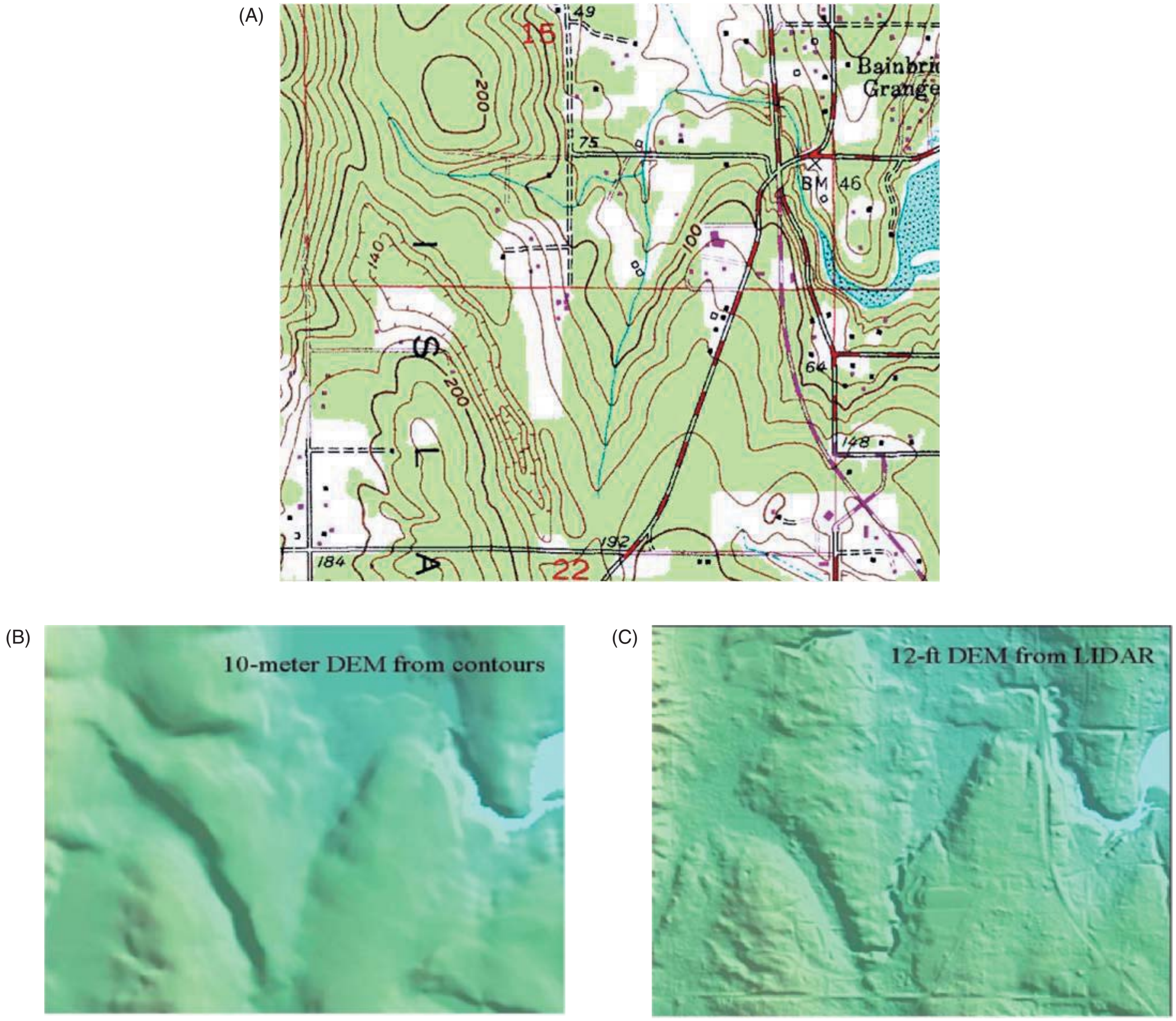


Figure 8.7 Comparison of three datasets for 1 square mile of Bainbridge Island, Washington State: (A) scanned USGS 1:24,000 topographic map sheet; (B) 10-m digital elevation model (DEM) derived from contours digitized from a map sheet; (C) 12-ft (365-cm) resolution DEM derived from a LiDAR survey. (From pugetsoundlidar.ess.washington.edu/example2.htm)

elevation model that is smaller in size and easier to work with in a GI system. Figure 8.7 presents a comparison of three datasets for the same area, one of which is derived from LiDAR.

8.3 Secondary Geographic Data Capture

Geographic data capture from secondary sources is the process of creating raster and vector files and databases from maps, photographs, and other hardcopy documents. Scanning is used to capture raster data. Heads-up digitizing, stereo-photogrammetry, and COGO data entry are the most widely used methods for capturing vector data.

8.3.1 Raster Data Capture Using Scanners

A scanner is a device that converts hardcopy analog media into digital images by scanning successive lines across a map or document and recording the amount of light reflected from a local data source (Figure 8.8). The differences in reflected light are normally scaled into bi-level black and white (1 bit per pixel) or multiple gray levels (8, 16, or 32 bits). Color scanners typically output data into 8-bit red, green, and blue color bands. The spatial resolution of scanners varies widely from as little as 200 dpi (8 dots per mm) to 2400 dpi (86 dots per mm) and beyond. Most GIS scanning

Source: Contex

Figure 8.8 A large-format roll-feed image scanner.

is in the range 400–800 dpi (16–40 dots per mm). Depending on the type of scanner and the resolution required, it can take from 30 seconds to 30 minutes or more to scan a map.

Scanned maps and documents are used extensively as background maps and data stores.

There are three main reasons to scan hardcopy media for use in GI systems:

- Documents, such as building plans, CAD drawings, property deeds, and equipment photographs are scanned to reduce wear and tear, to improve access, to provide integrated database storage, and to index them geographically (e.g., building plans can be attached to building objects in geographic space).
- Film and paper maps, aerial photographs, and images are scanned and georeferenced so that they provide geographic context for other data (typically vector layers). This type of unintelligent image or background geographic wallpaper is very popular in systems that manage equipment and land and property assets (Figure 8.9).
- Maps, aerial photographs, and images are scanned prior to vectorization (see Section 8.3.2.1) and sometimes as a prelude to spatial analysis.

