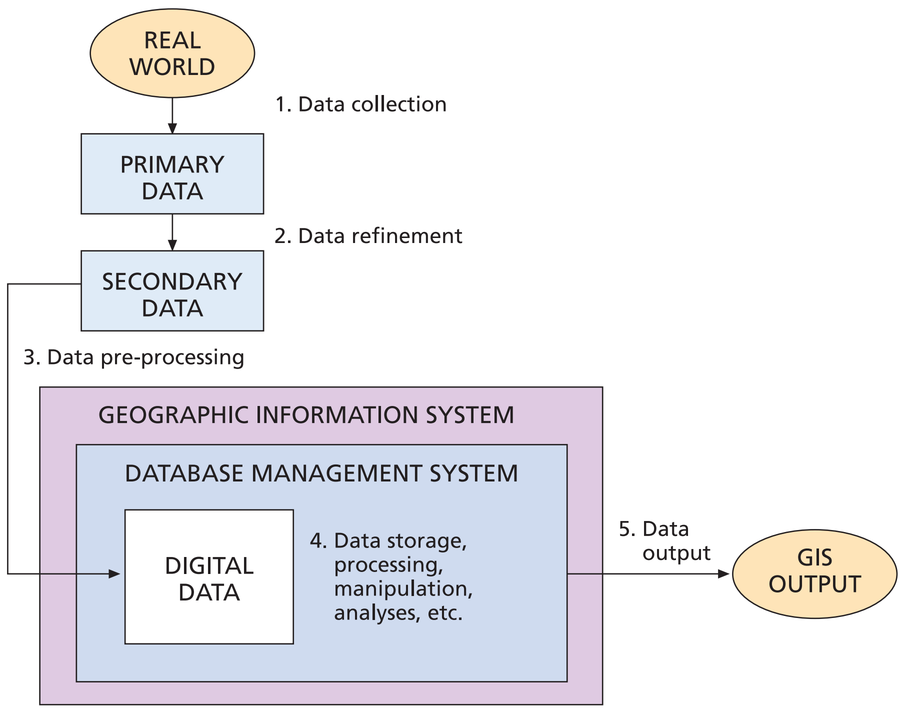
5. Preparing data for GIS use

G.J. Meaden (FAO consultant, Canterbury, United Kingdom), J. Jenness (FAO consultant, Flagstaff, Arizona, United States of America), and S.Walker (NOAA Coastal Services Center, Charleston, United States of America)

5.1 INTRODUCTION

GIS data can be collected using a variety of primary and secondary data collection methods, as discussed in Chapter 3. The next stage is to assemble and prepare both primary and secondary GIS data to support the GIS-based project.[113] Usable fisheries or aquaculture data come in many forms and formats. Some data exist via "cyberspace", i.e. online at a Web site but not yet downloaded. Other data are stored on digital data collecting devices, and yet other data are in the form of paper-based maps or tables, with some primary data written down in the form of a list. Whatever the data form, the data need to be prepared so that they are permanently accessible to the GIS software in an appropriate digital format. Many workers in the GIS field would agree that getting the data entered in a suitably formatted way, and having these data well edited and stored securely in a database, is the most important (and often time-consuming) task of the whole GIS project. This qualitative concern with data is one primary theme of this chapter, a concern that can be conceptualized by looking at the flow of data through a GIS from the "real world" to the "GIS output" (Figure 5.1).

FIGURE 5.1
Stages in data transformation when used for GIS purposes



[113] It will occasionally also be necessary to identify and collect "proxy" data (see Section 3.5.3).

Figure 5.1 illustrates the five stages at which some transformation of the data will occur; in this chapter, the focus is on the first three stages (data collection, data refinement and data pre-processing), plus some preliminaries of stage four[114] (data storage, manipulation, analyses, etc.). The other main focus of this chapter is to show how data can best be modelled and structured so that a GIS can properly function, i.e. so that the GIS software can use and present the data in a meaningful way. A consideration of how data is stored in databases and how databases function via database management systems is also made, and a brief look is taken at metadata – which is, essentially, information about data.

5.2 METHODS OF DATA INPUT

The methods of data input will vary according to the form or format in which raw data are held. This section examines scanning and digitizing, keyboard data entry and digital data transfer, and also addresses the types of information that need to be recorded, etc. It is important to mention data formats. Format refers to the way that data have been coded and stored in a data file or database. Whatever data that are input to any GIS, they have to be in a format that is readable by the GIS software. Although many GISs handle several formats, some will have their own formatting requirements. Some data formats will require special tools to import data into the GIS. Any specific data format will have strict rules on matters such as how a georeference should be stated (see Section 3.3.4), the range of values that are permissible and the number of decimal places to be recorded. A large number of data formats have evolved, mainly because data are collected for many different purposes. Over the last two decades, attempts have been made to standardize data formats, and much progress has been made on this (see Section 3.4 on Data Quality and Standards). There are now a range of software applications that convert data between formats.[115] Methods of handling remotely sensed imagery data are discussed in Chapter 6.

5.2.1 Scanning and digitizing

Scanners were discussed as one form of useful hardware for GIS in Chapter 2. The overriding purpose of scanning as an aid to GIS is to create digital map images. A simple A4-size scanner can easily create a digital copy of a paper map, aerial photograph, or any printed or hand-drawn diagram. Typically, these digital images will be recorded at resolutions of 200 to 400 dots per inch (dpi) on a flatbed scanner.[116] The resulting captured image can then be displayed on the computer monitor and the size of the image can be adjusted according to the degree of detail required. The scanned image per se cannot be used directly for GIS purposes because it lacks "intelligence". Thus, the scanned image is simply a copy of what was scanned, with the screen image consisting of rows of pixels to match the original paper map image. The information at each pixel point lacks any georeferencing or other information such as colour, attribute and line thickness and, therefore, the scanned image is incapable of being manipulated in any meaningful way. However, the scanned image is an essential prerequisite for digitizing. It should be noted that the accuracy of the scanned image also depends on the condition of the paper being scanned; for example, it may be creased, warped, torn or contain stains. Many of these problems, however, can be resolved in the editing process (as described in Section 5.3.1). After the image has been georeferenced, i.e. assigned real world coordinates so that the GIS knows exactly what area the image represents, the image can be used for either digitizing any selected features or as a background image to any overlaid map (see Section 7.3).

[114] Stage 1 (data collection) includes an array of methods as described in Chapters 3 and 6.

[115] See: Safe Software (www.safe.com) if information is required on converting data between formats.

[116] More detailed scanning of 500 to 2500 dpi can be obtained via the use of more sophisticated drum scanners.

Digitizing is a method of adding intelligence to data so that the data can be usefully used for GIS work. Chapter 2 discussed the use of a traditional large digitizing table for digitizing; however, today, almost all digitizing is carried out using so-called "heads-up" digitizing directly on the computer monitor,[117] using specialized on-screen digitizing software or, more typically, the digitizing function available in most popular GIS software. Whatever features of the scanned map (or image) are required, it is necessary to first georeference the image so that every point on the map can be accurately referenced to a coordinate (see Section 3.3.4). This can either be done "manually" by referring to the paper source or it can be done by overlaying the scanned image onto some other georeferenced data set. Digitizing itself is accomplished by progressively moving the on-screen cursor along any desired map outline and clicking on the mouse as digitizing proceeds. Digitizing procedures allow the user to record details of what has been captured, including feature names/types, registration positions, line thicknesses or colour (see Section 5.8.1 for further details). The digitizing program will build up a series of files containing all of the lines (or points) digitized, including their georeferenced locations and any attributes about them that need recording.

All maps and analyses performed later in a GIS project depend on the accuracy and precision of these data and it is, therefore, critical that the person doing the digitizing takes great care to perform the digitizing operations as accurately as possible. For greater accuracy, especially in areas where the map may be complex, it is possible to zoom into the scanned image while digitizing. As was shown in Figure 2.4, dual or split screens can be used so that both the original map and the already digitized lines can be viewed. In this figure, the green (woodland) areas have been digitized from the original map on the left. Use of the digitizing function within a GIS software package will ensure that the digital map data have been captured in a suitable format for that software.

