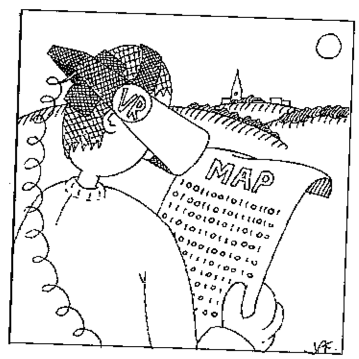
CHAPTER 4

Getting the Map into the Computer

"Roll up that map; it will not be wanted these ten years."

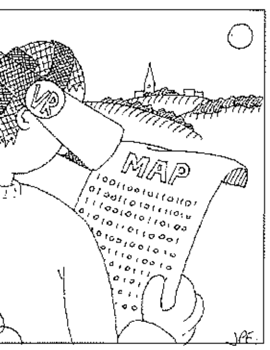
—(William Pitt, 1805)



They say "to err is human but to really foul things up requires a computer."
But who, when lost, has not been heard to say:
"My map shows clearly that a road is here;
the land is wrong, why ask that trooper?"

4.1 ANALOG-TO-DIGITAL MAPS

Most people think of maps as drawings on paper. Maps hang on walls, lie in map drawers, and fill the pages of books, atlases, street guides, newspapers, and magazines. Maps roll off the nation's printing presses in the millions each year, and they fill the spaces in every car's glove compartment, neatly folded or not! The traditional paper maps of our everyday world can be called *real maps*, because they are touchable. We can hold them in our hands, fold them up, and carry them around. The computer, in contrast, has forced us to reconsider this simple definition of a map. In the digital era, and especially within GISs, maps can be both real and *virtual*.



walls, lie in map drawers,
nd magazines. Maps roll
d they fill the spaces in
ional paper maps of our
able. We can hold them
r, in contrast, has forced
ra, and especially within

A *virtual map* is a map waiting to be drawn. It is an arrangement of information inside the computer in such a way that we can use the GIS to generate the map however and whenever we need it. We may have stored map information about roads, rivers, and forests, for example, but may decide that only the forests and rivers need be shown on any map that the GIS produces. Every real map is simply a conversion of the virtual map into a medium, the form that the map will take. In most cases, the medium we use is paper.

Using maps within GISs means that somehow they have already been turned from real into virtual maps. Another way to say this is that a paper map has gone through a conversion, from a paper or analog form into a digital or number form. We start with paper, or sometimes film, Mylar, or some other medium, and we end up with a set of numbers inside files in the computer. This conversion process is called *geocoding*, which we can define as the conversion of spatial information into computer-readable form. Some GIS vendors would be pleased to help you acquire the data you need, but at an immense price. Studies have shown that finding the right maps, and converting these maps from real to virtual form by geocoding, takes up anywhere between 60% and 90% of both the time and money spent on a typical GIS project. Fortunately, this is a once-only cost. As soon as we have the map in digital form, we can use it in a GIS over and over again for different uses and projects unless it needs an update.

Digital map data for use in GIS really falls into two categories. Either the data already exist and all we have to do is find or buy them, or they don't exist and we have to geocode paper maps or maps on some other medium. A third case is that the maps don't even exist, and here we often turn to remote sensing, aerial photography, or field data collection by surveyors or the global positioning system (GPS), to get our first map of a new location. Also, sometimes the maps we need already exist, but whoever geocoded them is not interested in sharing the data with you, even for a price! Even when we can get the maps we need in digital form, they may not suit our particular type of GIS, or may be out of date or not show the features we want. The bottom line is that sooner or later, and usually sooner, we end up geocoding at least some of our own maps.

Before we cover the ways that maps can be converted into numbers, scanning and digitizing, we will take a look at how we might go about finding digital map data that already exist. If we are successful, with a little effort, some conversion programs, and a knowledge of GIS data formats, we can reuse one of the many excellent maps already available for us. Many of these maps can be read straight into a GIS, sometimes without any need to research the way the files and numbers are structured. In this chapter we take a guided tour of the various flavors of data, their formats, and the way the information in the maps has been structured during geocoding.

These days, very few GIS projects have to start with no data at all. The vast amount of data collected and made available by the various branches of government is an excellent base on which to start building. The trick is knowing where to look, what to do when you find what you want, and how to get the data into your GIS.

4.2 FINDING EXISTING MAP DATA

The search for paper maps is often started in a map library. The libraries most likely to carry maps and to support cartographic research are the research libraries in the largest cities and those attached to major universities. Map librarians make use of computer networks to share information and conduct searches, and they are increasingly making census and other digital maps available both in libraries and via computer networks.

Another place to look for map information is in books. An excellent starting point is the *Map Catalog* by John Makower (1986). *Maps for America*, by M. M. Thompson (1987) of the USGS, is a good survey of existing published maps for the United States. Another information source, especially internationally, is the *Inventory of World Topographic Mapping* (Bohme, 1993). The appendices in John Campbell's book *Map Use and Analysis* (Campbell, 1993) show how to use map series and their indices, and many other information sources are listed, especially Chapter 21, "U.S. and Canadian Map Producers and Information Sources."

In many cases, state and local governments keep collections of paper maps. A local planning or building permit office can often find maps of your property or of parks and business properties. Make sure to call ahead. How good the service of providing maps to the public is depends a great deal on the office and its policies and services. Some larger agencies have their own map division. A state highway authority, park service, or industrial development organization may have its own maps available, sometimes free or at little cost.

Commercial companies sell cartographic data and some will conduct map data searches. Imagery from most commercial vendors can be searched and browsed using an online database. Many commercial services offer not only packaged existing data for your use, but will digitize or scan data and even write the data in GIS format for you at a cost. Two companies offering such services are ETAK and Geographic Data Technology.

Obviously, each company has its own strengths and types of map for sale. Commercial companies are not, however, for the novice. They are primarily used by large corporations, governments, the real estate industry, and so on. For a first cut, the usually free public data are the best starting point, and in many cases enough, even many times more, than you will ever need to work with your GIS.

Digital map data by public agencies have been dominated by data from the federal government. In the United States, digital map data created at the federal level for its own use are the property of the American people, with the obvious exception of sensitive data of use in national security—although recently even spy satellite data have been made available. The Freedom of Information Act guarantees every American the right to get copies of digital map data used by the federal government, subject to a distribution or copying cost that may not exceed a reasonable marginal cost of providing the data.

Not all data has to be extracted from the government using the act, however. Government agencies have made it their mission to make map data as freely accessible as possible to any interested party. Computer networks have made this not only accessible to almost any computer user but have also made it more flexible.

4.2.1 Finding Data on the Networks

An excellent way to begin a data search is to use a computer network. Several computer packages allow you to do this over the various network access methods, such as America

Online and CompuServe. The most sophisticated tools, however, are those available on the Internet. Among the various tools, such as Archie, Veronica, WAIS, and Gopher, is a computer program called Mosaic, from the National Supercomputing Center at the University of Illinois. Mosaic allows you to search the World Wide Web (WWW), an interlinked set of computers and servers, or data repositories on the Internet. Similar and more widely used programs are Netscape Navigator and Microsoft Explorer. Each major agency has a World Wide Web server, or *gateway*, through which data can be searched and downloaded. Simply enormous amounts of data are available through this simple mechanism.

While many U.S. government agencies create and distribute digital maps, data from three agencies, each one with its own different types, have been most used in GISs. The agencies are the U.S. Geological Survey (USGS), part of the Department of the Interior; the U.S. Bureau of the Census; and the National Oceanic and Atmospheric Administration (NOAA), both part of the Department of Commerce. Data they supply cover the land and its features, the population, and the weather, atmosphere, and oceans across the United States.

Each of these agencies is worth covering here, although there are many others. Finding information in any of them has been made much easier by several public information service and computer network services, especially over the Internet. The Internet is a network of computer networks and is accessible to all users through a computer that is attached to the system. More information can be found on the World Wide Web site for this book, given in Chapter 1.

4.2.2 U.S. Geological Survey

Digital cartographic data from the USGS are distributed by the Earth Science Information Centers as part of the National Mapping Program. Information is available by calling 1-888-ASK-USGS in the United States and Canada or by writing to the addresses listed in the study guide for this chapter. The USGS digital data fall into six categories: *digital line graphs* (DLGs), *digital elevation models* (DEMs), *land-use and land-cover digital data*, *digital cartographic text* (Geographic Names Information System, GNIS), *digital orthophotoquads* (DOQ), and *digital raster graphics* (DRG). The USGS Web portal is shown in Figure 4.1.

The USGS continues to improve coverage of the United States and distributes the map data products on computer tape, floppy disk, and CD-ROM. These data formats are covered in this chapter, and many GIS packages support them directly. Arc/Info, for example, will recognize a digital line graph file and read the data accordingly.

Many of the USGS data sets are now available directly through the Internet by the file transfer protocol (FTP), a tool for moving files over computer networks. Data sets are maintained on USGS servers, usually listed alphabetically by the name of the map quadrangle concerned. This means that to retrieve the data set, you should both know what data are needed and what format the data are to be found in.