An 8-bit (256 gray level) 400 dpi (16 dots per mm) scanner is a good choice for minimum resolution for scanning maps to be used as a background

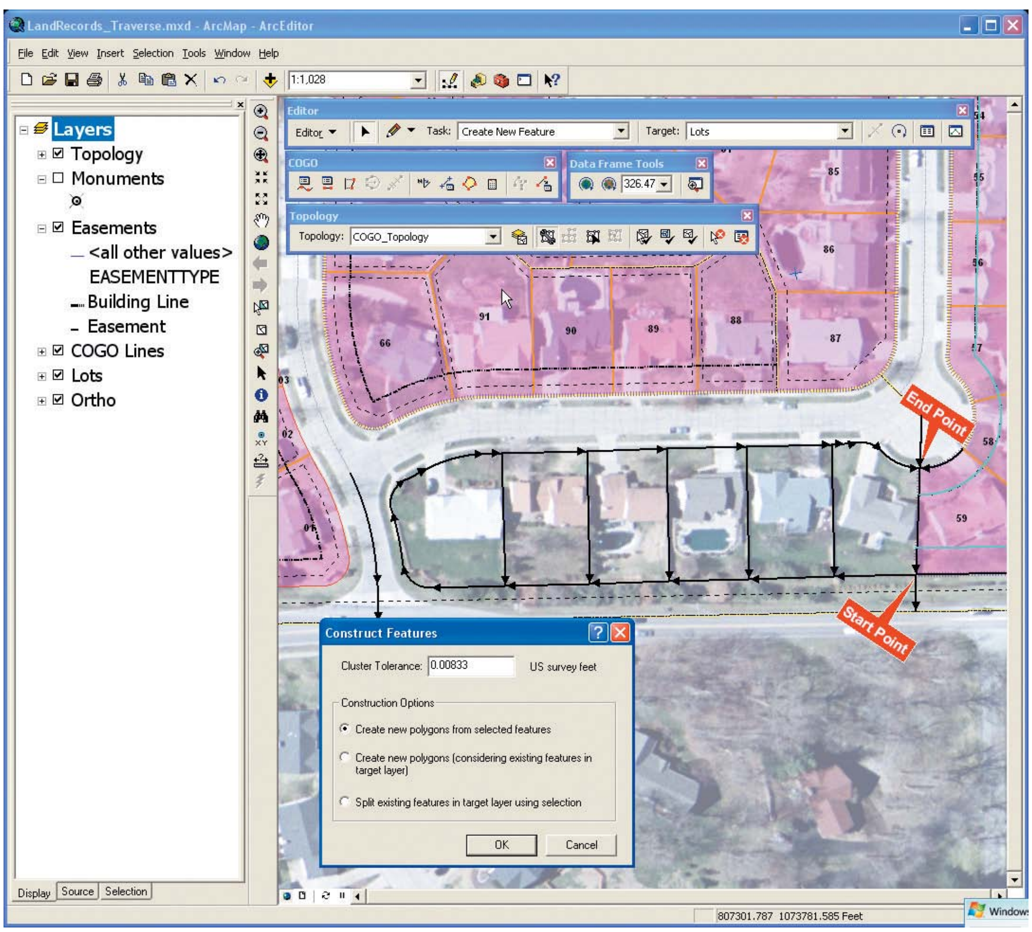


Figure 8.9 An example of raster background data (color aerial photography) underneath vector data (land parcels) that are being digitized on-screen.

reference layer. For a color aerial photograph that is to be used for subsequent photo interpretation and analysis, a color (8 bit for each of three bands) 900 dpi (40 dots per mm) scanner provides a more appropriate minimum resolution. The quality of data output from a scanner is determined by the nature of the original source material, the quality of the scanning device, and the type of preparation prior to scanning (e.g., redrafting key features or removing unwanted marks will improve output quality).

8.3.2 Vector Data Capture

Secondary vector data capture involves digitizing vector objects from maps and other geographic data sources. The most popular methods are heads-up digitizing and vectorization, photogrammetry, and COGO data entry. Historically, manual digitizing from digitizing tables was the most popular way of secondary vector data capture, but its use has now almost entirely been replaced by the other more modern methods.

8.3.2.1 Heads-up Digitizing and Vectorization

One of the main reasons for scanning maps (see Section 8.3.1) is as a prelude to vectorization—the process of converting raster data into vector data. The simplest way to create vectors selectively from raster data is to digitize vector objects manually straight off a computer screen using a mouse or digitizing cursor. This method is called heads-up digitizing because the map is vertical and can be viewed without bending the head down. It is widely used for capturing, for example, land parcels, buildings, and utility assets.

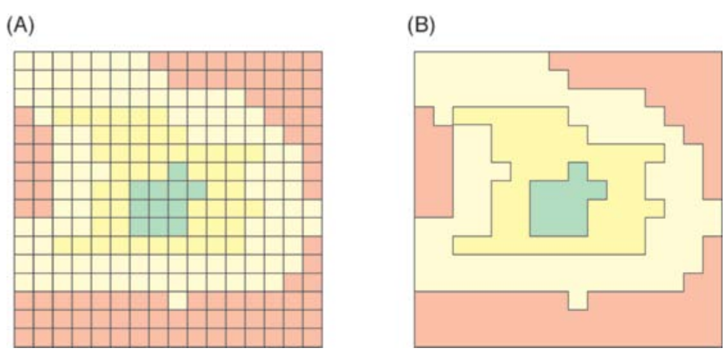
Vectorization is the process of converting raster data into vector data. The opposite is called rasterization.

Any type of image derived from either primary sources, such as satellite images or aerial photographs, or secondary sources, such as maps or other documents, can be used in heads-up digitizing and vectorization. After loading a scanned image into a GI database, it must be georeferenced (the term *georegistration* is also sometimes used synonymously: see Section 4.12) before digitizing can begin. This involves a geometric transformation process that uses well-known algorithms to convert image coordinates into database coordinates. The algorithms use both image and database coordinates from a minimum of three well-defined reference points. Once the transformation has been set up, any future coordinates digitized from the image will be in the database coordinate reference system.