5.2.2 Keyboard data entry

If data have been collected manually using no equipment, they are likely to be in a written, hard copy (or analogue[118]) format (see Section 3.1). Transforming these data to a digital format typically requires manually typing the information via keyboard entry, although in some cases it can be scanned and converted to text using optical character recognition (OCR) software. Data can be entered into a spreadsheet or into a database that has been designed for a specific purpose. For example, a typical fisheries or aquaculture survey should have a database set up with all the predetermined attributes listed (see Figure 5.2). Use of a spreadsheet is simple with each line (row) representing data from one sampling point, and the columns representing different attributes of the data. For GIS purposes, it is essential that one (or two) column(s) contains a georeference for each sampling point, using actual coordinates or a coded reference to a place name, zip code or some other areal unit. Any data entered manually must be saved as a file in a format suitable for accessing to the GIS program being used (see Section 5.5.1). GIS software will either contain its own inbuilt database program that allows the user to enter data directly to the GIS, or it can connect to a proprietary database or spreadsheet package such as Microsoft's Excel or Access. The keyboard will also be used to edit data in existing files as and when necessary.

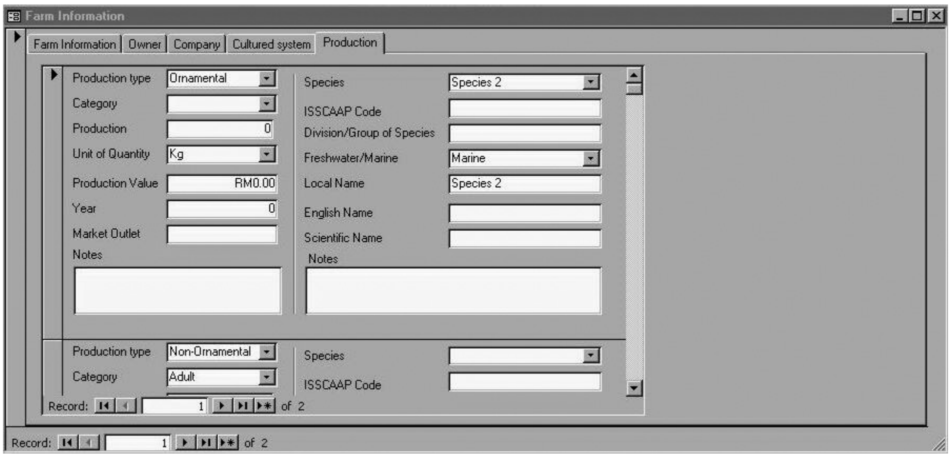
---

[117] It should be noted that there are other forms of digitizing that may use specially prepared "colour separate" map sheets, or which may rely on automated line following technology. However, these are unlikely to be appropriate in fisheries or aquaculture GIS tasks.

[118] Analogue format can include text, numeric or alphanumeric forms.

FIGURE 5.2

Example of a database set up for recording aquaculture production



Source: Sabah Integrated Coastal Zone Management (1999).

5.2.3 Digital data transfer

Although some data will be input to the GIS via scanning and digitizing or via the keyboard, today the bulk of data is likely to already be in a digital form. There are many mediums by which digital data may be transferred, including CD–ROMs, DVDs, data loggers, memory sticks, internal networks or the Internet. Because similar issues arise with all of these mediums, they can be discussed together.

A large proportion of digital data being transferred will not have been collected specifically for a GIS project. This means that data may have been collected in a neutral way by a recognized authority for a census or for topographic mapping, or data may have been collected for an entirely different purpose, so the user must work with the data available. These facts illustrate that some data will be reliable and accurate whereas other data will not be. While circumspect data may be usable, it is likely that the data will need updating, editing, reclassifying, etc., and, moreover, the data may not have been collected at a suitable scale or resolution. For these reasons, care should be taken wherever possible, to source data from reliable providers.

All data acquired will need to be in an appropriate format and, as indicated earlier, it will be essential to know the format(s) acceptable to the GIS software being used. Therefore, data downloaded from a global positioning system (GPS) or a data logger should be stored in the computer or server in the appropriate format, and it may be necessary to obtain conversion software to translate data into that format. Data that are being transferred use a standard file format called file transfer protocol (FTP). This means that a standardized set of rules (the FTP) has been established that allows data to be input to an exchange medium (CD–ROM, DVD, memory stick, the Internet, etc.), to be stored, to be transferred, and to be reassembled by a distant user.[119] Users need to be aware that some data sets that are downloaded over the Internet or available on CD–ROMs can be very large, as, for instance, remotely sensed imagery, and thus digital storage capacity may be an issue. If large data or information files need to be transferred, they can now be sent over the Internet by using a specialist file delivery service such as YouSendIt (www.yousendit.com) or ShareFile (www.sharefile.com). The YouSendIt service enables users to store and send files to others. YouSendIt offers several account types, which fall into the following categories: free, personal and corporate. Free accounts are offered at no charge and up to 100 MB can be sent by

[119] Data being transferred online will use the Hyper Text Transfer Protocol (HTTP).

e-mail. The other account types are subscription services requiring periodic payments (monthly or annually, depending on the account type) for sending larger amounts of data. A detailed description of the service and the features associated with different account types can be found at www.yousendit.com/compare-plans. ShareFile allows the creation of a custom-branded, password-protected area where business files can be easily, securely and professionally exchanged with clients; up to 2 GB of data can be sent by e-mail per delivery. Delivery costs can be low, though they vary quite substantially according to the amount of usage that is made and, therefore, whether a single delivery is paid for or whether a time period subscription is taken out.

Sometimes data sets are simply too large to transfer efficiently over the Internet using the technology most people have. Some data sets are hundreds of gigabytes in size, and the only way to efficiently transfer them is to load them onto a stand-alone hard drive and physically mail or ship the hard drive.

5.3 DATA VALIDATION AND EDITING

The data that have been collected and transferred to a digital format now need to be suitably organized and stored. However, before this is done users should confirm that the data are up to date and accurate and are, therefore, trustworthy. This may require validating, editing and correcting the data, and may also require other manipulations such as data transformations or reclassifications. Ideally, these edits and manipulations should be done before any analysis is performed, although constant reviews of data are always beneficial. Data validation is always a good habit for maintaining good quality data; however, a lack of time and funds often forces users to simply trust the quality of the data sources. As long as users review the data before their analysis they can decide whether to conduct a thorough validation or editing exercise or not. In this section, only the necessary editing and corrections to make sure that the data are trustworthy are considered; in Chapter 7, other optional manipulations are looked at, i.e. that may be necessary to transform data so that the data can be used for specific GIS projects. Additionally, the updating of data is not discussed because the importance of this is self-evident. However, users should know that maintaining current data can be a time-consuming and expensive task.

With respect to data entry, the sources of errors are numerous: the source material may have been incorrect or incomplete; mistakes are likely to have been made during the digitizing process; georeferencing might be incorrect or incomplete; data formatting might be unsuitable; and there could be many uncertainties regarding attribute classification and coding. These uncertainties need to be addressed before GIS work starts and before the errors begin to contaminate other data sets. Most GIS packages have their own set of editing tools, ranging from basic functions such as copying and deleting to other more advanced tools such as topological editing and attribute validation tools (see Section 5.8.1). The most comprehensive GIS software allows users to graphically "walk through" and edit the spatial errors. Other GIS software simply identifies the type and coordinates of the error. Some errors can be sorted out before they are input to the GIS, i.e. by means of editing functions within the digital data source such as in image processing software. Because validation and editing is often labour intensive and time consuming, users should consider the error correction capabilities carefully when evaluating any GIS software. Although the software should provide the ability to edit both graphical and attribute data, for many types of editing it may be necessary to do this manually. The editing processes are categorized under two headings: (i) graphical (spatial); and (ii) non-graphical (non-spatial or attributes), though editing may also be necessary if the spatial data are incorrectly linked to the non-spatial data[120].