The USGS also now distributes data on land cover derived from classifications of NOAA's AVHRR (*advanced very high resolution radiometer*) measurements. These data are distributed by the EROS Data Center on CD-ROM and on the Internet, with a ground resolution of 1 kilometer. Biweekly composites showing a vegetation index for North

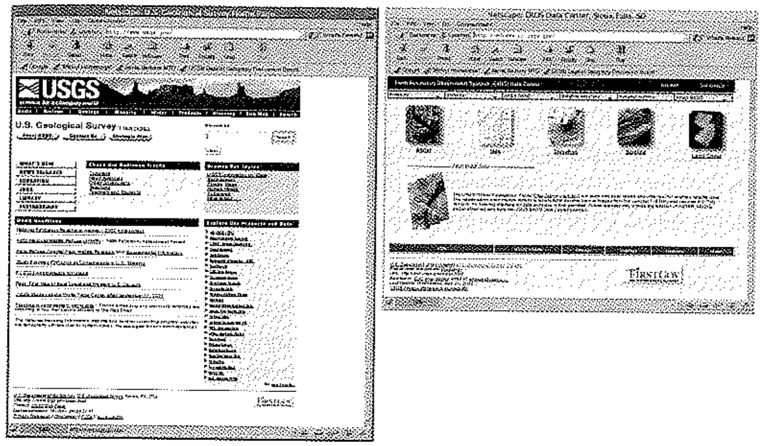


FIGURE 4.1: World Wide Web gateway for the U.S. Geological Survey. Left: The main home page `http://www.usgs.gov`. Right: The EROS data center home page at `http://edcwww.cr.usgs.gov`. (Used with permission of the U.S. Geological Survey.)

America are available, and efforts are under way to release a global AVHRR data set at this resolution.

An important global data set is the *Digital Chart of the World* (DCW). This digital map contains coverage of the entire world, including major cities, rivers, lakes, coastlines, contours, vegetation, and transportation routes. The DCW is a digital version of the Defense Mapping Agency's operational navigation chart (ONC) and jet navigation chart (JNC) at a scale of 1 : 1,000,000. The data are stored and distributed on four CD-ROMs, which include software for viewing the data files on IBM-PC compatible microcomputers. Distribution of the data is via the USGS's Earth Science Information Centers.

4.2.3 National Oceanic and Atmospheric Administration

The NOAA concentrates on marine and aeronautical navigation systems that electronically integrate digital charts, global positioning system-based locations, and real-time environmental information. Examples are the daily weather map, satellite and radar images, and maps used by pilots and air traffic control. The NOAA charts must be carried aboard all large ships in U.S. waters. The National Geophysical Data Center, part of NOAA, has released numerous digital map data sets on CD-ROM, most recently including detailed bathymetry of the ocean and land-surface topography as well as geodetic and magnetic data for the earth's surface (Figure 4.2).

4.2.4 U.S. Bureau of the Census

The mapping of the U.S. Census Bureau is to support the decennial census by generating street-level address maps for use by the thousands of census enumerators. For the 1990 census, the Census Bureau developed a system called TIGER (*topologically integrated geographic encoding and referencing*). The TIGER system (Figure 4.3) uses the block face or street segment as a geographic building block and recognizes cartographic objects of different dimensions, points (nodes), lines (segments), and areas (blocks, census tracts,

y. Left: The main home page
/edcwww.cr.usgs.gov. (Used

global AVHRR data set at

orld (DCW). This digital
s, rivers, lakes, coastlines,
a digital version of the
) and jet navigation chart
buted on four CD-ROMs,
mpatible microcomputers.
mation Centers.

on

on systems that electron-
l locations, and real-time
map, satellite and radar
e NOAA charts must be
Geophysical Data Center,
CD-ROM, most recently
pography as well as geo-

nial census by generating
numerators. For the 1990
(*topologically integrated*
igure 4.3) uses the block
nizes cartographic objects
eas (blocks, census tracts,

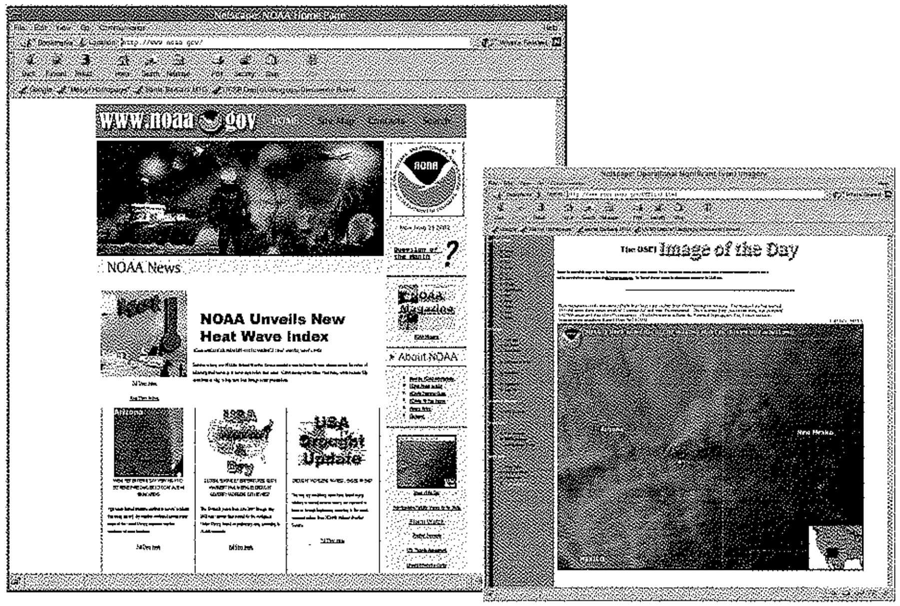


FIGURE 4.2: World Wide Web gateway for NOAA, and a sample satellite image map extracted over the Internet at `http://www.noaa.gov`. (Used with permission.)

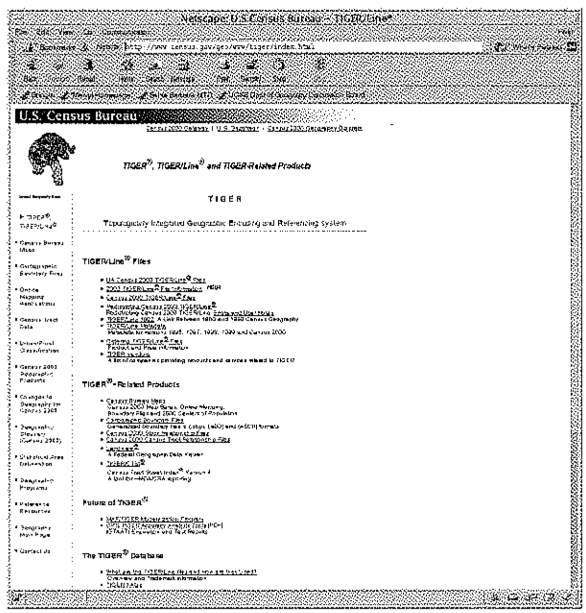


FIGURE 4.3: World Wide Web gateway for the census TIGER data at the U.S. Census Bureau. The TIGER logo is a copyright of the U.S. Census Bureau. (Used with Permission.)

or enumeration districts). In TIGER terminology, points are *zero cells*, lines are *one cells*, and areas are *two cells*.

A large-scale cooperative effort prepared these maps for the 1990 census. Map digitizing was performed in collaboration with the U.S. Geological Survey (see "People in GIS" at the end of this chapter). The maps are distributed along with the census data on computer tape, separately on CD-ROM, and over the Internet. Virtually every GIS allows TIGER files to be imported directly into the system, although not all GISs handle the attribute data as well. The TIGER was the first comprehensive GIS database at street level for the entire United States. An important ability of TIGER is to do *address matching*: the search for street addresses through the attribute files to match a block or census tract in the TIGER graphic files; that is, finding its geographic location on the map solely from a street address listing.

4.2.5 Creating New Data

Wonderful as it is to find an existing digital map, the myriad of different data formats is usually the least of the GIS analyst's problems. Digital maps, like their analog sources, are specific to a given map scale. Boundary lines, coastlines, and so on all reflect the degree of generalization applied to the lines when the map was originally digitized. In addition, maps were usually digitized with different levels of precision, from source maps that were out of date, or that have become out of date since they were captured by computer, or sometimes even have errors or problems with their accuracy. Two different maps of the same area rarely agree over every detail, yet the computer is unable to resolve the differences in the same way that the human mind can reason about the reliability of information, its timeliness, and so on.

In summary, like it or not, sooner or later if you are involved with GIS you will find yourself digitizing a map. Although this is a tedious, time-consuming, and potentially frustrating exercise, the learning process involved will greatly increase your awareness of the limitations of digital maps for GIS use. It is far better to persevere and learn, than to make a million errors and misjudgments for the lack of a little hands-on experience. Time, then, to get a little digital (or at least virtual) "mud" on our boots!

4.3 DIGITIZING AND SCANNING

Historically, many different means have been used to geocode. At first, some very early GIS packages required maps to be encoded and entered by hand. The hours of monotonous work required for this task made errors common and their correction difficult. Since special-purpose digitizing hardware became available, and especially since the cost of this hardware fell substantially, virtually all geocoding has been performed by computer.