Vertices defining point, polyline, and polygon objects are captured on-screen using point- or stream-digitizing methods. Point digitizing involves placing the screen cursor at the location for each object vertex and then clicking a button to record the location of the vertex. Stream-mode digitizing partially automates this process by collecting vertices automatically every time a distance or time threshold is crossed (e.g., every 0.02 inch (0.5 mm) or 0.25 second). Stream-mode digitizing is a much faster method, but it typically produces larger files with many redundant coordinates (although these can be generalized using standard algorithms).

A faster and more consistent approach is to use software to perform automated vectorization in either batch or semi-interactive mode. Batch vectorization takes an entire raster file and converts it to vector objects in a single operation. Vector objects are created using software algorithms that build simple (spaghetti) line strings from the original pixel values. The lines can then be further processed to create topologically correct polygons (Figure 8.10). A typical map will take only a few minutes to vectorize using modern hardware and software systems. See Section 9.7 for further discussion of structuring geographic data.

Figure 8.10 Batch vectorization of a scanned map: (A) original raster file; (B) vectorized polygons. Adjacent raster cells with the same attribute values are aggregated. Class boundaries are then created at the intersection between adjacent classes in the form of vector lines.



Unfortunately, batch vectorization software is far from perfect, and postvectorization editing is usually required to clean up errors. To avoid large amounts of vector editing, it is useful to undertake a little editing of the original raster file prior to vectorization in order to remove unwanted noise that may affect the vectorization process. For example, text that overlaps lines should be deleted, and dashed lines are best converted to solid lines. Following vectorization, topological relationships are usually created for the vector objects. This process may also highlight some previously unnoticed errors that require additional editing.

Batch vectorization is best suited to simple bi-level maps of, for example, contours, streams, and highways. For more complicated images and maps and where selective vectorization is required (for example, digitizing electric conductors and devices, or water mains and fittings off topographic maps), interactive vectorization (also called semiautomatic vectorization, line following, or tracing) is preferred. In interactive vectorization, software is used to automate digitizing. The operator snaps the cursor to a pixel on the screen, indicates a direction for line following, and the software then automatically digitizes lines. Typically, many parameters can be tuned to control the density of points (level of generalization), the size of gaps (blank pixels in a line) that will be jumped, and whether to pause at junctions for operator intervention or always to trace in a specific direction (most systems require that all polygons are ordered either clockwise or counterclockwise). Interactive vectorization is still quite labor intensive, but generally it results in much greater productivity than manual or heads-up digitizing. It also produces high-quality data, as software is able to represent lines more accurately and consistently than can humans. For these reasons specialized data capture groups much prefer vectorization to manual digitizing.

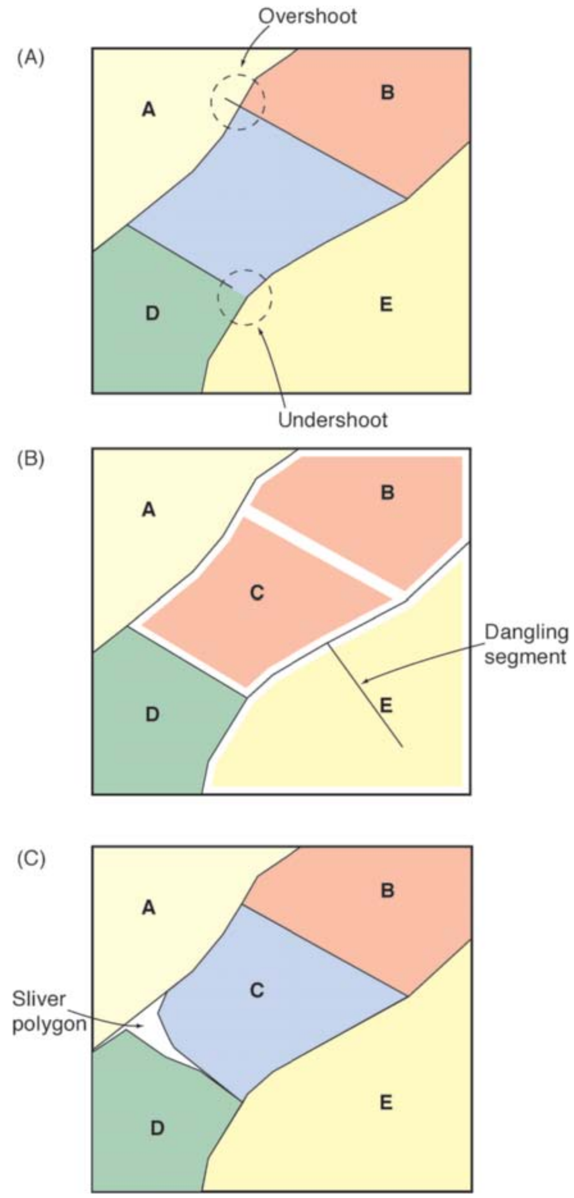


Figure 8.11 Examples of human errors in digitizing: (A) undershoots and overshoots; (B) invalid polygons; and (C) sliver polygons.

8.3.2.2 Measurement Error

Data capture, like all geographic workflows, is likely to generate both measurement and operator errors. Because digitizing is a tedious and hence error-prone practice, it presents a source of operator errors—as when the operator fails to position the cursor correctly or fails to record line segments. Figure 8.11 presents some examples of human errors that are commonly introduced in the digitizing process. These errors are overshoots and undershoots where line intersections are inexact (Figure 8.11A); invalid polygons that are topologically inconsistent because of omission of one or more lines, or omission of attribute data (Figure 8.11B); and sliver polygons, in which multiple digitizing of the common boundary between adjacent polygons leads to the creation of additional polygons (Figure 8.11C).