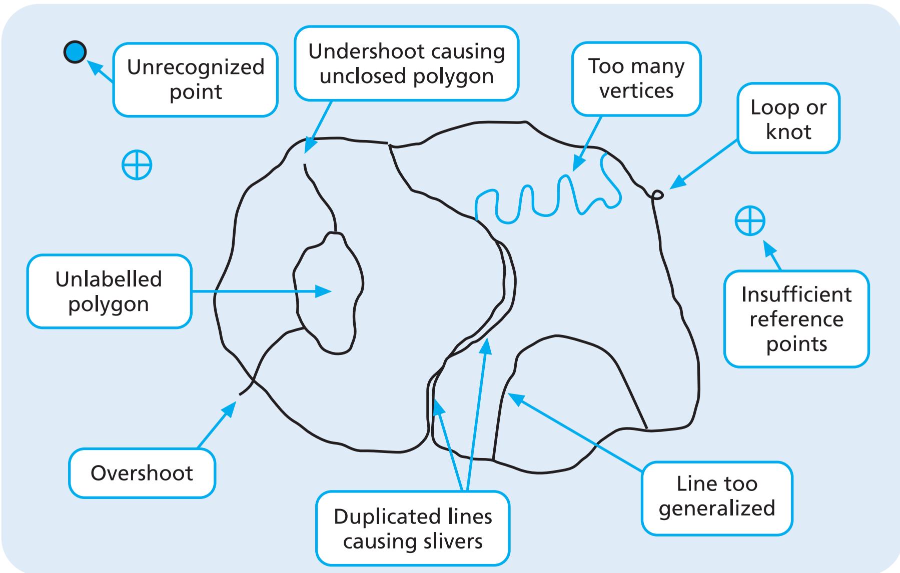
---

[120] These mistakes usually occur because an incorrect unique identifier (label) has been allocated during the digitizing process.

5.3.1 Graphical editing

Graphical or spatial editing refers to checking and correcting the digital data that are employed to draw the points, lines and polygons that appear on maps. Data obtained from a national mapping agency are usually reliable, though may require updating. However, data that may have been digitized, usually by a GIS team member, are frequently subject to many errors. Figure 5.3 illustrates some of the errors that typically occur during the digitizing process. When digitizing existing paper maps, it is best to validate graphical data during the data capture process.

FIGURE 5.3
Typical errors caused through the map digitizing process



The most common graphical fault in need of editing are so-called slivers. These usually occur because neighbouring polygons are individually digitized and the common boundaries do not match up exactly, although they may also occur when combining (overlaying) data from different sources. Slivers are easily eliminated through GIS editing procedures. Loops or knots occur when the person digitizing pauses while following a line and fails to restart in exactly the same place. Overshoots and undershoots are common when lines are not "snapped" to link with other lines.[121] It is particularly important to correct for undershoots because polygons will not be correctly identified. And at least four reference points must be used when geo-positioning the map; otherwise, the mapped outline may not be located in an exact position. Many more reference points may be necessary if the data must be mathematically warped or projected to conform to real-world coordinates. Other problems occur when the person who is digitizing forgets to add labels to lines or when the standard of digitizing is insufficiently accurate. Most GIS packages have tools that help identify digitizing errors and once identified the errors can be easily removed.

5.3.2 Non-graphical editing

The focus in this section is with checking and correcting the attribute data that correspond to the mapped features. Errors in attribute data may be caused by a lack of

[121] Users are generally able to select a snapping distance in which the end of a line will automatically snap to join up to the nearest alternative line. This distance will depend on the scale of the work and the accuracy required.

knowledge concerning the classification or categorization of data, in which case they may be difficult to correct, or they may also come from incorrect keyboard entries.

Attribute errors are usually more difficult to detect than graphical errors and, although it may be impossible to verify that every mapped feature has been described correctly, Heywood, Cornelius and Carver (2006) outline a number of methods that can be used to verify the data:

- **Impossible values.** Most data sets contain values that are only within a certain range. Data values lying outside the normal range may need verifying.
- **Extreme values.** These values may need to be checked against the source document.
- **Internal consistency.** Do the data that are in summary tables or that may represent mean or median values correspond with those given in source documents?
- **Scattergrams.** If two or more variables in an attribute table were correlated in the form of a scattergram, do any values depart noticeably from the regression line?
- **Trend surfaces.** Data often exhibits a regional trend, e.g. air temperatures will decrease towards the poles. Are there values that go against the trend?

These data verification methods apply to values that are assigned to attributes, but some of the methods could also be applied to checking the codes given to attributes. For example, Table 5.1 shows part of an imaginary file recording fishing vessel registrations in the Federative Republic of Brazil. The columns headed 1, 2, 3, 4 and 5 are codings given to the previous “attribute” column; in column headed “2”, for example, the numbers shown represent a code for the date (year) in which the fishing vessel was registered. It is relatively easy to see that the attribute codes conform to the known range of fishing vessels or the known range of construction dates, etc. Therefore, in Table 5.1, the codings in the column headed “2” may need to be verified because the coding “15” (data row 3) clearly indicates that the vessel was registered in a year ending in “15”; possibly 1915, which means that the vessel is nearly 100 years old yet it is constructed from glass fibre – this is almost certainly a mistake. Similarly, the coding “A” in the column headed “3” indicates a vessel that is not registered in the Federative Republic of Brazil, and this needs to be checked and resolved. Note that, as a means of updating data, it will be simple to add extra rows or columns of data to data sets such as this. A useful way to validate tabular data when they are to be entered manually is to ask two people to enter the same data and then compare the results. Each person is liable to make errors, but they are unlikely to make exactly the same errors. A search is made to find any cases where the two data sets differ, and the correct values can be determined using the source data.

TABLE 5.1
Hypothetical table (file) showing illustrative fishing vessel registrations in Brazil

Registration no.	1	Date registered	2	Home port	3	Overall length (m)	4	Hull construction material	5
RV 1443	43	10.06.97	97	Recife	7	16	16	Wood	1
RV 1447	47	15.07.96	96	Santos	3	31	31	Metal	3
RV 1451	51	23.03.15	15	Recife	7	20	20	Glass fibre	2
RV 1456	56	28.11.01	01	Niteroi	6	28	28	Wood	1
RV 1462	62	04.06.99	99	Montevideo	A	35	35	Metal	3

After the data have been validated and edited where necessary, they can be stored awaiting further use. However, before examining the storage of data in files and databases, it is first necessary to review data modelling. The structure and design of a good data model can be a critical factor in how well data can be used and analysed in a GIS.

5.4 GEOGRAPHIC DATA MODELLING

Initially, some words of caution are needed regarding the confusing terminology that abounds in the areas of data, databases and their associated geographic modelling. Readers might come across the term "data model" being used to portray at least three different concepts:[122]

i. Bernhardsen (1999) discusses "data models" in terms of the ways in which geographic data can best be structured in a GIS so as to represent the real world. This approach is essentially spatial data modelling and refers to the two main models – raster and vector – that are discussed in Section 5.8.
ii. Heywood, Cornelius and Carver (2006) discuss "data models" in terms of the way in which database management systems are structured in order to optimize data handling. This is essentially database modelling and is briefly mentioned in Section 5.5.
iii. Longley *et al.* (2005b) and Wright *et al.* (2007) use "data models" to indicate the ways in which the real world is conceived for digital mapping purposes. This is essentially concerned with the levels of abstraction and visualization required to create industry-specific digital database designs that can best be used within a GIS, and it is this concept of data modelling that is discussed here[123].

In order to create digital maps that will be used in any GIS, it is critically important that a data modelling process has been undertaken. Most GIS users will not be involved in this process because it will already have been undertaken in order that any specific GIS can function as effectively as possible. What is being done in the data modelling process is to undertake a series of steps or stages that create a set of rules allowing for any aspect(s) of the real world to be represented in a digital map form. Modified sets of rules can be used to create a range of geographic data models that can be used for different types of digital mapping purpose, e.g. the creation of topographic maps, marine charts and archaeology maps. Data modelling shows how real world information can be structured in such a way as to allow the portrayal (visualizing) of features on a digital map, or how spatial information can be intelligently structured as the basis for map production in a GIS. It is important to realize that there is no such thing as a "correct" geographic data model; every spatial area or project can be represented with many possible data models. Box 5.1 shows the stages in a hierarchy from the reality of the real world at Stage 1 to the increasingly abstract world of the map at Stage 5.