Two technologies have evolved to get maps into the computer. Digitizing mimics the way maps were drafted by hand and involves tracing the map over using a cursor while it is taped down onto a sensitized digitizing tablet. The second method involves having the computer "sense" the map by scanning it. Both approaches work and have their advantages and disadvantages. Most important, the method of geocoding stamps its form onto the data in such a way that many other GIS operations are affected afterwards.

o *cells*, lines are *one cells*,

or the 1990 census. Map
gical Survey (see "People
ed along with the census
Internet. Virtually every
m, although not all GISs
nprehensive GIS database
of TIGER is to do *address*
files to match a block or
eographic location on the

f different data formats is
like their analog sources,
and so on all reflect the
as originally digitized. In
f precision, from source
ce they were captured by
r accuracy. Two different
nputer is unable to resolve
on about the reliability of

ed with GIS you will find
nsuming, and potentially
increase your awareness
persevere and learn, than
ttle hands-on experience.
our boots!

ode. At first, some very
d by hand. The hours of
and their correction dif-
able, and especially since
ding has been performed

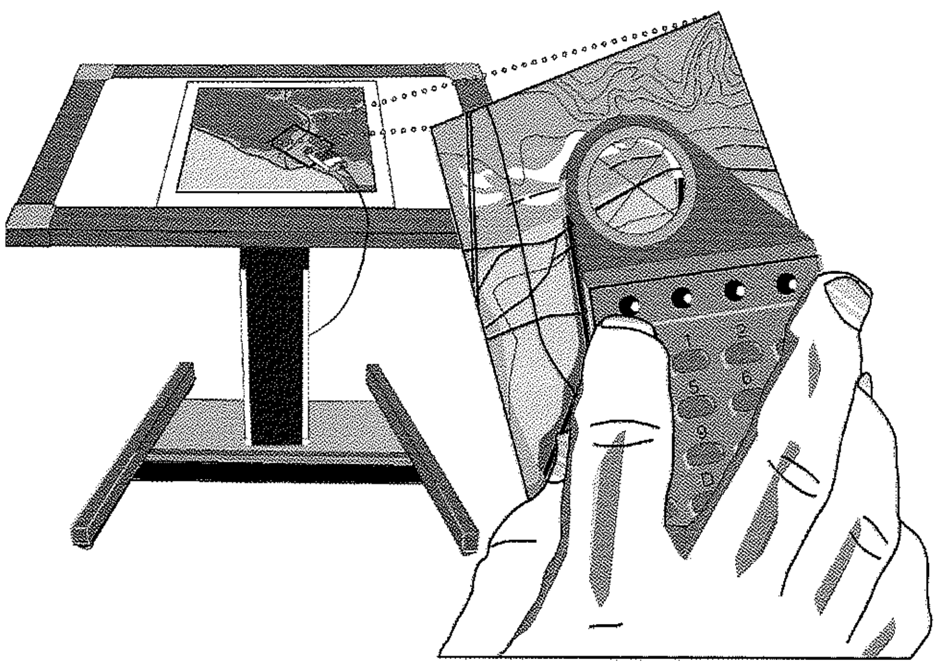
mputer. Digitizing mim-
g the map over using a
blet. The second method
t. Both approaches work
the method of geocoding
IS operations are affected

4.3.1 Digitizing

Geocoding by tracing over a map with a cursor is sometimes called *semiautomated digitizing*. This is because in addition to using a mechanical device, it involves a human operator. Digitizing means the use of a digitizer or digitizing tablet (Figure 4.4). This technology has developed as computer mapping and computer-aided design have grown and placed new demands on computer hardware.

The digitizing tablet is a digital and electronic equivalent of the drafting table. The major components are a flat surface, to which a map is usually taped, and a stylus or cursor, with the capability of signaling to a computer that a point has been selected (Figure 4.4). The mechanism to capture the location of the point can differ. Many systems have connected arms, but most have embedded active wires in the tablet surface that receive an electrical impulse sent by a coil in the cursor. In some rare cases, the cursor transmits a sound, which is picked up and recorded by an array of microphones.

The actual process of digitizing a map proceeds as follows (Figure 4.5). First, the paper map is tailored or preprocessed. If the map is multiple sheets, the separate

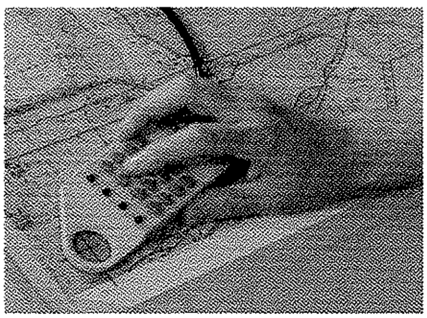


1. Digitizer cursor transmits a pulse from an electromagnetic coil under the view lens.
2. Pulse is picked up by nearest grid wires under tablet surface.
3. Result is sent to computer after conversion to x and y units.

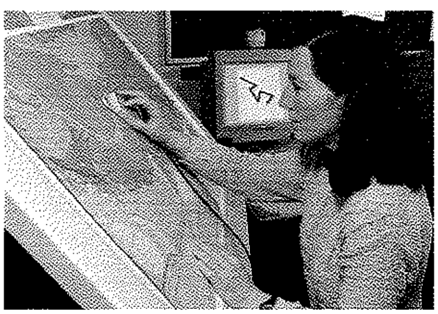
FIGURE 4.4: Digitizing tablet system of operation.

Stable base copy of map is prepared by choosing digitizing control points at known locations. Any features to be selected should be marked in advance.

Map is firmly taped or fixed down to tablet. No movement of the map should be possible. Surface should be flat and free of folds, bubbles, and so forth. Double tape over edges that will be rubbed by elbows and forearms.



Control points to be used in registering the map to the tablet are entered one at a time, along with their map coordinates in geocoded coordinate space such as latitude and longitude or UTM meters.



Digitizing begins. Map features are traced out using the cursor. Care is taken to capture features accurately, with a suitable level of detail. Points can be selected one at a time or in a stream turned on and off from the cursor. Attributes can be entered as features are completed.

FIGURE 4.5: The digitizing process.

sheets should be digitized independently and digitally merged (zipped) later. Unless annotations have to be made onto the map to assist the geocoding, the next major step is to derive a coordinate system for the map. Most applications use Universal Transverse Mercator coordinates (UTM), or latitude and longitude (see Chapter 2), but many people use hardware coordinates or map inches or millimeters. Map units are sometimes used when precise matching between the digitized map and its source is required. The map is then transformed into geographic coordinates when both the editing and the proofing are complete.

As a minimum, the coordinate locations of three points are required, usually the upper right easting and northing, the lower left easting and northing, and at least one other corner. From these points, with their map coordinates and their raw digitizer coordinates, all the parameters can be computed for converting the data into the map's coordinates. Many GIS map entry and digitizing software packages require four of these control points for computing the map geometry, and it is advisable to repeatedly digitize control points and to average the coordinates to achieve higher accuracy.

The beginning of the digitizing sequence involves selecting the control points and interactively entering their world coordinates. This is a very important step, because an error at this stage would lead to an error in every pair of coordinates. After the map is taped to the tablet, it should not be moved without entering the control points again, and it is preferable to perform this step only once per map. Ideally, the entire digitizing process should be finished at one sitting.

Tape should be placed at each map corner after smoothing the map, and care should be taken to deal with folds and the crinkles that develop with certain papers during periods of high humidity. A stable base product such as Mylar or film is preferable for digitizing. The lower edge, which will have the cursor and your right sleeve (if you are right-handed) dragged over it many times, should be taped over its entire length. You should always permanently record the x and y values of the map control points, ideally digitally and with the geocoded data set. This may allow later recovery of lost resolution or systematic errors.

Digitizing then proceeds with the selection of points. The cursor may have multiple buttons and may be capable of entering text and data without using the keyboard. Voice data entry and commands are also sometimes used. On specialized workstations, there may even be a second tablet with its own mouse or cursor for commands. Errors can be reduced during this process by reading the documentation in advance and by occasionally stopping to review the actual data being generated on the screen.

Points are usually entered one at a time, with a pause after each to enter attributes such as labels or elevations. Lines are entered as strings of points and must be terminated with an end-of-chain signal to determine which point forms the node at the end of the chain. This signal must come from the cursor in some way, either by digitizing a point on a preset menu area or by hitting a preset key.

Areas such as lakes or states are usually digitized as lines. Sometimes an automatic closure for the last point (snapping) can be performed. Finally, the points should be checked and edited. The digitizing software or GIS may contain editing features, such as delete and add a line or move and snap a point. The software may also support multiple collection modes. *Point mode* simply digitizes one point each time the button on the cursor is pressed. *Stream mode* generates points automatically as the cursor is moved, either one point per unit of time or distance. This mode can easily generate very large data volumes and should be avoided in most cases. Error correction is especially difficult in this mode. *Point select mode* allows switching between point mode and stream mode. This mode is sometimes used when lines are both geometric and natural, such as when following a straight road and then a river.

At this point the data are ready for direct integration into the GIS. Usually, a separate module of the GIS is used for digitizing and editing, and the map can now be passed on for use. Digitizing should be approached with caution and a desire for accuracy. Errors in digitizing can usually be eliminated using some very simple procedures and rules.

4.3.2 Scanning

The second digitizing process is *automated digitizing* or more usually, just *scanning*. The scanner you may have seen at a computer store or in an advertisement, or perhaps the one you use for scanning documents, is a *desktop scanner*. The *drum scanner* is most commonly used for maps. This type of scanner receives an entire sheet map, usually clamped to a rotating drum, and scans the map with very fine increments of distance, measuring the amount of light reflected by the map when it is illuminated, with either a spot light source or a laser (Figure 4.6). The finer the resolution, the higher the cost and the larger the data sets. A major difference with this type of digitizing is that lines, features, text, and so on, are scanned at their actual width and must be preprocessed for the computer to recognize specific cartographic objects. Some plotters can double as scanners, and vice versa.