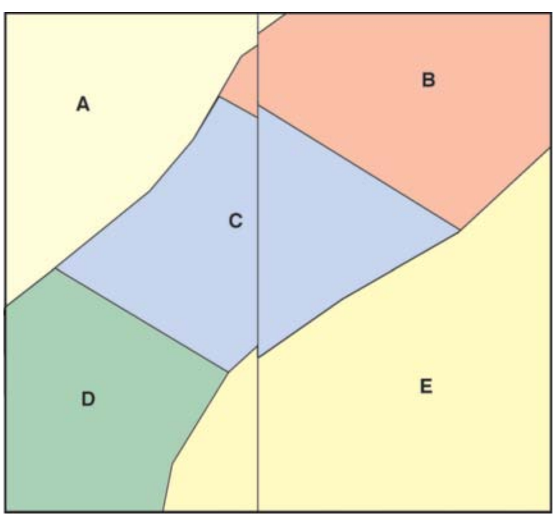
Most GI software packages include standard functions which can be used to restore integrity and clean (or rather obscure, depending on your viewpoint!) obvious measurement errors. Such operations are best carried out immediately after digitizing, so that omissions may be easily rectified. Data-cleaning operations require sensitive setting of threshold values, or else damage can be done to real-world features, as Figure 8.12 shows.

Many errors in digitizing can be remedied by appropriately designed software.

Figure 8.12 Error induced by data cleaning. If the tolerance level is set large enough to correct the errors at A and B, the loop at C will also (incorrectly) be closed.

Further classes of problems arise when the products of digitizing adjacent map sheets are merged. Stretching of paper base maps, coupled with errors in rectifying them, give rise to the kinds of mismatches shown in Figure 8.13. Rubber sheeting is the term used to describe methods for removing such errors on the assumption that strong spatial autocorrelation exists among errors. If errors tend to be spatially autocorrelated up to a distance of x, say, then rubber sheeting will be successful at removing them, at least partially, provided control points can be found that are spaced less than x apart. For the same reason, the shapes of features that are less than x across will tend to have little distortion, whereas very large shapes may be badly distorted. The results of calculating areas (Section 14.1.1), or other geometric operations that rely only on relative position, will be accurate as long as the areas are small, but will grow rapidly with feature size. Thus it is important for the user to know which operations depend on relative position, and over what distance; and where absolute position is important (of course, the term *absolute* simply means relative to the Earth frame, defined by the Equator and the Greenwich Meridian, or relative over a very long distance; see Section 4.7). Analogous procedures and problems characterize the rectification of raster datasets, be they scanned images of paper maps or satellite measurements of the curved Earth surface.

Figure 8.13 Mismatches of adjacent spatial data sources that require rubber sheeting.



8.3.2.3 Photogrammetry

Photogrammetry is the science and technology of making measurements from pictures, aerial photographs, and images. Although in the strict sense it includes 2-D measurements taken from single aerial photographs, today it is almost exclusively concerned with capturing 2.5-D and 3-D measurements from models derived from stereopairs of photographs and images. In the case of aerial photographs, it is usual to have 60% overlap along each flight line and 30% overlap between flight lines. Similar layouts are used by remote-sensing satellites. The amount of overlap defines the area for which a 3-D model can be created.

Photogrammetry is used to capture measurements from photographs and other image sources.

To obtain true georeferenced Earth coordinates from a model, it is necessary to georeference photographs using control points (the procedure is essentially analogous to that described for digitizing in Section 8.3.2.1). Control points can be defined by ground survey, or nowadays more usually with GPS (see Section 8.2.2.1 for discussion of these techniques).

Measurements are captured from overlapping pairs of images using stereoplotters. These build a model and allow 3-D measurements to be captured, edited, stored, and plotted. Stereoplotters have undergone three major generations of development: analog (optical), analytic, and digital. Mechanical analog devices are seldom used today, whereas analytical (combined mechanical and digital) and digital (entirely computer based) are much more common. It is likely that digital (softcopy) photogrammetry will eventually replace mechanical devices entirely.

There are many ways to view stereo models, including a split screen with a simple stereoscope and the use of special glasses to observe a red/green

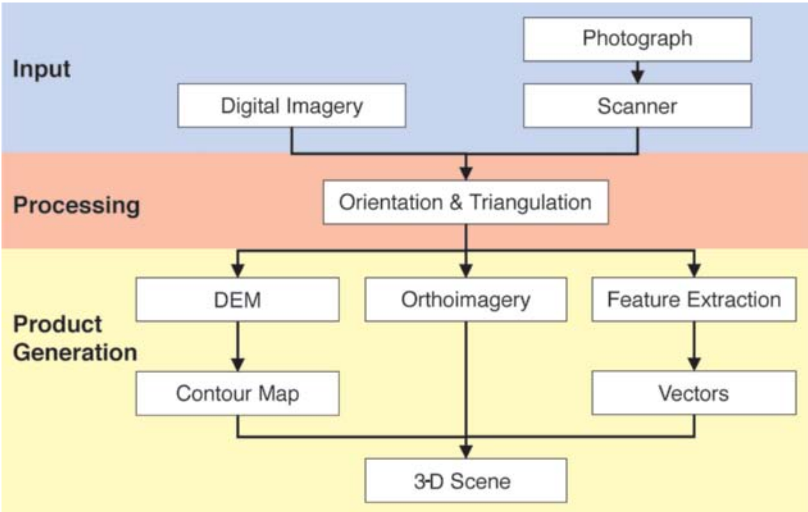


Figure 8.14 Typical photogrammetry workflow.

display or polarized light. To manipulate 3-D cursors in the *x*, *y*, and *z* planes, photogrammetry systems offer free-moving hand controllers, hand wheels and foot disks, and 3-D mice. The options for extracting vector objects from 3-D models are directly analogous to those available for manual digitizing as described earlier: namely, batch, interactive, and manual (Section 8.3.2.1). The obvious difference, however, is that there is a requirement for capturing *z* (elevation) values.

Figure 8.14 shows a typical workflow in digital photogrammetry derived from the work of Vincent Tao. There are three main parts to digital photogrammetry workflows: data input, processing, and product generation. Data can be obtained directly from sensors or by scanning secondary sources. Orientation and triangulation are fundamental photogrammetry processing tasks. Orientation is the process of creating a stereo model suitable for viewing and extracting 3-D vector coordinates that describe geographic objects. Triangulation (also called block adjustment) is used to assemble a collection of images into a single model so that accurate and consistent information can be obtained from large areas.