For many main areas of GIS, data models have already been developed that provide optimum ways of establishing databases. In the area of fisheries and aquaculture specifically, these models have not been developed, though Wright *et al.* (2007) provide evidence of extensive data modelling for the marine environment more generally.[124] Additional information on the rather complex topic of "data modelling" can be found in Bernhardsen (1999), Lo and Leung (2002), Rigaux, Scholl and Voisard (2002), or Manolopoulos, Papadopoulos and Vassilakopoulos (2005), and online information is available at Urban and Regional Information Systems Association (www.urisa.org/node/534).

[122] Causing even greater confusion is that the same authors often use the term "data model" in at least two of the three ways described here.
[123] For examples of this meaning of data models, see Environmental Systems Research Institute (http://support.esri.com/en/downloadsdatamodel).
[124] See ESRI ArcGIS Marine Data Model (http://dusk.geo.orst.edu/djl/arcgis/ArcMarine_Tutorial/).

BOX 5.1

Levels of abstraction in data modelling for a fisheries or aquaculture mapping scenario

1. The first level is the *real world*, i.e. the way that a vast array of natural and constructed features exist in any real world fisheries and/or aquaculture landscape or seascape.

2. The second *data or conceptual model* level is the abstraction of the real world to include the objects and processes that are seen as important to a particular problem area or theme. For example, if the GIS project was concerned with lake fisheries, special attention may be paid to the fishing vessels, fish markets, preferred species, the shoreline, etc., and little attention needs to be given to the range of tree species around the lake. This step also introduces human bias. Two people developing a data model for the same topic may derive rather different sets of data to be placed in the database, but eventually agreement must be reached. It is also important at this stage to consider the relationship between objects (or entities and attributes), and to keep all considerations completely isolated and independent from the computing environment.

3. The third level is the *data structure or logical model* – sometimes referred to as database design. This is concerned with the organization of data in a way that allows the data to be efficiently stored, managed and manipulated, and the data must also relate to the database software being used and be able to meet user requirements, as expressed in level 2. Here, it will be necessary to work out what sub-themes within, for instance "lake fisheries", will need to have separate files or databases and what attributes will be recorded for each sub-theme. It will also be necessary to work out the symbolism to be used to show all items to be mapped, including colours, fonts and legends.

4. The fourth level is the actual *file structure, or physical model*, referring to how the data are physically stored in the computer, plus the computing environment that will best prevail. This so-called "physical schema" is usually determined by the database management software's data definition language, i.e. the GIS user will have little say in this part of the data model.

5. The final level represents the creation of the *digital maps* that should be possible given that the databases have been carefully designed and structured. If data modelling has been done well, these maps should have the necessary intelligence to be optimally used for GIS purposes.

Note: The names shown in italic are the conventional terms for each stage or level.

Source: Adapted from Longley *et al.* (2001).

5.5 THE STORAGE AND MANAGEMENT OF DIGITAL DATA FOR GIS PURPOSES

As the needs for research into fisheries and aquaculture are being increasingly addressed, it is certain that data will accumulate at an accelerating rate. For maximum efficiency, all data collected need to be appropriately stored, organized, managed and shared, and this is accomplished via the use of files, databases and database management systems (DBMS). Once data are in a database, the data can be added to, edited, searched, ordered, etc. Specific database packages can be linked to GIS software, although major GIS packages now incorporate their own database functionality. This section examines the main factors regarding the storage and management of data.

5.5.1 Data files

As described in Section 5.2.2, data may be collected, stored and saved for future use in the form of files and/or databases. Files are the most basic form of data storage and common examples are text files, tables, shapefiles[125] and images. File-based tables commonly exist as comma- or tab-delimited text files, as dBASE files or as Excel tables.[126] Table 5.1 illustrates the basic structure for tables. The reader will associate this structure as the one adopted in spreadsheets. Each row in a table typically represents one record or one item from all those that are being surveyed or examined. Each column in the table represents an attribute (or field) of those surveyed items. Attributes, which may be numeric or textual, are usually assigned codings (or keys) by the database or GIS software, i.e. as a means of storing the information in an alphanumeric way that can be handled by the software. Columns headed 1 to 5 in Table 5.1 give examples of this. Tables are typically developed for a specific project or survey, though each table might contain records for a particular time period or a particular area. Theoretically, the number of rows or columns is unlimited; however, restrictions exist in the form of data storage and handling capacity, data collection costs, etc. When initially assembled, the records in tables do not need to be in any specific order, but it is essential that tables within a single database are assembled (formatted and structured – see Section 5.2) in the same way. Individual tables do not need to contain the same set of fields (columns), though if tables are to be joined and/or linked they must contain a specific linking or joining field (see Box 5.2) that is common to both tables.

Most tables used for GIS purposes are unique in the sense that they must include a column (or columns) that provides some kind of georeference. Without this, it would be impossible to map the data. The georeference (or unique spatial identifier) in Table 5.1 is shown in the column headed "3", i.e. because this is the only column that contains geographic (spatial) data that could be mapped. Here, the spatial data are in the form of place names, but in some files, tables or databases, the georeference may be in the form of a grid reference or latitude and longitude coordinates. All the files that make up a single database will need to be based on the same georeferencing code so that the files can be linked (see example in Box 5.2). It can be inferred from Table 5.1 that each fishing home port in the Federative Republic of Brazil will have a different number by which it can be identified. The spatial codings in attribute files nearly always refer to points or polygons on a map. Points may be very specific locations, such as a town position, road junction or any other specific feature. Polygons are more spatially extensive areas, such as counties or states, though note that the scale of mapping will affect whether a mapped feature appears as a point or a polygon.

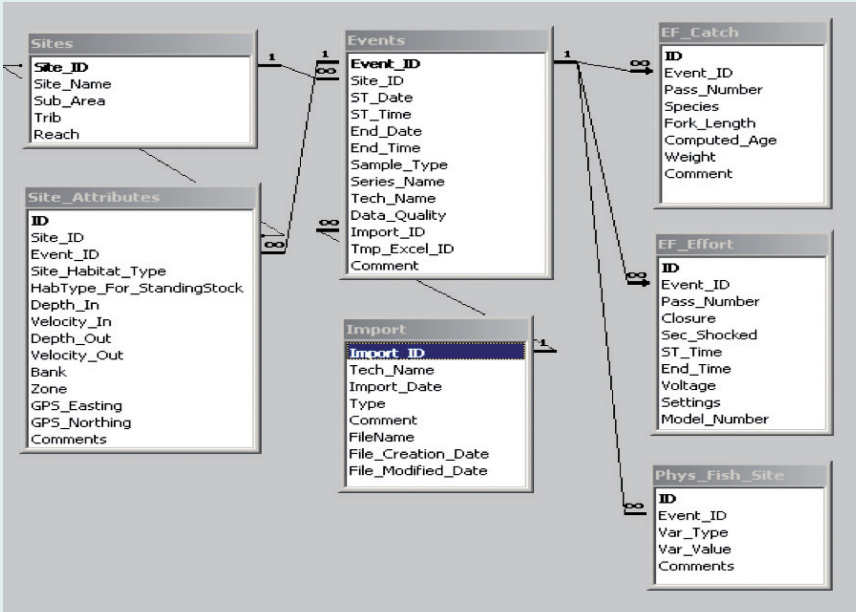
Shapefiles are similar to tables except that they also include a vector object (such as a point, line or polygon) that is paired with each row of data. While a basic table may include a pair of coordinates for each row or a coded value that refers to some spatial location, a shapefile includes the actual spatial object itself. There are other types of files that contain georeferencing information on the locations of point, line and polygon features on a map, such as field boundaries, roads and rivers – this is referred to as topological data and is covered in Section 5.8. It is the georeference unique identifier in files that allows separate files to be intelligently merged, joined or linked.

[125] The ESRI Shapefile (or simply Shapefile) is a popular vector data format for geographic information systems software. It was developed and regulated by ESRI as a (mostly) open specification for data interoperability among ESRI and other software products. The Shapefile format was introduced with ArcView GIS version 2 in the beginning of the 1990s. It is now possible to read and write shapefiles using a variety of free and non-free programs.

[126] In this section, the description of files concentrates on tables because they are easy to conceptualize.

BOX 5.2

Part of the database design for a 2007 fish census on the Lower Bridge River, British Columbia, Canada



Source: Lower Bridge River Database (2005).