For scanning, maps should be clean and free of folds and marks. Usually, the scanned maps are not the paper products but the film negatives, Mylar separations, or the scribed materials that were used in the map production. An alternative scanner is

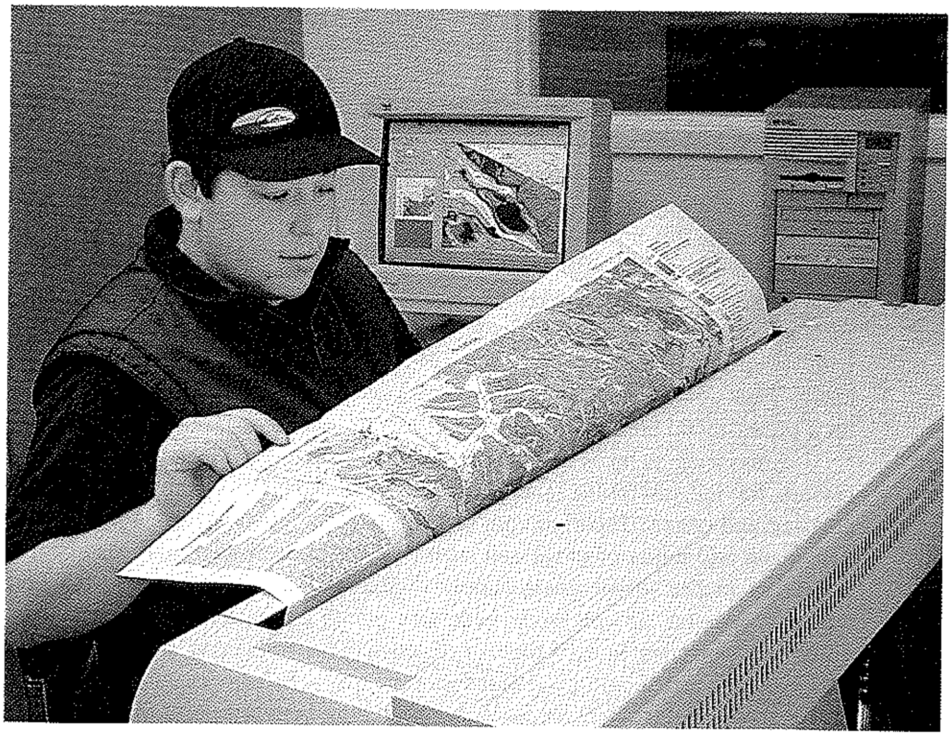


FIGURE 4.6: Digitizing by the map scanning process. (Photo by Susan Baumgart/Matt Eimers.)

the GIS. Usually, a sepa-
he map can now be passed
desire for accuracy. Errors
e procedures and rules.

usually, just *scanning*. The
ertisement, or perhaps the
The *drum scanner* is most
entire sheet map, usually
ne increments of distance,
is illuminated, with either
lution, the higher the cost
of digitizing is that lines,
and must be preprocessed
ome plotters can double as

s and marks. Usually, the
ves, Mylar separations, or
An alternative scanner is

san Baumgart/Matt Eimers.)

Left: Map is cut or folded to frame the section which is to be scanned. Center: Map is placed in scanner, aligned to edge as upright. Top is closed to clamp map and prevent movement. Right: Software is used to frame area to be scanned, to set resolution and mode (color, monochrome, etc.). When all is set, map is scanned, and saved on hard disk. Scan is checked against original.

FIGURE 4.7: Scanning a map with a desktop scanner. (Images courtesy of David Lawson and Jeannette Candau.)

the *automatic line follower*, a scanner that is manually moved to a line and then left to follow the line automatically. Automatic line followers are used primarily for continuous lines, such as contours. These and other scanners are very useful in CADD (computer-aided drafting and design) systems, where input from engineering drawings and sketches is common.

Simple desktop scanners are becoming important geocoding devices as their resolutions improve and their prices fall. The process of scanning usually begins with preparing the section of map, which obviously needs to be as clean and with as solid and crisp lines as possible. Next, the map is placed on the desktop scanner. The software is told which window to scan, the scan is previewed, and the scan is then saved to the resultant scan file (Figure 4.7). The process can be very quick; nevertheless, care and attention can save considerable work later on. Desktop or low-resolution scanning is rarely adequate for GIS purposes but can be used to put a rough sketch into a graphic editing system for reworking. In this way, a field sketch can be used as the primary source of information for developing the final map for the GIS.

It is important to have a clear concept of *scale* and *resolution* when scanning. In Figure 4.8, the same map, part of the Goleta, California, 7.5-minute USGS quadrangle map, was scanned at two different resolutions. The two larger squares outlined on the map are 100 millimeters on a side. At 1 : 24,000, this distance is $24{,}000 \times 100 = 2{,}400{,}000$ millimeters, or 2400 meters.

Although Figure 4.8 shows a scanned 100-mm-by-100-mm segment of the map, the pixel density of the scan is given per inch. A scan at 400 dots per inch (DPI) translates to 15.748 pixels per millimeter, making the scanned map square about 1575 by 1575 pixels. The same area was also scanned at 100 DPI, or 3.937 pixels per millimeter, for an image

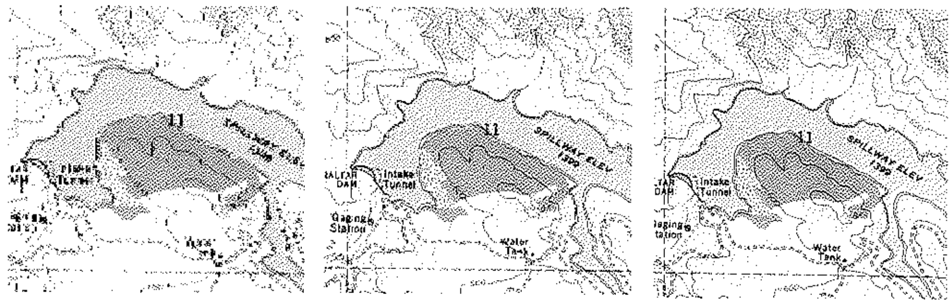


FIGURE 4.8: Drum scanner images of two sections of the Little Pine Mountain, California topographic quadrangle at 1 : 24,000. Left at 50 dpi, center at 100 dpi, and right at 200 dpi.

that is 394 by 394 pixels. On the two scans at their ground equivalent, one pixel is 1.524 meters on the first but 6.096 meters on the second. It is not this print density but the equivalent scale that is important for the map's accuracy. The width of a very thin line on the map, such as a small stream or a contour, is about 0.1 millimeter.

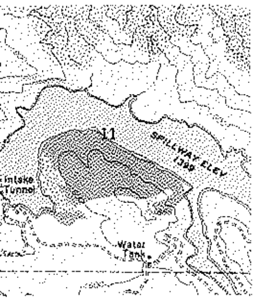
At 1 : 24,000, this means that the contour line would be 2.4 meters wide if it were painted on the ground, more than the pixel size on the 400 DPI scan but less than a single pixel of 6.1 meters on the 100 DPI scan. Most of the line would be skipped, and only occasionally would the pixel and the line coincide. This can be shown clearly in the insets in Figure 4.8. Losing features in this way is called *dropout*. Dropout can virtually eliminate a feature on the map, or at best make it seem like background "noise." Another fact to note about scanning is that folds in the map, pencil lines, coffee stains, paper discoloration, and, in particular, wrinkles all show up. This can lead to problems, as we will see in Section 4.6.

4.4 FIELD AND IMAGE DATA

4.4.1 Field Data Collection

An increasing amount of data for GIS projects comes from a combination of field data, global positioning system data, and imagery. Field data are collected using standard surveying methods, in which locations are established in the field as *control points* and then additional locations, tracing out features or covering terrain, for example, are traced out by large numbers of measurements using instruments designed to measure angles and distances. The highest accuracy instruments, called *total stations* are digital recorders as well as measurement instruments and use laser ranging to prism reflectors to calculate distance.

Less expensive instruments such as theodolites, engineers' transits, and levels often use a technique for measuring distances called *stadia*, which involves reading the numbers on a calibrated pole through the lens of the instrument. Data are recorded in notebooks, and usually the data are then entered into a computer program to turn the bearings, angles, and distances into eastings, northings, and elevations. The type of software used is called COGO, for "coordinate geometry," and many COGO packages either write data directly into GIS format or are capable of writing files that can transfer directly. Figure 4.9 shows a typical field mapping project, the mapping of a lake in upstate New York.



ıntain, California topographic

equivalent, one pixel is
ıot this print density but
The width of a very thin
).1 millimeter.
4 meters wide if it were
PI scan but less than a
e would be skipped, and
ı be shown clearly in the
ıt. Dropout can virtually
ground "noise." Another
ıes, coffee stains, paper
lead to problems, as we

ombination of field data,
ected using standard sur-
s *control points* and then
example, are traced out
l to measure angles and
s are digital recorders as
m reflectors to calculate

transits, and levels often
lves reading the numbers
e recorded in notebooks,
turn the bearings, angles,
of software used is called
either write data directly
irectly. Figure 4.9 shows
e New York.