Photogrammetry workflows yield several important product outputs, including digital elevation models (DEMs), contours, orthoimages, vector features, and 3-D scenes. DEMs—regular arrays of height values—are created by matching stereo image pairs together using a series of control points. Once a DEM has been created, it is relatively straightforward to derive contours using a choice of algorithms. Orthoimages are images corrected for variations in terrain using a DEM so as to appear as if every point was seen from vertically above. They have become popular because of their relatively low cost of creation (when compared with topographic maps) and ease of interpretation as base maps except where tall buildings and other dramatic topographic features are present. They can also be used as accurate data sources for heads-up digitizing (see Section 8.3.2.1). Vector feature extraction is still an evolving field, and there are no widely applicable fully automated methods. The most successful methods use a combination of spectral analysis and spatial rules that define context, shape, proximity, and the like. Finally, 3-D scenes can be created by merging vector features with a DEM and an orthoimage (Figure 8.15).

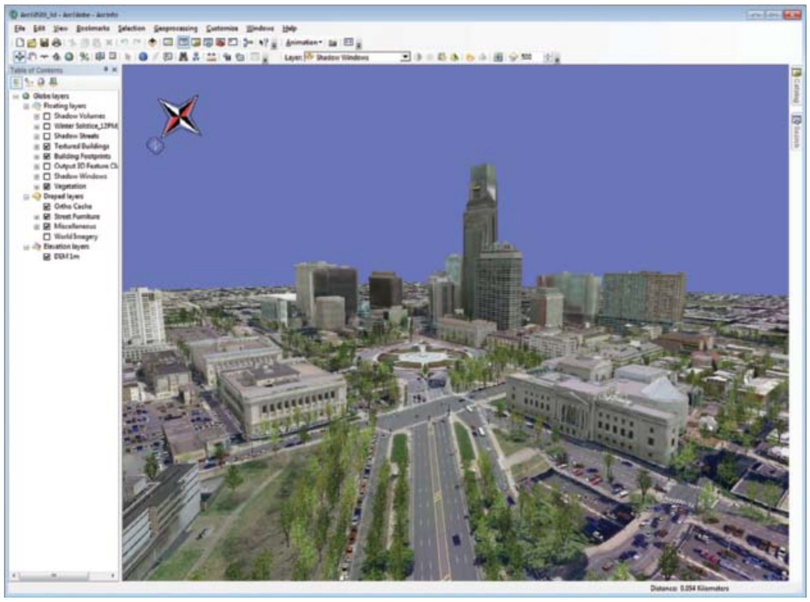
In summary, photogrammetry is a very cost-effective data capture technique that is sometimes the only practical method of obtaining detailed topographic data about an area of interest. Unfortunately, the complexity and high cost of equipment have restricted its use to primary data capture projects and specialist data capture organizations where very detailed information is required.

8.3.2.4 COGO Data Entry

COGO, which as noted earlier is a contraction of the term *coordinate geometry*, is a method for capturing and representing geographic data. COGO uses survey-style bearings and distances to define each part of an object in much the same way as described in Section 8.2.2. Figure 8.9 shows how land-parcel features can be created using COGO tools and then formed into topologically correct polygons. Some examples of COGO object-construction tools are shown in Figure 8.16. The Construct Along tool creates a point along a curve using a distance along the curve. The Line Construct Angle Bisector tool constructs a line that bisects an angle defined by a from-point, a through-point, a to-point, and a length. The Construct Fillet tool creates a circular-arc tangent from two segments and a radius.

The COGO system is widely used in North America to represent land records and property parcels (also called lots). Coordinates can be obtained

Figure 8.15 Automatically created 3-D model of central Philadelphia.

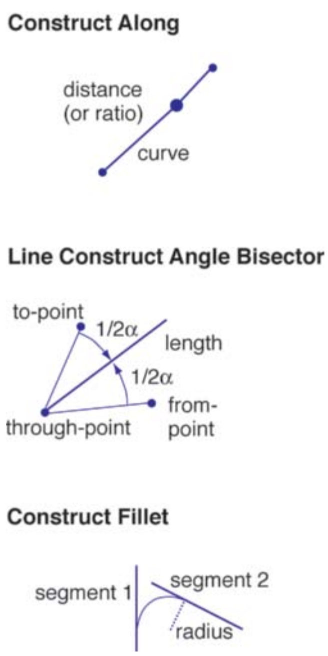


Source: Esri

from COGO measurements by geometric transformation (i.e., bearings and distances are converted into *x*, *y* coordinates). Although COGO data obtained as part of a primary data capture activity are used in some projects, it is more often the case that secondary measurements are captured from hardcopy maps and documents. Source data may be in the form of legal descriptions, records of survey, tract (housing estate) maps, or similar documents.

> **COGO stands for coordinate geometry. It is a vector data structure and method of data entry.**

Figure 8.16 Example COGO construction tools used to represent geographic features.



COGO data are very precise measurements and are often regarded as the only legally acceptable definition of land parcels. Measurements are usually very detailed, and data capture is often time consuming. Furthermore, commonly occurring discrepancies in the data must be manually resolved by highly qualified individuals.

8.4 Obtaining Data from External Sources (Data Transfer)

One major decision that needs to be faced at the start of a GI project is whether to build or buy part or all of a database. All of the preceding discussion has been concerned with techniques for building databases from primary and secondary sources. This section focuses on how to import or transfer data into a GI system that has been captured by others. Some datasets are freely available, but many of them are sold as commodities, mainly over the Web. Increasingly geographic data are being made available as direct-use services that can be applied in geographic analysis and mapping without having to import the data.

There are many sources and types of geographic data. Space does not permit a comprehensive review of all geographic data sources here, but a small selection of key sources is listed in Table 8.3. In any case, the characteristics and availability of datasets are constantly changing, so those seeking an up-to-date list should consult one of the online sources described in this section. Kerski and Clark offer a useful guide to public domain data. Chapter 17 also discusses the characteristics of geographic information and highlights several issues to bear in mind when using data collected by others.

Table 8.3 Examples of some digital data sources that can be imported into a GI system. NMO = National Mapping Organizations, USGS = U.S. Geological Survey, NGA = U.S. National Geospatial-Intelligence Agency, NASA = National Aeronautics and Space Administration, DEM = Digital Elevation Model, EPA = U.S. Environmental Protection Agency, WWF = World Wide Fund for Nature, FEMA = Federal Emergency Management Agency, Esri BIS = Esri Business Information Solutions.