The fishing authorities in British Columbia contract out regular fish surveys on a number of rivers and the data collected are stored in a relational database. This box shows the seven "internal index tables" that make up one section of the database called "Electrofishing". Each index table shows the file names in that section of the database and the attributes for which data are collected in each file. For instance, the "Sites" file will have the five attribute headings of "Site ID" (possibly an allocated code number); "Site Name" (perhaps the name of a place or a bridge); "Sub-Area" (the name of political areal unit, e.g. a county); "Tributary" (the name of the river); and "Reach" (the name of the particular stretch of the river). Other databases for the same river section include "population estimates", "snorkel surveys", "benthic" and "habitat" data. Together with the "Electrofishing" database shown above, these data sets make up one large database that is continually updated. The linkages show common identifiers (ID) that allow the tables to be connected to one another. For instance, the file "Events" can be connected to "Import" because they both have the common attribute column named "Import ID". It is essential that each index table includes at least one ID that is common to at least one other index table. As this is one of ten databases that make up the complete fish census database for this single river, it is clear that much thought will have been given to the conceptual model that was devised in order to best develop this particular database structure.

5.5.2 Databases

As with "data modelling", the term "database" covers a spectrum of related meanings. At one extreme, Longley *et al.* (2001; p. 18) say that a database "consists of a digital representation of selected aspects of some specific area of the earth's surface or near surface, built to serve some problem solving or scientific purpose". At the other extreme, a database can be seen as a small collection of perhaps half a dozen related files. Whatever the description, a database here refers to an integrated set of data (nowadays usually in digital format) covering a specific subject area. The information stored in a database can refer to: (i) graphical (locational) data, which is digitally georeferenced to show where things are (see Section 5.3.1); (ii) attribute (entity) coding, which is coding to show what things are (see Section 5.3.2); and (iii) relational coding, which shows how features are related to one another. A database that contains locational data is often referred to as a geodatabase. For GIS purposes, the focus is on spatial databases, meaning that they must have georeferencing capabilities. Databases can consist of a collection of data files that are in tabular spreadsheet format (Table 5.1) and/or in special database formats (Figure 5.2), and the collection of files in a single database will all relate to the same topic or theme. Box 5.3 outlines the advantages of using digital databases.

BOX 5.3
Some advantages of using digital databases

For GIS to operate effectively, data need to be organized in databases. This data organization method offers the following advantages:

- All data can be assembled at the same place, thereby allowing efficient data sharing, management and reduction of duplication.
- Maintenance costs can be greatly reduced, and shortcomings in data holdings can be easily recognized.
- Multiple applications will use the same data, so this gives confidence that GIS output will not show conflicting results – it will minimize inconsistencies.
- Security and standards for data and data access can be established and enforced.
- Access to data is controlled and centralized, though non-programmers have a predefined interface to help them.
- Standardized query languages (especially SQL, or "structured query language") are available that allow for efficient posing of a huge range of questions.

Sources: Healy (1991) and Date (1995).

Databases are essential to GIS software packages because they provide the structure for the data stored so that it can be manipulated in a multitude of ways. From a practical viewpoint, database usage may function at different scales. Thus, a very small fisheries or aquaculture institute may hold all of their data in one computer or data server. This is convenient in many ways because one person can easily manage the data in terms of their editing, updating, file ordering and security. However, it is likely that in many regions or countries, fisheries-related GIS data holdings will now be at a larger scale. Here, a special department in charge of database information is common and data are distributed from a central server via the Internet or a LAN. In the case of a departmental database, matters relating to data management, security and upkeep may be operated on a totally different basis than for the small-scale GIS. For instance, because organizations can hire database specialists, they can benefit from economies of scale and thus data can be more easily shared allowing for improved data performance to be recorded.

5.5.3 Database management systems

A database management system (DBMS) is a software package for storing, accessing, modifying, maintaining and retrieving data from databases. There are a large number of DBMS software packages available, but until relatively recently the majority of them lacked the ability to handle spatial data. Until the mid-1990s, GIS-related database management functioned on the basis of external links to separate DBMS, but more recently specialist DBMS vendors have either developed spatial database applications or the GIS software developers have integrated DBMS functions into their packages. DBMS can work with different types of data, they allow databases to be manipulated, to be queried or searched, they provide programming tools, they have particular file

BOX 5.4
The main functions of a database management system (DBMS)

- **Data model** – DBMS include standard general purpose data models suitable for representing several types of themes and objects.
- **Data load** – The provision of tools to load data in standard data types or formats.
- **Index** – An index is a data structure used to speed up searching.
- **Query language** – All DBMS support a standard data query and/or manipulation language called structured query language (SQL).
- **Security** – DBMS provide controlled access to data, e.g. some people have no access rights and others may be restricted to limited sections of the data.
- **Controlled update** – Allows for the management and updating of data sets or databases. This is important when updates affect several different data sets.
- **Back-up and recovery** – In the case of a system problem or failure, such as an electrical failure or incorrect (or accidental) updates or deletions, the data will be recovered.
- **Database administration** – Includes a number of tasks, such as setting up the database schema, creating and maintaining indexes, tuning to improve performance, allocating access rights. Specialized internal software applications conduct this work.
- **Applications** – DBMS include standard general-purpose tools for creating, using and maintaining databases.
- **Programmable API (application programming interface)** – Although DBMS have general-purpose applications for standard use; some specialist applications may require further customization.

Source: Adapted from Longley *et al.* (2001).

structures, and generally they maximize the efficiency in which the data can be utilized for carrying out the whole range of GIS tasks. Box 5.4 provides a more detailed list of the main functions required of a DBMS. It is clear from this information that the average user of a GIS will expect these functions to occur, yet he or she needs to have little knowledge of the functions that are built into the DBMS or the GIS software. In larger organizations, the database manager oversees all the DBMS operations, including adding new data sets or files, regulating access to the system, understanding and adjudicating on legal aspects of data usage, and coping with system problems or failures.

Database management systems themselves can be structured according to four different basic models: networked, hierarchical, relational and object-oriented models. Of these, the relational model has become the most widely used for GIS and for other purposes (Heywood, Cornelius and Carver, 2006) and only this model will be summarized here. Details on the other DBMS structures can be found in Bernhardsen (1999), Longley *et al.* (2011) and in many basic DBMS texts. The structuring of the DBMS relates to the way each database is structured in terms of its efficiency in performing database queries.

The relational database management system (RDBMS) requires that data are organized in a series of tables having columns and rows (as per Table 5.1 or as per a digital spreadsheet). Each cell in the table can contain a "value" that can consist of alphanumeric data having a permissible range of values or meanings. Individual data tables in the DBMS can be linked by common data fields known as "keys". Box 5.2 shows the computer-generated "internal index" tables (the data structure) showing files that make up part of a database that has been developed to record fish surveys in a river in British Columbia in Canada. In the Electrofishing database, the linking "keys" between tables could be "Event ID", "Site ID" or "Import ID", i.e. these "keys" occur in more than one file. Because individual files or databases can be readily linked, the database as a whole can be queried and quite complex questions can be asked. So-called "structured query language" (SQL) has been developed to enable the querying of databases.[127] As Box 5.2 shows, the "Site_Attributes" table is linked to the "Catch" table (via the "Events" table). Based on this link, the user could write an SQL query that would calculate, for instance, the number of fish of a certain weight that were caught in water flowing within a specified velocity range, and through the use of GIS software with its spatial database, this information could readily be mapped. Additional queries can be made on any number of linked tables as required to provide answers, and the queries themselves can be constructed on the basis of Boolean logic.[128] Most proprietary DBMS software generate dialogue boxes providing information on what queries can legitimately be made, and this will depend on what data are stored and whether the data can be linked. Because of the complex tabular linkages that can be created in RDBMS, the system may be inherently slower than other DBMS and efforts have been made to speed operations up through the use of object-relational DBMS.[129]

5.6 METADATA

When collecting data for use in GIS, it is important that the collector has information about the data that are being gathered. This information describing data sets is called metadata. The contents of the metadata will inform the user on whether any data set is worth acquiring – is it reliable to use? Each set of metadata may include a range of variables; for instance, Box 5.5 illustrates the United States Federal Geographic Data Committee recommendations of the major features of geospatial[130] metadata. This list would vary only slightly from country to country, though in many countries metadata quality and standards are still being developed and updated (see Chapter 3).[131] Metadata are stored in a database (a meta-database), and this is intended to be an effective link between the user and the data producer in that they provide relevant information about the various data sets that are held.