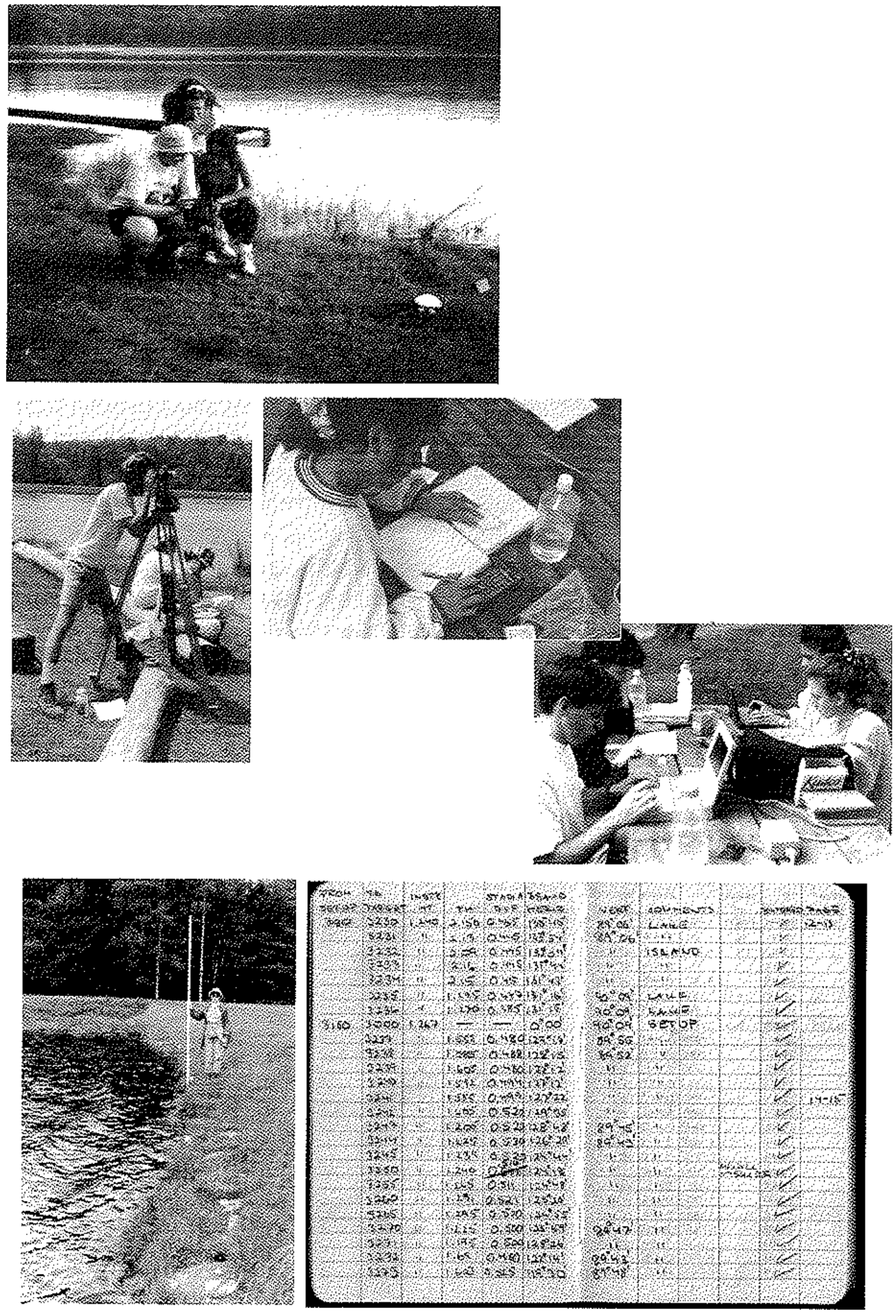


FIGURE 4.9: Getting field data into the GIS. Top left: Collecting base control data with a GPS receiver. Center left: Surveying detailed lake edge information using a theodolite to measure angles and (bottom left) stadia to measure distance. Bottom right: Field note-book data from the surveying. Center middle: Transcribing and averaging the field notes. Center Right: Entering the data into a surveying COGO package to translate distances and angles into coordinates. (Bowman Lake State Park, NY: Hunter College GTECH 350 students.)

4.4.2 GPS Data Collection

The first stage in the surveying process, that of setting the control, is usually done by locating a USGS control point (a bench mark) or by using GPS. Using two GPS receivers together in *differential mode* it is possible to locate control points to sub-meter accuracy. These points are then used as the basis for continued extension of the survey network going outward and between these points. A GPS unit being used for control surveying at extremely high precision is shown in Figure 4.10.

The GPS is a system of 24 orbiting satellites in medium earth orbits (about 20,000 km), each transmitting a time signal. At any given time, at least four of the satellites are above the local horizon at every location on earth 24 hours a day. When a GPS receiver is activated, the nearest satellites are located and the signals are received from each visible satellite. By decoding the time differences between the signals from each satellite, combined with data from the satellite itself about its orbit (called *ephemeris data*) it is possible to solve the three unknowns of latitude, longitude, and elevation. Many receivers can do direct conversion into any of several coordinate systems and datums, and most can download the data directly to a computer. Some GPS equipment can download directly in common GIS formats.

Prior to May 2000, the GPS signal was accurate in its coarse acquisition (C/A) code mode to only about 75 to 100 meters because the signal was deliberately degraded under a system called *selective availability*. The use of selective availability has now been determined to be no longer of national security interest and has consequently been turned off, so that accuracies of 10 to 25 meters are normally possible.

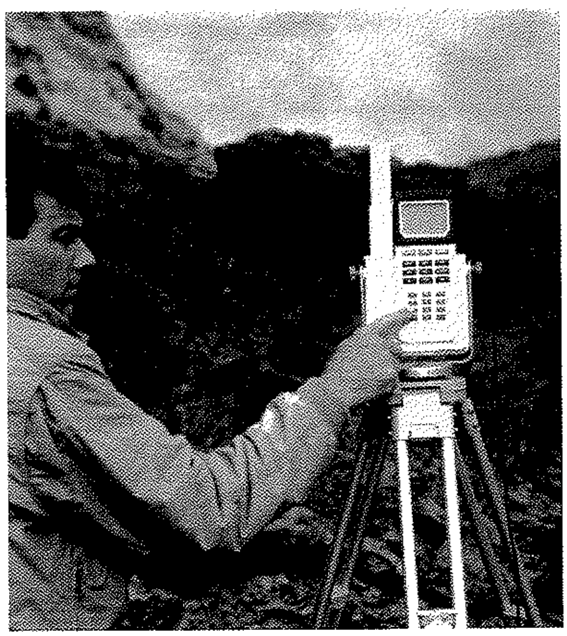


FIGURE 4.10: GPS control being established using a high-precision differential GPS receiver. (Photo courtesy of Magellan Systems Corporation, San Dimas, CA. Used with permission.)

By using two units, one at a known location and one "rover" unit, the degradation can be measured and eliminated, usually by processing the data from the two units on a computer after collecting the data. This is called *differential-mode use* of the GPS. It is possible by using a radio receiver or a cellular telephone link to receive real-time differential corrections in the field. The corrections are broadcast as an aid to navigation in the United States and elsewhere, and are also available from private services.

Many hand-held GPS receivers are now capable of downloading their data to computer software, either for post-processing for accuracy enhancement or for direct integration into GIS. In some cases, GPS receivers have elaborate map displays integrated into the portable units. In others, GPS units work in conjunction with software (including GIS) to display GPS locations directly onto maps. Several GPS vendors now offer software for portable digital assistants, and even cellular telephones, which will soon come equipped with their own GPS receivers inside.

4.4.3 Image and Remote Sensing Data

Imagery data are very common input layers to a GIS. They are most frequently air photos such as the USGS's orthophotos or satellite images. Air photos are usually black and white and are produced from scanning or accessed from CD-ROM. The National Airphoto Program makes photography at a variety of scales available in the United States, and private vendors also sell images. Digital orthophotos are at an equivalent scale of 1 : 12,000 and have a 1-meter ground resolution. The current program calls for national coverage soon after the turn of the century and a 5-year update after that. Examples of a digital orthophoto and satellite data from the Landsat program are shown in Figure 4.11.

The Landsat program has been generating imagery of many locations in the world since 1972. Two scanners, the multispectral scanner and the thematic mapper, image areas on the ground at 79 and 30 meters, respectively. The images are geometrically corrected into the space oblique Mercator projection and are available for use in GIS projects. Coverage can be quite discontinuous due to gaps in the program. Other satellites also generate imagery, including the French SPOT satellite series and the Canadian RADARSAT. In addition, GIS projects at small scales also often use the NOAA polar orbiting satellites carrying the AVHRR (advanced very high resolution radiometer) and the geostationary GOES weather satellite, the one seen every evening on the television weather report.

4.5 DATA ENTRY

Geocoding is the part of GIS data input that results in getting a map into the computer. It is not the entire story, however, for as yet we have not dealt with getting the attributes into the GIS. An attribute is a value, usually a number, containing information about the features contained in the GIS. If the feature we are geocoding is a road, for example, then capturing the route of the road from a map as it winds from intersection to intersection is pure geocoding. We also have to tell the computer what this long and winding line is: a road, and anything else that the GIS needs to know about it. Relevant attributes for a road might be its state route number, the year it was built, what the surface is made of, how many traffic lanes are on the road, if the road is one-way or two-way, how many bridges it goes over, how many cars travel along the road per hour, and so on. These

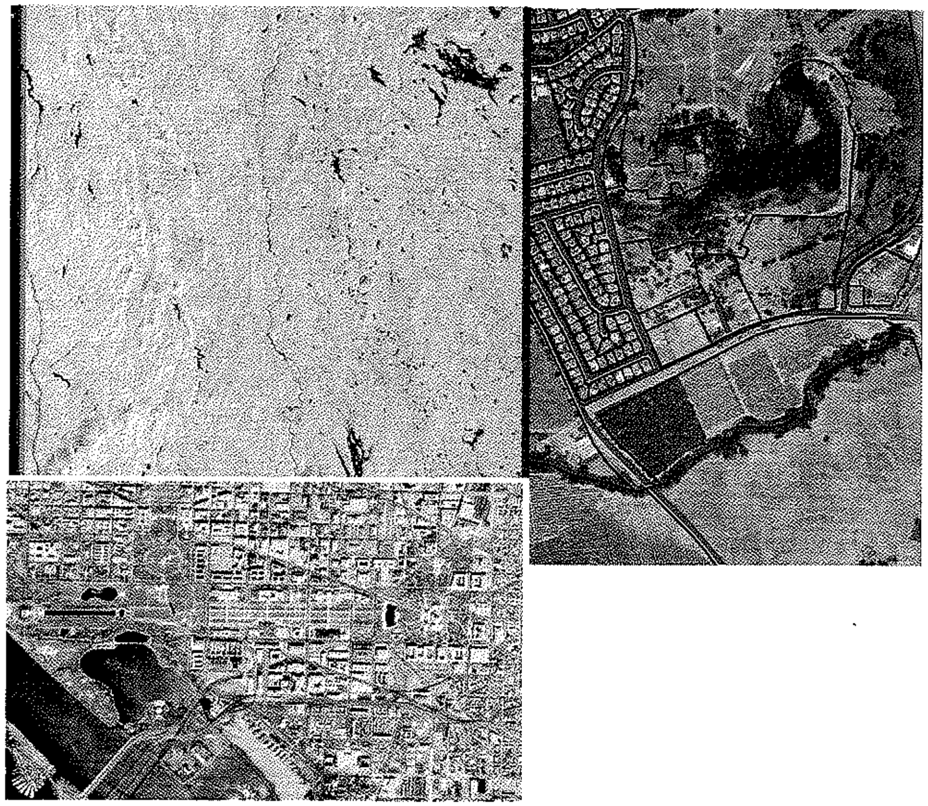


FIGURE 4.11: Imagery often used in GIS. Lower left: Digital orthophoto of Washington, DC. Ground resolution of 1 meter. Upper left: Landsat 7 thematic mapper image of New England. Ground resolution of 30 meters. Upper right: High resolution air photography matched with vector data. Sources: USGS EROS Data Center and county of Santa Barbara.

values are the road's *attributes*. They are the very meat and potatoes of GIS analysis. Somehow, we have to get them into the computer, too.