Type	Source	Details
	Base Maps	
Geodetic framework	Many NMOs, e.g., USGS and Ordnance Survey	Definition of framework, map projections, and geodetic transformations
General topographic map data	NMOs and military agencies, e.g., NGA	Many types of data at detailed to medium scales
Elevation	NMOs, military agencies, and several commercial providers, e.g., USGS, SPOT Image, NASA	DEMs, contours at local, regional, and global levels
Transportation	National governments and several commercial vendors	Highway/street centerline databases at national levels
Hydrology	NMOs and government agencies	National hydrological databases are available for many countries
Toponymy	NMOs, other government agencies, and commercial providers	Gazetteers of place-names at global and national levels
Satellite images	Commercial, government, and military providers, e.g., EROS Data Center, IRS, NASA, SPOT Image, i-cubed, and DigitalGlobe	See Figure 8.2 for further details
Aerial photographs	Many private and public agencies	Scales vary widely, typically from 1:500–1:20,000
	Environmental	
Wetlands	National agencies, e.g., U.S. National Wetlands Inventory	Government wetlands inventory
Toxic release sites	National environmental protection agencies, e.g., EPA	Details of thousands of toxic sites
World ecoregions	Conservation agencies, e.g., WWF	Habitat types, threatened areas, biological distinctiveness
Flood zones	Many national and regional government agencies, e.g., FEMA	National flood-risk areas
	Socioeconomic	
Population census	National governments, with value added by commercial providers	Typically, every 10 years with annual estimates
Lifestyle classifications	Private agencies (e.g., CACI and Experian)	Derived from population censuses and other socioeconomic data
Geodemographics	Private agencies (e.g., Experian, Esri BIS)	Many types of data at many scales and prices
Land and property ownership	National governments	Street, property, and cadastral data
Administrative areas	National governments	Obtained from maps at scales of 1:5,000–1:750,000

The best way to find geographic data is to search the Internet. Several types of resources and technologies are available to assist searching, and these are described in detail in Section 10.2. They include specialist geographic data catalogs and stores, as well as the sites of specific geographic data vendors. These sites provide access to information about the characteristics and availability of geographic data. Some also have facilities to purchase and download data directly. Probably the most useful resources for locating geographic data are the geolibraries and geoportals (see Section 10.2.2) that have been created as part of national and global spatial data infrastructure initiatives (SDI).

The best way to find geographic data is to search the Internet using one of the specialist geolibraries or SDI geographic data geoportals.

A major challenge of using data obtained from the Web is evaluation of fitness for purpose. Too often inexperienced practitioners download data from the Web and assume that its accuracy and licensing terms are adequate for use in a GI project. It is essential that the suitability of all datasets be checked before they are used. A good starting point is to examine the metadata records associated with the dataset (Section 10.2); these records should indicate age, provenance, projection, and a range of other relevant properties. Simple checks include overlay of the data on top of a base map of known and acceptable accuracy and independent verification (e.g., by fieldwork or by comparison with other datasets) of the geometric and attribute properties of a representative sample of objects. It is also best to assume that all data are proprietary until open access/use is confirmed.

8.4.1 Geographic Data Formats

One of the biggest problems with data obtained from external sources is that they can be encoded in many different formats. There are so many different geographic data formats because no single format is appropriate for all tasks and applications. It is not possible to design a format that supports, for example, both fast rendering in police command and control systems and sophisticated topological analysis in natural resource information systems: the two are mutually incompatible. Also, given the great diversity of geographic information, a single comprehensive format would simply be too large and cumbersome. The many different formats that are in use today have evolved in response to diverse user requirements.

Given the high cost of creating databases, many tools have been developed to move data between systems and to reuse data through open application programming interfaces (APIs). In the former case, the approach has been to develop software that is able to translate data (Figure 8.17), either by a direct read into memory or via an intermediate file format. In the latter case, software developers have created open interfaces to allow access to data.

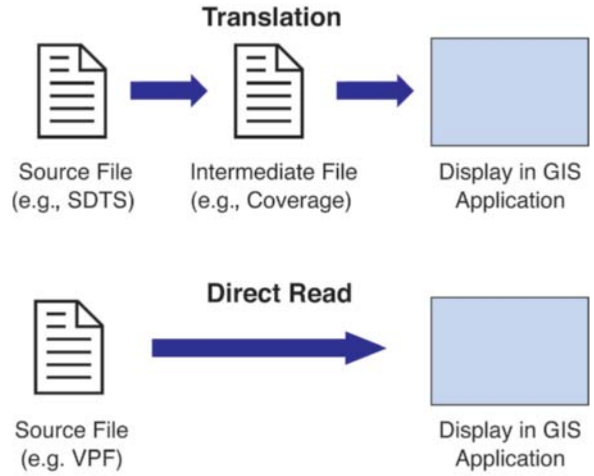


Figure 8.17 Comparison of data access by translation and direct read.

Many GI systems are now able to read directly AutoCAD DWG and DXF, Microstation DGN, Esri Shapefile, VPF, and many image formats. Unfortunately, direct read support can only be easily provided for relatively simple product-oriented formats. Complex formats, such as SDTS (Spatial Data Transfer Standard), were designed for exchange purposes and require more advanced processing before they can be viewed (e.g., multipass read and feature assembly from several parts).

Data can be transferred between systems by direct read into memory or via an intermediate file format.

More than 25 organizations are involved in the standardization of various aspects of geographic data and geoprocessing; several of them are country and domain specific. At the global level, the ISO (International Organization for Standards) is responsible for coordinating efforts through the work of technical committees TC 211 and 287. In Europe, CEN (Comité Européen de Normalisation) is engaged in geographic standardization. At the national level, there are many complementary bodies. One other standards-forming organization of particular note is OGC (Open Geospatial Consortium), a group of vendors, academics, and users interested in the interoperability of geographic systems. To date, there have been promising OGC-coordinated efforts to standardize on simple feature access (simple geometric object types), metadata catalogs, and Web access (see Chapter 10 for further details).

The most efficient way to translate data between systems is usually via a common intermediate file format.