From the user perspective, comprehensive metadata have the following advantages:

- Provides a means by which the person(s) producing any data can maintain an overview of what has been compiled over time. This avoids possible duplication of data. Proper documentation creates an institutional record of data holdings.
- Allows the sharing of reliable information with others by allowing for the inclusion of the user's metadata records on metadata clearinghouse servers.
- Helps with the publication and promotion of the data creator's work.
- Reduces workloads in the long run by allowing the creation of templates for an organization's metadata records.
- Helps users to understand the strengths and limitations of the data simply by reading the descriptive information, and helps them to decide whether a particular data set is appropriate for their purposes.
- Helps organizations protect their investment. Data creation or collection is a time-consuming process and it uses resources that could be used elsewhere.
- Documents the process used to create a data set that allows others to replicate that process if the original data collector leaves the organization.

---

[127] SQL is a standard computer language designed for accessing, querying and manipulating databases. It allows the user to search for specific information, retrieve it, delete it or add to it, update it and to combine queries.

[128] Boolean logic is commonly used in search engines and it implies the use of delimiting words in queries, such as "and", "or", "equal to", "greater than" and "less than".

[129] See SearchOracle.com (http://searchoracle.techtarget.com/definition/object-oriented-database-management-system) or University of Liverpool (www.csc.liv.ac.uk/~dirk/Comp332/COMP332-ORDB-notes.pdf) for an introduction to object-relational database management systems.

[130] Geospatial tends to be used as a way of differentiating between "geographic" data and data that might be concerned with "space". For simplicity, geospatial can also be referred to as "spatial data".

[131] Other metadata standards include the International Organization for Standardization and the United Nations Environmental Programme standards, plus the standards that are compatible with existing metadata databases, such as the National Aeronautics and Space Administration, Center for International Earth Science Information Network, and the World Conservation Monitoring Centre.

BOX 5.5

Main characteristics of metadata as defined by the United States Federal Geographic Data Committee

- Data set identification – Details on the name, ownership, themes, etc.
- Data quality – How accurate and complete are the data, plus consistency, lineage, timeliness and derived data sources.
- Spatial data organization – The data model used, the number of variables collected and details on non-coordinate methods of location encoding.
- Spatial referencing information – Coordinate system used, plus projection and datums, and what spatial area is being recorded.
- Data exchange format – Information on the structure of the data – how it is stored.
- Entity and attribute information – For each attribute, what information is recorded and any coding used.
- Distribution information – Where and from whom can the data be obtained, including available formats, online or other availability, copyright details and costs.
- Metadata reference information – Contact address for organization producing the data. Responsibility for data upkeep.

Source: Adapted from Federal Geographic Data Committee (2012).

As a GIS user, the first thing that should be done when collecting primary geospatial data is to create metadata about this data.[132] Metadata should always be reviewed when collecting secondary data and, if no metadata are present, then the user should give serious consideration as to whether to use the data or not. Can the user have confidence in its reliability? With inadequate metadata, there is no assurance that the data were collected with enough accuracy or precision to use in a serious project. More information on metadata can be found at the following sites:

- Federal Geographic Data Committee (www.fgdc.gov/metadata);
- Open Geospatial Consortium, Inc. (http://opengeospatial.org);
- Environmental Systems Research Institute (ESRI) (www.esri.com/library/whitepapers/pdfs/metadata-and-gis.pdf);
- International Organization for Standardization (www.iso.org/iso/catalogue_detail.htm?csnumber=26020).

5.7 FROM REAL WORLD FEATURES TO A MAPPED WORLD

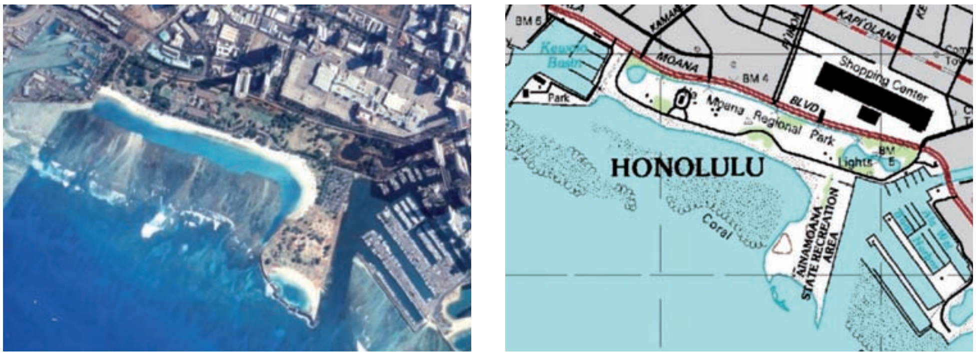
A brief look at geographic data modelling was made in Section 5.4. The focus now is to look at the basics of digital mapping before exploring in more detail four important data models that have evolved for operational GIS. From the GIS perspective, at its most basic level, geographic data are nothing more than an electronic representation of real world features. For the GIS technicians who will be creating spatial data and using data created by others, it is vital that they understand the process of generating mapped information from real world features. It is incumbent upon them to try presenting reality as accurately as possible when data are created. This means that GIS technicians must have a good understanding of the features that they are trying to represent as geospatial data (maps), and this can be quite difficult to achieve.

Figure 5.4 illustrates how real world features (as detected from a remotely sensed image) might be represented as mapped data; it is clear that to derive the map from the real world a series of decisions must be made – this decision series is shown as the five levels of abstraction in Box 5.1, and this modelling is described in more detail in Bernhardsen (1999) and Lo and Yeung (2002). Box 5.1 clearly illustrates that the mapping processes involve considerable generalization and classification. Thus, the real world is complex compared with the mapped world, and the art of the cartographer is to identify the important information that needs to be displayed

[132] It will also be necessary to update the metadata whenever editing or updating of any data occurs.

FIGURE 5.4
Real world spatial features represented as geospatial (mapped) data

Realworld space **Spatial representation**

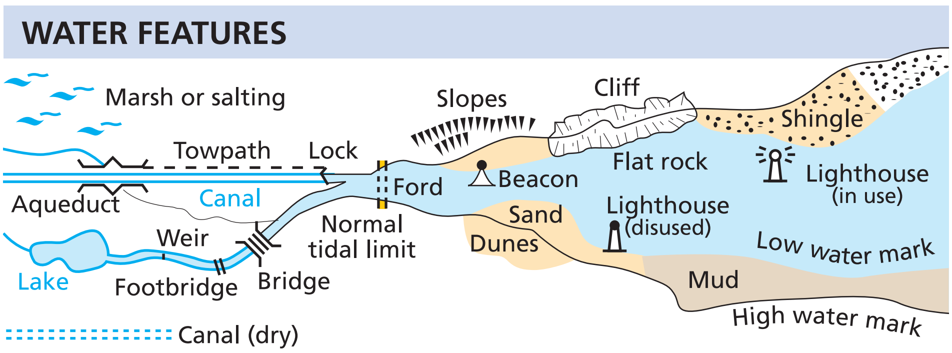


Source: Feinholz and Walker (2007).

and to establish ways of symbolically representing this important information.[133] The process of generalization in GIS is a recognized (and quite skillful) means of simplifying reality while at the same time maintaining as much as possible the true position and shape of the features to be mapped.[134] Cartographic classification forms an associated process, i.e. how best to classify groups of similar objects (entities).

For instance, in the real world space of Figure 5.4, the roads are different sizes, and the spatial representation of these roads has classified them into four categories: (i) a main "boulevard" shown as a double red line; (ii) a dual highway shown as a double black line; (iii) main roads shown as single thick black lines; and (iv) smaller roads shown as thin black lines. Good maps will have legends that convey what object classifications have been made for the construction of that particular map (Figure 5.5). Through the processes of classification and generalization, the cartographer will identify all those features of the real world for which classes of objects (entities) have to be described. As well as the roads just mentioned, there will be buildings, forests, waterways, etc. Each entity type may have subcategories. Thus, buildings might be hospitals, offices, factories, houses, etc., and each entity or entity subcategory must also have attributes. Attributes are terms that describe the entity. For instance, houses could be brick,

FIGURE 5.5
Legend to show water-related features on a topographic map



Source: Ordnance Survey (2012).