The simplest way to think of attributes is in a flat file. A flat file is really like a table of numbers. The columns of the table are the attributes, and the rows of the table are the features themselves. Each line in the table is a record, but the name used depends on who you talk to. A computer scientist would call a row a *tuple*, a statistician would call it a *case* or an *observation*. A programmer might call it an *instance of a geographic object*. They are all pretty much the same. *Record* sounds simpler.

Take a look at the flat file in Figure 4.12. The records and attributes relate to the example we discussed above, a road. The attribute table then consists of several parts. First, it has attributes with their names. Setting up the attributes means deciding what values are going to be associated with each of our features. At the time of setup, it is easy to anticipate something we may want to collect in the future and to leave a column in the table for it.

Second, there are records. A record usually has a value in every one of the columns. Software programs such as spreadsheets and some databases allow you to click into a

oto of Washington, DC. Ground
v England. Ground resolution of
data. Sources: USGS EROS Data

l potatoes of GIS analysis.

. A flat file is really like a
, and the rows of the table
but the name used depends
a *tuple*, a statistician would
an *instance of a geographic*
mpler.

and attributes relate to the
en consists of several parts.
butes means deciding what
. At the time of setup, it is
uture and to leave a column

in every one of the columns.
es allow you to click into a

Features on Adams, NY Map

ID #	Feature	Name	Surface	Lanes	Traffic per hour
1	Road	US 11	tarmac	3	113
2	Road	I 81	concrete	4	432
3	Road	Lisk Bridge Road	tarmac	2	12

***value*, is the number or text associated with a record for an attribute**

***attribute* has a name and a value for each record**

***record*, all attributes for one feature**

FIGURE 4.12: An attribute table organized as a flat file.

cell in the table and put in a value. Nevertheless, setting up the table has to be a little more formal than that.

Each attribute has more than simply a name associated with it. For example, if we tried to put "US11" into the attribute column "Surface," something is obviously wrong. Each attribute should have several characteristics, all of which usually have to be known in advance. The following is a list of what has to be considered.

1. What is the *type* of the value? For example, values could be text, number, decimal value or units such as meters, vehicles per hour, and so on.
2. What is the legitimate *range* of the values? For example, percentages should be between 0 and 100. Are negative numbers allowed? For text values, what spellings or range of choices (known as *categories*) are allowed? For text, how many characters long is the longest string?
3. What happens when there is a missing cell in the table? For example, a record could be missing an attribute such as the traffic count in Figure 4.12 because nobody was available to make the count. Often a *missing value flag*, such as the value −999 or NULL, is used in these cases. We obviously would not want these to count if we summed or averaged the rows or columns.
4. Are duplicates allowed? What if we had two road entries for Interstate 81 on Figure 4.15, one for the northbound and one for the southbound lanes? The traffic counts, road surfaces, and so on may be different and worthy of their own record. In this case, the values entered in the attribute column under "Name" would be identical.
5. Which attribute is the *key*? The key is the link between the two databases. So in the example of Figure 4.15, the attribute "ID #" should match the tag that was placed on the road when it was digitized from the map. Otherwise, all our attributes would be "lost in space."

Many of these questions must be answered when we set up the database to begin with. The tool within the database manager that allows this attribute setup is called the *data definition module*. It often has its own menus, language, and so on, and may need a programmer rather than a typical GIS user to set it up. In some cases, just as with the digital map data, the attribute data will have been found from an existing source, such as the Census Bureau's data files that link to the TIGER files. In this case, the links will already be made between the attributes and the features on the map. If there are new data, however, or if we make our own database for our own purposes, we have to make the links and check them ourselves.

A complete listing of all of the above information is called a *data dictionary*. Having the data dictionary in advance allows the part of the GIS that handles data entry, or the spreadsheet or database program we choose to use, to check each value as we enter it. Sometimes we enter the numbers and values one by one into a special part of the database manager called the *data-entry module*.

Often, we import into our GIS data manager all the records in a preexisting setup. Some of the more common databases and spreadsheets support specific formats for data exchange to allow this. The simplest form is to write a file with each of the attributes and their labels written as text, one per line, sometimes separated by commas and quote marks so that blanks and other symbols can be included. For example, the data in Figure 4.12 could be "unraveled" into the file listed in Figure 4.13.

```
Attribute_labels = "ID #", "Feature",
"Name" , "Surface" , "Lanes", "Traffic" , "per hour"
"1",
"Road",
"US 11",
"tarmac",
"3",
"113"

"2",
"Road",
"I 81",
"concrete",
"4",
"432"

"3",
"Road",
"Lisk Bridge Road",
"tarmac",
"2",
"12",
"4"
```

FIGURE 4.13: An unraveled or ASCII text from version of the flat file in Figure 4.12.

For new GIS data, the process of entering attributes eventually comes down to someone (usually the lowest-paid person) entering the attribute values one by one into a database manager's data-entry module. The data usually come from a data form of some kind, onto which they had been recorded painstakingly by the person collecting the data.

Some data-entry systems are better than others. At the very least, the system should check the type and range of the value for each attribute at the time of entry. At best, it is helpful if the software allows things like copying a record but then changing it to reflect a new value, deletion, or changing of values that are wrongly entered at the time of entry, and if the software brings errors to your attention with beeps and messages so that correction can take place immediately. No software package should allow data to be lost if the computer crashes, the file fills up, or the user presses the wrong button.

Most GIS packages allow the use of almost any spreadsheet, such as Microsoft Excel, or database systems such as DBase IV or Borland. Some require that you use the database entry system that comes with the GIS and no other. Each is slightly different, although all share the items discussed in this section.

4.6 EDITING AND VALIDATION

Many early geocoding systems had only limited editing capabilities. They allowed data entry, but error detection was by after-the-fact processing, and correction was by deletion of records or even whole data sets and reentry. Anything we can do in the geocoding process that reduces errors, or that makes errors easily detectable, we should indeed do. As an absolute minimum, data for lines and areas can be processed automatically for consistency, and any unconnected lines or unclosed polygons can be detected and signaled to the user. The connection between lines, known bordering of areas, and inclusion of points in areas is called map *topology*. Topology really comes into its own during the map validation stage.

The easiest way to avoid errors in geocoding is to ensure that errors are detected as soon as possible and then to make their correction easy. Video display during digitizing and audio feedback for error messages is essential. GIS software should spell out exactly what will happen in the case of an error. A common geocoding error is to overflow a hard or floppy disk or perhaps a disk-size quota while digitizing. It helps also to be able to recognize errors when they appear and to be able to understand their origin.

Some easy-to-detect errors are *slivers*, *spikes*, *inversions*, *lines that are not ended*, and *unsnapped nodes*, which we discussed in Chapter 3. Scaling and inversion errors are when the map appears squashed, like the titles at the beginning of a wide-screen movie shown on TV, or flipped. These are usually due to an incorrect digitizer setup procedure; that is, they are systematic errors caused by incorrectly entering the control points for establishing the map geometry. *Spikes* are random hardware or software errors in which a zero or extremely large data value erroneously replaces the real value in one of the coordinates. Spikes are also sometimes known as *zingers*. Errors in topology, missing or duplicate lines, and unsnapped nodes are operator errors.

Plotting the data becomes a useful aid because unplottable data often have bad geocodes. Similarly, attempting to fill polygons with color often detects gaps and slivers not visible in busy polygon networks. The best check for positional accuracy is a check against an independent source map of higher accuracy.

The equivalent of a plot for the attribute data is a *data listing* or *report*. Most data management systems have the ability to generate a report, listing the attributes as a table, or formatting them neatly for printing and checking. You should go line by line, checking the attributes and their values. However, even when the attributes and the map are validated by checking, it is still likely that errors exist in the links. One New York City database had more than 20 spellings for a single street name, for example.

The GIS often allows check plots to be generated that simply plot the label or identification number of the key within a polygon or next to a line. These maps and the tedious process of checking them should never be skipped. Moving straight on to making elegant graphics or doing a GIS-based analysis with erroneous data can be anything from embarrassing to dangerous, or even life-threatening.

A data set that is correctly geocoded both positionally and with attributes is not necessarily logically consistent. Logical consistency can be checked most easily for topological data. Topologically, data can be checked to see that all chains intersect at nodes, that chains cycle correctly in a ring around polygons, and that inner rings are fully enclosed within their surrounding polygons. Otherwise, attributes can be checked to ensure that they fall within the correct range and that no feature has become too small to be represented accurately.