Having obtained a potentially useful source of geographic information, the next task is to import it into a GI database. If the data are already in the native format of the target GI system, or the software has a direct read capability for the format in question, then this is a relatively straightforward task. If the data are not compatible with the target software, then the alternatives are to ask the data supplier to convert the data to a compatible format or to use a third-party translation software system, such as the Feature Manipulation Engine from Safe Software (www.safe.com lists over 300 supported geographic data formats) to convert the data. Geographic data translation software must address both syntactic and semantic translation issues. Syntactic translation involves converting specific digital symbols (letters and numbers) between systems, whereas semantic translation is concerned with converting the meaning inherent in geographic information. Although syntactic translation is relatively simple to encode and decode, semantic translation is much more difficult and has seldom met with much success to date.

Although the task of translating geographic information between systems was described earlier as relatively straightforward, those that have tried this in practice will realize that things on the ground are seldom quite so simple. Any number of things can (and do!) go wrong, ranging from corrupted media to incomplete data files, incompatible versions of translators, and different interpretations of a format specification, to basic user error.

Two basic strategies are used for data translation: one is direct and the other uses a neutral intermediate format. For small systems that involve the translation of a small number of formats, the first is the simplest. Directly translating data back and forth between the internal structures of two systems requires two new translators (A to B, B to A). Adding two further systems will require 12 translators to share data between all systems (A to B, A to C, A to D, B to A, B to C, B to D, C to A, C to B, C to D, D to A, D to B, and D to C). A more efficient way of solving this problem is to use the concept of a data switchyard and a common intermediate file format. Systems now need only to translate to and from the common format. The four systems use only 8 translators instead of 12 (A to Neutral, B to Neutral, C to Neutral, D to Neutral, Neutral to A, Neutral to B, Neutral to C, and Neutral to D). The more systems there are, the more efficient the result. This is one of the key principles underlying the need for common file-interchange formats.

8.5 Capturing Attribute Data

All geographic objects have attributes of one type or another. Although attributes can be collected at the same time as vector geometry, it is usually more cost effective to capture attributes separately. In part, this is because attribute data capture is a relatively simple task that can be undertaken by lower-cost clerical staff. It is also because attributes can be entered by direct data loggers, manual keyboard entry, optical character recognition (OCR), or increasingly, voice recognition—methods that do not require expensive hardware and software systems. By far the most common method is direct keyboard data entry into a spreadsheet or database. For some projects, a custom data-entry form with built-in validation is preferred. On small projects single entry is used, but for larger, more complex projects data are entered twice and then compared as a validation check.

An essential requirement for separate data entry is a common identifier (also called a key, or object-id) that can be used to relate object geometry and attributes following data capture (see Figure 7.10 for a diagrammatic explanation of relating geometry and attributes).

Metadata are a special type of nongeometric data that are increasingly being collected. Some metadata are derived automatically by the GI system (for example, length and area, extent of data layer, and count of features), but some must be explicitly collected (for example, owner name, quality estimate, and original source). Explicitly collected metadata can be entered in the same way as other attributes, as described earlier. For further information about metadata, see Section 10.2.

8.6 Citizen-Centric Web-Based Data Collection

New developments in Web technology have opened up a new vista of opportunities for distributed geographic data collection. A raft of new Web 2.0 technologies has enabled organizations and individual projects to use citizens to collect data very rapidly across a wide variety of thematic and geographic areas that represent a wide spectrum of viewpoints. It is now very simple to create a Web site with a form-based interface to collect geographic data about many types of phenomena, events, and activities. Locations can be obtained by asking the user to digitize the location on a map or to upload coordinates collected in any of the ways outlined earlier in this chapter. This type of approach to data collection by volunteers is discussed in Section 2.5. Box 8.2 presents an example of a Web 2.0, citizen-centric data collection application.

Applications Box 8.2

E-Flora British Columbia, Citizen-Centric Data Collection

The British Columbia, Canada E-Flora BC project is a good example of the way data collection is changing to incorporate citizen-centric data input. E-Flora BC is an online, Web-accessible electronic atlas of plant species that allows interactive data collection, reporting, and mapping (www.eflora.bc.ca).

The public is encouraged to participate in an Invasive Alien Plant Program run by the Project by reporting suspected new occurrences of invasive plants using an interactive form-based Report-a-Weed tool. This tool (Figure 8.18) allows citizen–scientists to collect and enter pertinent information about invasive species, which is then delivered to an appropriate botanist for review and then entered into the organization's master database.

The Web has radically transformed the way that databases of this type are collected and maintained: no longer is this process the sole preserve of the official government organizations. This type of Web 2.0 data collection and public empowerment is especially good at handling local (in space and time) phenomena. There is a rapidly growing list of examples of Web projects that rely on volunteers to collect geographic information (see also Section 10.2 for further discussion).

(A)

(B)

(C)

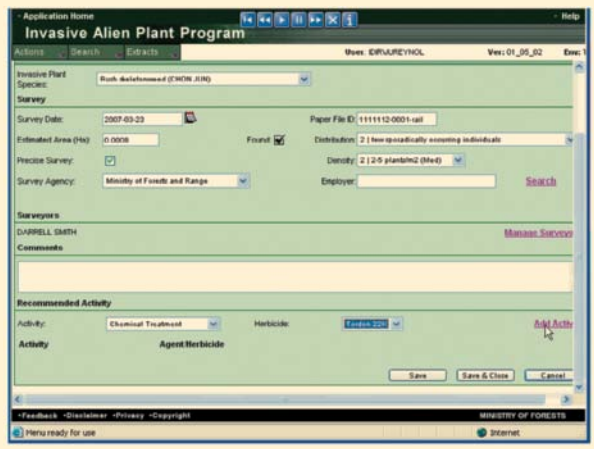


Figure 8.18 E-Flora information about *Senecio jacobaea* (the "stinking willie" or tansy ragwort): (A) photograph; (B) distribution map; and (C) citizen-centric Web data input form.

8.7 Managing a Data Collection Project

The subject of managing a GI project is given extensive treatment in Chapters 16–18. The management of data capture projects is discussed briefly here, both because of its critical importance and because it involves several unique issues. That said, most of the general principles for any GI project apply to data collection: the need for a clearly articulated plan, adequate resources, appropriate funding, and sufficient time.