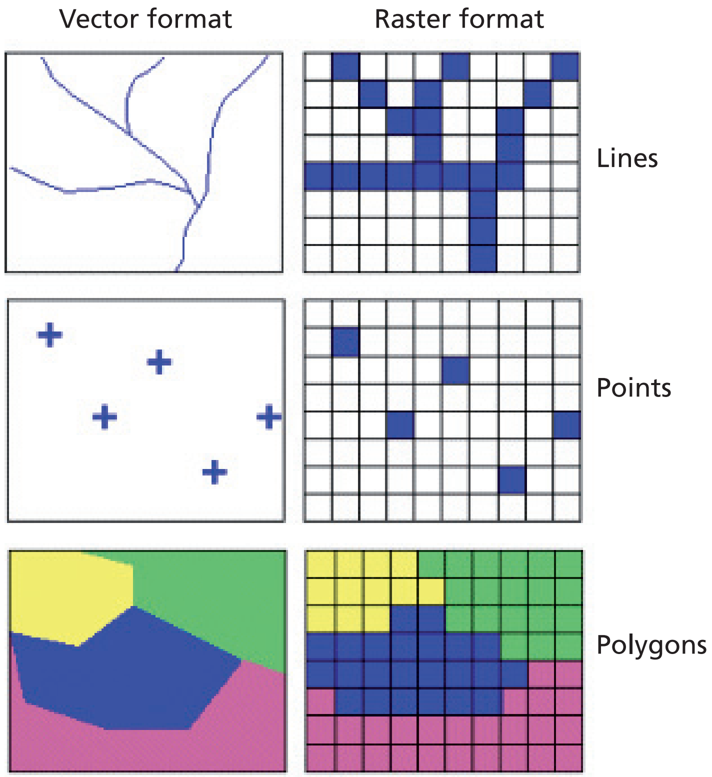
---

133 Information that is considered as important to map will vary according to the purpose of the mapping being undertaken, i.e. this is the basis on which "thematic" mapping occurs.

134 The process of generalization is very important in cartography; further details can be found in Longley *et al.* (2005b) and Harvey (2008).

or wood, and they could be detached, three storey or terraced. Clearly, map construction has to be able to "accommodate" the depiction of a vast array of entities and their attributes. Tables of data are constructed to convey this information to the computer (see Section 5.8.1 and Table 5.2). Moreover, the data being conveyed must contain both graphical information (that associated with where and how to draw the mapped information) and the entity and attribute data associated with the description of the features being mapped.

FIGURE 5.6
Line, point and polygon representations in vector and raster formats



Source: National Informatics Centre (2012).

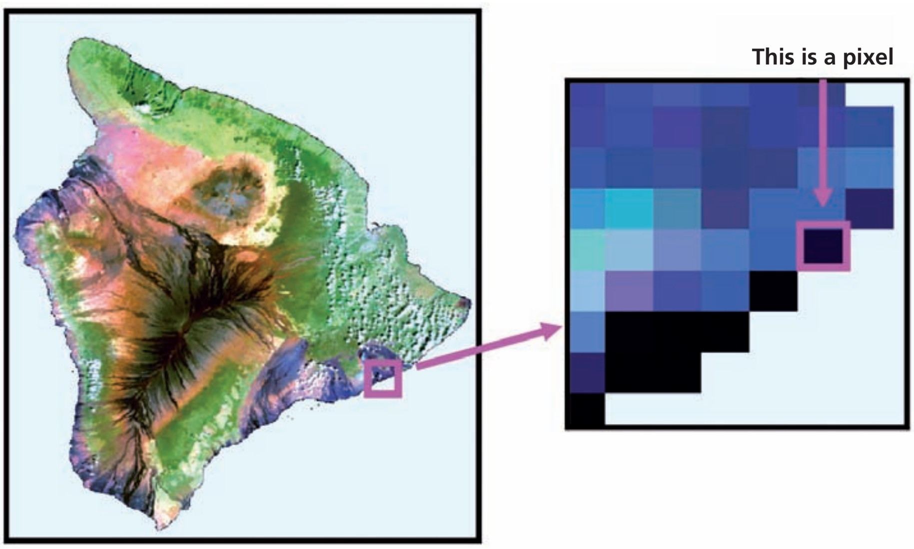
Figure 5.5 shows the legend that describes all the water-related features appearing on a United Kingdom Ordnance Survey topographic map at a 1:50 000 scale. A close inspection of either the map (spatial) representation in Figure 5.5 or the map legend in Figure 5.6 shows that all the graphical (drawn) features shown on maps can be conveniently categorized as being either point, line or polygon features, as illustrated in Figure 5.6. This means that in terms of the map drawing, or the data modelling processes shown in Box 5.1, all features in the real world can be readily depicted on a digital map as some kind of point, line or polygon (areal) feature.[135] The legend in Figure 5.5 illustrates point features such as beacons, lighthouses, footbridges, locks and weirs; line (or linear) features such as the high water marks, canals, towpaths and streams; and polygon (areal, spatially extensive or 2D) features such as mudflats, lakes, shingles, and sand and water areas in the estuary. It is important to know that, with a change in scale, point features can become polygons or lines and vice versa. For instance, on a small-scale map (showing a large area – typically >1:250 000), the aqueduct shown in Figure 5.5 might only be a point feature, whereas on a large-scale map (<1:50 000) this might be shown as a linear feature. Equally, the lake could change from a polygon on a large-scale map to a point on a small-scale map. These points, lines or polygons may represent any of an almost infinite number of entities, and their correct location is achieved through some form of georeferencing having been assigned to all entities that are saved in a GIS database.

[135] Some GIS texts identify a fourth form or dimension, i.e. that of "surface". This indicates the fact that land or seascapes have a vertical dimension that needs consideration in some GIS work. Some GIS software packages recognize time as another dimension and allow temporal processes to be mapped and analysed.

5.8 THE STRUCTURE OF SPATIAL DATA

Real world features, as discussed above, can be abstracted into three main feature types: points, lines and polygons (sometimes also called positions, networks and fields). These different feature types can be conveniently stored in two different data structures (also called models or formats) that allow them to be used by GIS software packages, and these are called vector and raster. In principle, both of these structures can be used for the mapping and analysis of points, lines and polygons. However, in practice, there is a strong association between the vector structure and points or lines, and between the raster structure and continuous surfaces.[136] Continuous data are data that either vary across a surface without discrete steps such as temperature or elevation, or data that uniformly and continuously cover an area such as land use or geology (Figure 5.7). Figure 5.6 shows how each structure can be used for this depiction, and a detailed account of the vector and raster structures follows. It should also be noted that the two other basic types of modelling examined in this section – network analyses and 3D analysis – can both be pursued in a vector or raster environment, although vectors are more commonly associated with networks, but 3D surface (or terrain) analyses can use either data structure. Traditionally, different GIS software packages were developed to either handle vector-based or raster-based functionality, but today all major GIS software is capable of handling both data structures and, indeed, they can readily switch between these and data can be converted from one structure to the other. Each data structure has certain advantages and disadvantages and these are examined below. A final introductory point is that much of what follows here describes functions that are embedded within any GIS and thus knowledge of these spatial structures might not be an essential prerequisite for GIS work. However, it is important that background information is provided so that at least the GIS operative is aware of the fundamentals upon which the GIS are operating.

FIGURE 5.7
Raster data set showing land cover on the island of Hawaii, United States of America



Source: Feinholz and Walker (2007).