Everyone would like to say that the data in his or her GIS are accurate and correct. Obviously, this means several things. Accuracy of position means that the locations shown on the map are in their correct locations with respect to the real world. Of course, there may be a difference between the map that was geocoded and the "best possible" map. Positional error is sometimes tested or measured, and this is best done against another map of higher accuracy or against accurate field measurements such as GPS fixes. Another aspect of data is the accuracy of the attribute. A map may be perfect as far as appearance is concerned, but the roads and rivers could both be mislabeled as power lines. This type of error can be treated as a misclassification. Testing can also be conducted and can even be automated as GIS data are already in a database management system.

A final issue is that of scale and precision. A map used for geocoding has a particular scale, such as 1 : 24,000. If this is the case, while the GIS allows us to compare data from another scale, say 1 : 250,000, it may not be appropriate to do so, as attributes, generalization of the features, and other properties of the map may be different at the two scales. Also, all data in the GIS have a degree of precision associated with them. If a highly detailed line is geocoded only to the nearest 10 meters on the ground, comparison with more detailed data becomes problematical. Generally, we should apply the same concerns and considerations of limitations to digital maps as we do to paper maps. Unfortunately, many people treat digital maps as absolutely correct maps instead of the digital alternative form of the analog maps to which they owe their humble origins.

The intelligent GIS user should know and understand the amount and distribution of error in a GIS database. Many of the sources of error are due to the method and process of geocoding. Some of the errors multiply as we move through the stages of data management, storage, retrieval, GIS use, and analysis. An understanding of error is essential to working effectively with GIS.

4.7 STUDY GUIDE

4.7.1 Summary

CHAPTER 4: Getting the Map into the Computer

Analog-to-Digital Maps (4.1)

- **A GIS depends on maps being available in digital rather than analog form.**

Finding Existing Map Data (4.2)

- **GIS data can be**

  (1) Purchased
  (2) Found from existing sources in digital form
  (3) Captured from analog maps by geocoding

- **Geocoding is the conversion of spatial information into digital form.**
- **Geocoding involves capturing the map, and also capturing the attributes.**
- **Existing map data can be found through a map library, via network searches, or on media such as CD-ROM and disk.**
- **Many major data providers make their data available via the World Wide Web, a network of file servers available over the Internet.**
- **Major federal agencies with WWW servers are the USGS, NOAA, and the Census Bureau.**

Digitizing and Scanning (4.3)

- **The method of geocoding can influence the structure and error associated with the spatial information that results.**
- **The two major geocoding methods for maps are digitizing and scanning.**
- **Digitizing on a tablet captures map data by tracing lines by hand, using a cursor and an electronically sensitive tablet, resulting in a string of points with (x, y) values.**
- **Scanning places a map on a glass plate and passes a light beam over it, measuring the reflected light intensity. The result is a grid of pixels. Image size and resolution are important to scanning. Small features on the map can drop out if the pixels are too big.**

Field and Image Data (4.4)

- **Much GIS data comes from field data collection, GPS data collection of positions, or from digital versions of air photos or satellite images.**

Data Entry (4.5)

- **Attribute data can be thought of as being contained in a flat file. This is a table of attributes by records, with entries called values.**
- **A database system contains a data definition module, which sets up the constraints on the attribute values, a data-entry module to enter and correct values, and a data management system for storage and retrieval.**

- **The legal data definitions can be listed as a data dictionary. A database manager can check values with this dictionary, enforcing data validation.**

Editing and Validation (4.6)

- **Validation and checking for map data is usually done using topology.**
- **Map and attribute data errors are the data producer's responsibility, but the GIS user must understand error.**
- **Accuracy and precision of map and attribute data in a GIS affect all other operations, especially when maps are compared across scales.**

4.7.2 Study Questions

Analog-to-Digital Maps

Define the following: *digital, analog, real, virtual,* and *geocoding*. Give some examples of maps that can be brought into a GIS free, and maps that would have to be geocoded.

Finding Existing Map Data

What GIS data are available to a user without access to the Internet? Give three examples of agencies that provide data over the World Wide Web, and list the type of data that each provides. What GIS applications might need each type?

Digitizing and Scanning

Discuss the differences between digitizing and scanning. What hardware and software would each require? What sorts of geocoding errors can each method generate?

Data Entry

What is the difference between map data and attribute data? Give the stages of setting up a database from scratch. Why is the data definition stage so important?

Editing and Validation

What software tools might be used in data editing? Name some of the common errors in geocoding that can be corrected by editing. Why might the value of an attribute in a record be invalid? What part of the database manager allows data editing and validation?

4.8 EXERCISES

1. *If you have access to the Internet, use the tools you have available (perhaps Netscape Navigator or Microsoft Explorer) to search for information about GIS and for digital map data online covering your town or city. Are there any attribute data for the maps, perhaps in gazetteers, almanacs, or data books? You may be able to get these from Internet providers. If you do not have access to the Web, visit a library and use its facilities, or look in the reference section for information. If you live near a map library, perhaps at a university, see if you can use this facility in your search. After a few searches, make an inventory of what data you were successful in locating. Put the inventory in the form of a list. Add a column to show which agencies supplied the data you found and how recent the data were. How are the data made available to the public?*

2. *Find the most detailed map you can of a place you know well. You can try your local planning office, a town or college map library, the county records office, or the town hall. Make a fact sheet about your search for other people to use, showing the source of the information, how you got the data, what the map showed, its scale, the date(s) of the information, and what features are shown. If you were to digitize this map for use in a GIS, what sort of applications would be possible? What other maps would you need?*
3. *Buy a copy of the U.S. Geological Survey 7.5-minute quadrangle covering the area where you live. You can find the maps in local camping stores, a map store, or a book store; copy them at a library, or you can get the name of map dealers near you by calling the Earth Science Information Center at 1-800-ASK-USGS (in the United States). Box out the section centered on your house, say 3 kilometers by 5 kilometers. Figure 4.8 is a section of a USGS quadrangle map scanned on a desktop scanner. This section of map was scanned, resulting in a file in TIF format that was 266,818 bytes in size. This was a file of color intensities between 0 and 255 for red, green, and blue in each of three layers spaced on a grid 0.25 millimeter apart. How much data would be necessary to capture the features on your map as vectors? Would it be more or less than the grid (raster) file?*
4. *Carefully read the documentation for any GIS package to which you have access. If you don't have a GIS, see if you can borrow the documentation, or get it as an online file. How does this package allow you to enter (1) data from other sources, such as the network, (2) raster data from a scanner, (3) vector data that you want to digitize from a paper map, or (4) data from other GIS packages? Make a log with entries for each of these capabilities. Keep this log, and compare it with logs for other GIS packages.*
5. *How does your GIS allow you to enter attribute information? Can you enter the data into a spreadsheet and move it into the GIS? What sort of user interface is used? What capabilities does the software have for data set definition? For the capture of errors during data entry? For validation of the data? For editing the data?*
6. *What steps are necessary for effective geocoding? Where did you make most errors in geocoding your first map? What steps in the geocoding process enabled you to (1) notice, (2) find or locate, and (3) eliminate these errors? How could these errors have been avoided to begin with?*

4.9 REFERENCES

4.9.1 Books

Bohme, R. (1993) *Inventory of World Topographic Mapping*. New York: International Cartographic Association/Elsevier Applied Science Publishers.

Campbell, J. (1993) *Map Use and Analysis*, 2nd ed. Dubuque, IA: Wm. C. Brown.

Clarke, K. C. (1995) *Analytical and Computer Cartography*, 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Makower, J. (ed.) (1986) *The Map Catalog*. New York: Vintage Books.

Thompson, M. M. (1987) *Maps for America*. 3rd ed. U.S. Geological Survey, Washington, DC: U.S. Government Printing Office.

4.9.2 Internet Addresses

U.S. Geological Survey `http://www.usgs.gov`

U.S. Census Bureau `http://www.census.gov/tiger/tiger.html`

NOAA `http://www.noaa.gov`

4.10 KEY TERMS AND DEFINITIONS

address matching: Using a street address such as *123 Main Street* in conjunction with a digital map to place a street address onto the map in a known location. Address matching a mailing list, for example, would convert the mailing list to a map and allow the mapping of characteristics of the places on the list.

analog: A representation where a feature or object is represented in another tangible medium. For example, a section of the earth can be represented in analog by a paper map, or atoms can be represented by Ping-Pong balls.

attribute: A characteristic of a feature that contains a measurement or value for the feature. Attributes can be labels, categories, or numbers; they can be dates, standardized values, or field or other measurements. An item for which data are collected and organized. A column in a table or data file.

data dictionary: A catalog of all the attributes for a data set, along with all the constraints placed on the attribute values during the data definition phase. Can include the range and type of values, category lists, legal and missing values, and the legal width of the field.

data entry: The process of entering numbers into a computer, usually attribute data. Although most data are entered by hand, or acquired through networks, from CD-ROMs, and so on, field data can come from a GPS receiver, from data loggers, and even by typing at the keyboard.

data-entry module: The part of a database manager that allows the user to enter or edit records in a database. The module will normally both allow entry and modification of values, and enforce the constraints placed on the data by the data definition.

digitizing: Also called semi-automated digitizing. The process in which geocoding takes place manually; a map is placed on a flat tablet, and a person traces out the map features using a cursor. The locations of features on the map are sent back to the computer every time the operator of the digitizing tablet presses a button.

digitizing tablet: A device for geocoding by semiautomated digitizing. A digitizing tablet looks like a drafting table but is sensitized so that as a map is traced with a cursor on the tablet and the locations are identified, converted to numbers, and sent to the computer.

drop-out: The loss of data due to scanning at coarser resolution than the map features to be captured. Features smaller than half the size of a pixel can disappear entirely.

drum scanner: A map input device in which the map is attached to a drum that is rotated under a scanner while illuminated by a light beam or laser. Reflected light from the map is then measured by the scanner and recorded as numbers.

editing: The modification and updating of both map and attribute data, generally using a software capability of the GIS.

flat file: A simple model for the organization of numbers. The numbers are organized as a table, with values for variables as entries, records as rows, and attributes as columns.

flatbed scanner: A map input device in which the map is placed on a glass surface, and the scanner moves over the map, converting the map into numbers.