In any data collection project there is a fundamental trade-off between quality, speed, and price

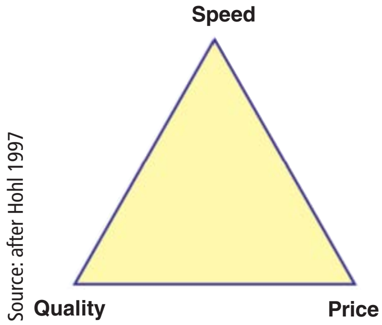


Figure 8.19 Relationship between quality, speed, and price in data collection.

(Figure 8.19). Collecting high-quality data quickly is possible, but it is also very expensive. If price is a key consideration, then lower-quality data can be collected over a longer period.

Data collection projects can be carried out intensively over a short period or less extensively over a longer period. A key decision facing managers of such projects is whether to pursue a strategy of incremental or very rapid collection. Incremental data collection involves breaking the data collection project into small manageable subprojects. This allows data collection to be undertaken with lower annual resource and funding levels (although total project resource requirements may be larger). It is a good approach for inexperienced organizations that are embarking on their first data collection project because they can learn and adapt as the project proceeds. At the same time, these longer-term projects run the risk of employee turnover and burnout, as well as changing data, technology, and organizational priorities.

Whichever approach is preferred, a pilot project carried out on part of the study area and a selection of the data types can prove to be invaluable. A pilot project can identify problems in workflow, database design, personnel, and equipment. A pilot database can also be used to test equipment and to develop procedures for quality assurance. Many projects require a test database to carry out hardware and software acceptance tests, as well as to facilitate software customization. It is essential that project managers are prepared to discard all the data obtained during a pilot data collection project, so that the main phase can proceed unconstrained.

A further important decision is whether data collection should use in-house or external resources. It is now increasingly common to outsource geographic data collection to specialist companies that usually undertake the work in areas of the world with very low labor costs (e.g., India, Thailand, and Vietnam). Three factors influencing this decision are cost/schedule, quality, and long-term ramifications. Specialist external data collection agencies can often perform work faster, cheaper, and with higher quality than with in-house staff, but because of the need for real cash to pay external agencies, this may not be possible. In the short term, project costs, quality, and time are the main considerations, but over time dependency on external groups may become a problem.

Box 8.3 introduces the work of Carmen Reyes, a researcher with the National Autonomous University of Mexico with over 40 years of experience in GI projects and data collection.

Biographical Box 8.3

Carmen Reyes

Carmen Reyes (Figure 8.20) is an expert and a researcher in GI science, cybercartography, geocybernetics, and geomatics; she has a BSc in Mathematics from the National Autonomous University of Mexico (UNAM) and a master's in Mathematics from the Metropolitan Autonomous University (UAM). She obtained her PhD in geographic information systems from Simon Fraser University, in Canada. During the past 40 years she has worked in over 80 geomatics/GI projects for local and international institutions, in both public and private sectors.

Dr. Reyes succeeded in her efforts to introduce geomatics and GI science in the Mexican scientific realm as founder in 1999 and, for a decade, general director of the J. L. Tamayo Center for Research in Geography and Geomatics (CentroGeo) under the (Mexican) National Science and Technology Council. She was presented with the Samuel Gill Gamble Award for Cartography by the Government of Canada through the PanAmerican Institute of Geography and History (PAIGH) of the Organization of American States (OAS). Dr. Reyes is member of national and international organizations, networks, and academic committees. She is in the Board of Directors of the Global Spatial Network (GSN) and member of the global advisory council of the Open Geospatial Consortium (OGC). Currently she is a senior researcher at CentroGeo, teaches graduate courses, is senior supervisor of Master and PhD students, is the chief editor of the *Journal on Geocybernetics*, and manages GI science projects.

On geocybernetics and knowledge-based GI systems, Carmen says:

> The term "geocybernetics" encompasses several avenues of research that explicitly incorporate the science of cybernetics, general systems theory, modeling, and complexity theory as theoretical building blocks. Currently, together

with a research group at CentroGeo I am conducting empirical and theoretical work in geocybernetics that include: (1) cybercartography, (2) complex solutions in geomatics, (3) collective mental maps, (4) the geomatics prototype, (5) the Strabo technique and (6) the Reyes method.

In the tradition of science, experimentation is a key and invaluable resource, for which data collection and the processing and application of information to sustain or reject hypotheses is a common approach. What makes this approach different is its main thesis— the assumption that conversations between scientific and societal actors should be based on cognitive knowledge frameworks. Through the process, new knowledge frameworks emerge out of the fusion of one or more knowledge domains; i.e., a transdisciplinary process evolves through which cognitive bridges are built not only within geomatics/GI science but also with other knowledge domains, such as public policy, landscape ecology, and criminology.

This knowledge-based approach to GI science and geomatics has been very effective for the interaction between science and society and has resulted in novel scientific findings and outcomes. It can be stated that the main driving force in the processes to design and implement solutions from a geomatics/GI science perspective is the K in knowledge rather than the I in information or the D in data.

Courtesy: Carmen Reyes

Figure 8.20 Carmen Reyes, GI Scientist.

Questions for Further Study

1. Evaluate the suitability of free geographic data for your home region or country for use in a GI project of your choice.
2. What are the advantages of batch vectorization over heads-up digitizing?
3. What quality assurance steps would you build into a data collection project designed to construct a database of land parcels for tax assessment?
4. Why do so many geographic data formats exist? Which ones are most suitable for selling vector data?

Further Reading

Hohl, P. (ed.). 1997. *GIS Data Conversion: Strategies, Techniques and Management*. Santa Fe, NM: OnWord Press.

Kerski, J. J. and Clark, J. 2012. *The GIS Guide to Public Domain Data*. Redlands, CA: ESRI Press.

Lillesand, T. M., Kiefer, R. W., and Chipman, R. W. 2008. *Remote Sensing and Image Interpretation* (6th ed.). Hoboken, NJ: Wiley.

Weng, Q. 2012. *An introduction to contemporary remote sensing*. New York: McGraw-Hill.