FTP (File Transfer Protocol): A standardized way to move files between computers. It is a packet switching technique, so that errors in transmission are detected and corrected. FTP allows files, even large ones, to be moved between computers on the Internet or another compatible network.

gateway: A single entry point to all the servers and other computers associated with one project or organization. For example, the U.S. Geological Survey, though spread across the country and throughout dozens of computers, has a single entry point or gateway into these information sources.

geocoding: The conversion of analog maps into computer-readable form. The two usual methods of geocoding are scanning and digitizing.

Internet: A network of computer networks. Any computer connected to the Internet can share any of the computers accessible through the network. The Internet shares a common mechanism for communication, called a protocol. Searches for data, tools for browsing, and so forth ease the tasks of "surfing" the Internet.

medium: A map medium is the material chosen on which to produce a map; for example, paper, film, Mylar, CD-ROM, a computer screen, a TV image, and so on.

network: Two or more computers connected together so that they can exchange messages, files, or other means of communication. A network is part hardware, usually cables and communication devices such as modems, and part software.

NOAA (National Oceanic and Atmospheric Administration): An arm of the Department of Commerce that is a provider of digital and other maps for navigation, weather prediction, and physical features of the United States.

point mode: A method of geocoding in semiautomated digitizing, in which one press of the cursor button sends back to the computer only one (the current) tablet location.

real map: A map that has been designed and plotted onto a permanent medium such as paper or film. It has a tangible form and is a result of all of the design and compilation decisions made in constructing the map, such as choosing the scale, setting the legend, choosing the colors, and so on.

report: A listing of all the values of attributes for all records in a database. A report is often printed as a table for verification against source material, and for validation by examination.

scanning: A form of geocoding in which maps are placed on a surface and scanned by a light beam. Reflected light from every small dot or pixel on the surface is recorded and saved as a grid of digits. Scanners can work in black and white, in gray tones, or in color.

server: A computer connected to a network whose primary function is to act as a library of information that other users can share.

stream mode: A method of geocoding in semi-automated digitizing, in which a continuous stream of points follows a press of the cursor button. This mode is often used for digitizing long features such as streams and coastlines. It can generate data very quickly, so is often weeded out immediately by automated line generalization within the GIS.

TIGER: A map data format based on zero, one, and two cells; used by the U.S. Census Bureau in street-level mapping of the United States.

topology: The numerical description of the relationships between geographic features, as encoded by adjacency, linkage, inclusion, or proximity. Thus a point can be inside a region, a line can connect to others, and a region can have neighbors. The numbers describing topology can be stored as attributes in the GIS and used for validation and other stages of description and analysis.

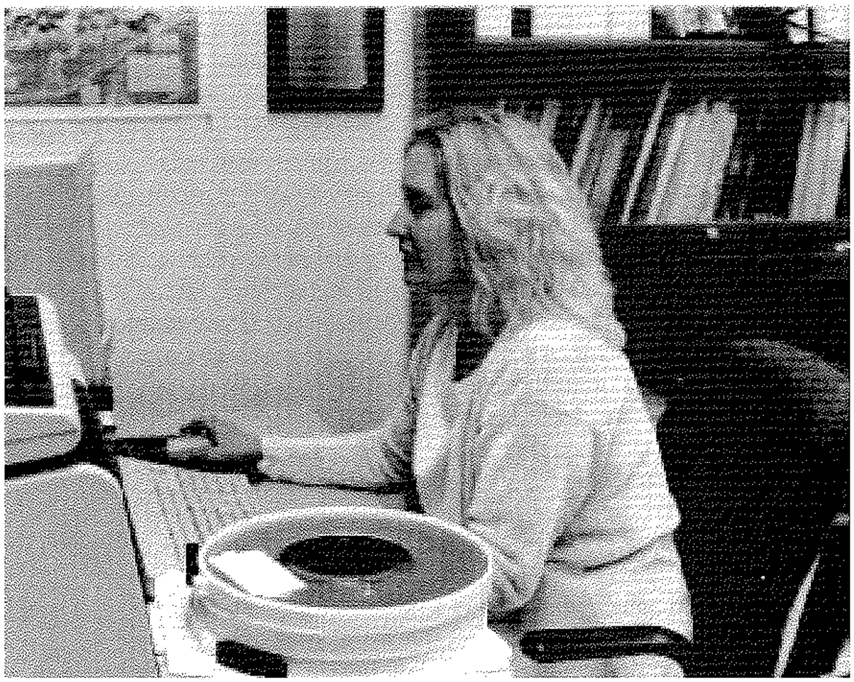
U.S. Census Bureau: An agency of the Department of Commerce that provides maps in support of the decennial (every 10 years) census of the United States.

USGS (United States Geological Survey): A part of the Department of the Interior and a major provider of digital map data for the United States.

validation: A process by which entries placed in records in an attribute data file, and the map data captured during digitizing or scanning, are checked to ensure that their values fall within the bounds expected of them and that their distribution makes sense.

virtual map: A map that has yet to be realized as a tangible map; it exists as a set of possible maps. For example, the same digital base map and set of numbers can be entire series of possible virtual maps, yet only one may be chosen to be rendered as a real map on a permanent medium.

PEOPLE IN GIS

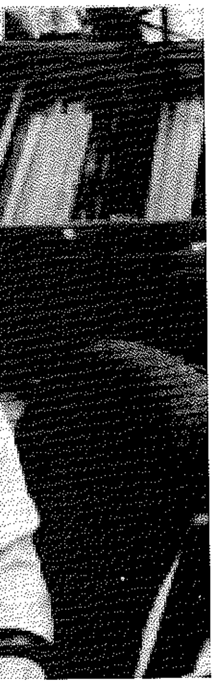


Susan Benjamin U.S. Geological Survey Physical Scientist

Susan Benjamin is a research physical scientist with the U.S. Geological Survey's National Mapping Division. She works at the EROS Data Center's field office at the NASA Ames Research Center at Moffett Field, California, near San Francisco. Raised and educated in the Bay Area, Susan has worked with remote sensing and GIS systems since graduating from San Jose State University with a B.A. and then an M.A. in geography. Susan has been involved in many research projects, including pioneering work on the new Digital Orthophoto Quadrangle images. She is the proud mother of a son and a daughter, ages 8 and 9.

ment of the Interior and

ı attribute data file, and
:ked to ensure that their
listribution makes sense.
ıap; it exists as a set of
set of numbers can be
hosen to be rendered as



st

J.S. Geological Survey's
enter's field office at the
ar San Francisco. Raised
sensing and GIS systems
nd then an M.A. in geog-
cluding pioneering work
e proud mother of a son

KC: Susan, how did your GIS career begin?

SB: After college I went to work on the ILIAC IV supercomputer at NASA Ames. My high school had emphasized math, and it was my minor in college. At that time, GIS courses existed at only one or two universities. You couldn't even get a degree in computer science; that was considered a tool, not a discipline!

KC: You picked math because you knew you were going to work with GIS?

SB: Yes. I knew that the math would be important in remote sensing and GIS. I started out in remote sensing. I became interested in the use of digitized photographs to identify features for making topographic maps. The National Mapping Division then was involved in making and revising large-scale topographic maps at the Western Mapping Center in Menlo Park, California.

KC: Is that where you first heard about GIS?

SB: I encountered GIS there for the first time in about 1985. In 1987 I was hired as part of the GIS Lab staff at USGS's Western Mapping Center. It was very chaotic; the lab had just been formed and equipment from other projects was being reassigned. We had SCITEX workstations that were being used by the USGS to produce digital maps for the 1990 census. Another workstation was used mostly by the Geologic Division to produce geologic maps and I was interested and wanted to work on that. It was the SCITEX that lured me here. At the time they were just setting up their Prime systems, getting Arc/Info up and running, and learning how to integrate Arc/Info into the map production process.

KC: How did you define GIS back then?

SB: I thought of GIS as computer programs that would make my job easier. I always thought that GIS was more analysis and that digital cartography was more production and digitizing of attributes. I thought of GIS as looking at what the attributes were, and drawing conclusions from them.

KC: Did you train as a photogrammetrist?

SB: Yes, I was the last person to be trained at Menlo Park on the PG2 stereo plotters. I learned how to take stereo pairs and set models and to use the floating dot to follow the ground and compile a contour map.

KC: What is a floating dot?

SB: A stereoplotter takes a set of two aerial photographs that have overlap between them, and lets you assemble the views to see three dimensions. The machine provides you with a floating dot that you can set at different elevations in the image. You have the dot follow the ground surface in three dimensions, tracing out a contour from the image. The kind of contours that are in the digital line graph and the digital elevation model data are made like this, only now it's all digital.

KC: What sort of GIS software have you used over the years?

SB: GRASS and Arc/Info. I have been through SPANS training. For remote sensing I've used ERDAS Imagine, which is often considered a GIS, and LAS, EDCs image analysis software.

KC: What's the plaque on your office wall?

SB: Back in 1990 I put out a sampler CD-ROM of digital orthophotos of Dane County, Wisconsin, the first CD-ROM that the mapping division produced with nothing but the DOQ images on it. I was responsible for assembling the data sets, organizing them on the disk, and putting them together for GRASS. I got all the descriptive information for premastering the CD; I took the disk to the pressing factory and made sure they were pressed and delivered to customers. The plaque on my wall is an award from URISA from the 1991 conference for the most innovative project in the project show case. It was for a hypercard stack that accesses the Dane County DOQ and lets you select digital photos for display and further use.

KC: That is great, thank you very much.

(Used with permission.)