5.8.1 Vector data structures

The vector data model is ideally suited for representing those real world points, lines or polygons[137] described in Section 5.7, and it does this by storing the feature's real world

[136] Continuous surfaces frequently appear as polygons on thematic or topographic maps.
[137] These are also described as discrete objects because the points, lines or polygons represent specific objects on the earth's surface.

locations as x and y coordinates using some form of georeference, such as latitude and longitude, grid references or state plane coordinates (see Section 3.3.4 and Box 3.5). In this format, a point is simply a coordinate pair giving the location of, for instance, a well or a pond. In the vector data format, a road (or any other linear feature) might consist of a string of nodes stored as x,y coordinate pairs and the software would "connect the dots" when drawing the feature. Likewise, a polygon has its nodes stored as x,y coordinates with the first and last coordinate pair being the same. By using precise georeferences, e.g. perhaps to several decimal places, then it is possible to gain very accurate locations for features, though to some extent the thickness of the drawn mapped line may impede true accuracy.

To store and use vector data for cartographic or GIS work, a number of specialized file formats have been developed. Although there are many formats, the three main types of formats are spaghetti data, topological and ESRI Shapefile, and within these three there are many variations on each format, some of which are specific to the main GIS software houses.

Spaghetti data. This is the simplest vector data format, and though useful for simple drawing or cartographic representation, it is unsuited for GIS work. Spaghetti data can best be described as digitized points, lines or polygons that lack intelligence. Thus, for basic cartographic drawing using the vector structure, the software needs to record a number of attributes about what is being drawn. Apart from the coordinates of every necessary point, line or polygon, the software needs to know what attribute is being drawn. This is achieved by the use of numerical coding[138] series that are compiled by the GIS software as mapped features are drawn, i.e. following the instructions given by the person digitizing the map. For instance, Table 5.2 shows typical numerical coding allocated to vector graphical data in order to identify entity (object) groups and object attributes. Other coding information will need to describe factors such as the line width and types to be drawn, the colour to be used and the size and shape for drawing points, and this information will simply be added to the attributes table (started as per Table 5.2). This information can be useful for basic map drawing purposes, but what is lacking is additional topological information showing how the objects digitized are related to each other. Without this information spatial query and analyses are impossible.

TABLE 5.2
Typical numerical coding allocated to vector graphical data

Numerical code series	Object group
1 000	Survey control stations
2 000	Terrain formation
3 000	Hydrography
4 000	Boundaries
5 000	Built-up areas
6 000	Buildings and facilities
7 000	Communications
8 000	Technical facilities
Numerical code	**Object type**
4 001	National border
4 002	County boundary
4 003	Township boundary
4 011	Property boundary
4 022	National park border

Source: Adapted from Bernhardsen (1999).

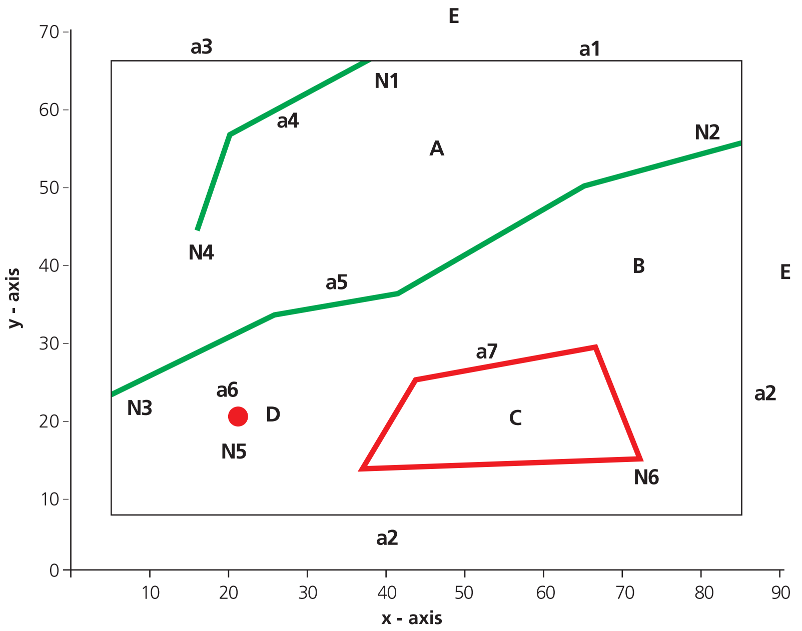
[138] Numerical coding is also called "feature coding". These codes are numeric attributes used in digital map data to describe each feature in terms either of the object surveyed or its representation on the map (or both).

Topological data formats. In order to gain the intelligence for GIS work, the traditional means of formatting vector data has been through the use of the so-called topological data format. Here, a description is made of the basis of this formatting, but there are more variations than the ones described. This data format is based on the concept of topology. Topology is the field of geometry concerned with the mathematics of spatial relationships and, as such, it defines the relationship between features (points, lines and polygons). These relationships are the foundation of spatial analyses and, while their importance cannot be understated, most GIS packages handle topological relationships behind the scenes leaving the user free to perform any analysis. Every GIS is likely to have its own internal means of handling topology, and this section describes only the general principles on which topology is based.[139]

Figure 5.8 shows the topological coding tables relating to the simple "map" at the top of the diagram. In the language of topology, lines are known as "arcs" and points are known as "nodes". The topological information may be stored in three types of "topological tables", in addition to a coordinate table (Figure 5.8). Arcs have a defined start and end node, plus intermediate nodes, and each node has a defined coordinate georeference. The first topology table (polygon topology table) defines the polygons by the arcs that create its boundary. The second table (arc topology) describes which end point nodes occur on each arc and which polygon is to the left or right of each arc. The third topology table stores information about the arcs that end at each of the nodes (node topology table). The fourth arc coordinate data table is not strictly one of the topological tables, but it is an essential associated table allowing the software to know the exact coordinates for drawing all lines or points. Because shared boundaries between polygons are stored as one arc, not the overlapping arcs used in older vector data formats, the storage space needed is reduced. Once the topological relationships have been established for a data set, the GIS package is able to determine all spatial relationships between points, lines and polygons efficiently and to perform an array of functional measurements and analyses. For instance, a set of topological tables containing the attribute and topology data for the river network in Australia would allow the basic river network to be mapped and calculations relating to network lengths could readily be made. If the river network database was then linked to the Australian state boundary information, then the total length of rivers in each state could be calculated. Bernhardsen (1999), Lo and Yeung (2002) and Longley *et al.* (2005b) give more detail on vector data structuring, formats and topology, as do most basic GIS texts, and ESRI offers an interesting background to topology (http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//006200000002000000.htm).

[139] It should be noted that as software evolves, so does the definition of topology. Some modern GIS systems implement topology in new and powerful ways. Topology in the latest versions of ESRI's ArcGIS, for example, no longer store shared boundaries in single arcs. Rather they store the entire vector object for each feature, and they enforce integrity, editing and validation among features within a single data set as well as among associated data sets.

FIGURE 5.8
Simple vector map with associated topological tables



POLYGON TOPOLOGY	
Polygon	Arcs
A	a1, a5, a3
B	a2, a5, 0, a6, 0, a7
C	a7
D	a6
E	Area is outside map

NODE TOPOLOGY	
Node	Arcs
N1	a1, a4, a3
N2	a1, a2, a5
N3	a2, a3, a5
N4	a4
N5	a6
N6	a7

ARC TOPOLOGY				
Arc	Start Node	End Node	Left Polygon	Right Polygon
a1	N1	N2	E	A
a2	N2	N3	E	B
a3	N3	N1	E	A
a4	N4	N1	A	A
a5	N3	N2	A	B
a6	N5	N5	B	B
a7	N6	N6	B	C

ARC COORDINATE DATA			
Arc	Start x, y	Intermediate x, y	End x, y
a1	38, 66	85, 66	85, 53
a2	85, 53	85, 7; 5, 7	5, 23
a3	5, 23	5, 66	38, 66
a4	17, 43	21, 57	38, 66
a5	5, 23	28, 33; 42, 36; 67, 49	85, 53
a6	21, 21		21, 21
a7	72, 14	37, 13; 44, 25; 67, 28	72, 14

The ESRI "Shapefile" format. Since the late 1990s, ESRI has been developing and refining the so-called "shapefile" format for handling vector data, and this has now become a standard data format that can be used by almost all major GIS software.[140] Shapefiles store non-topological geometry and attribute information for all spatial features in a data set and, as such, have no relational intelligence. The geometry for a feature is stored as a "shape" comprising a set of vector coordinates and links to any attribute. Because shapefiles do not have the processing overheads of a topological data structure, as well as lower storage requirements, they have advantages over other data

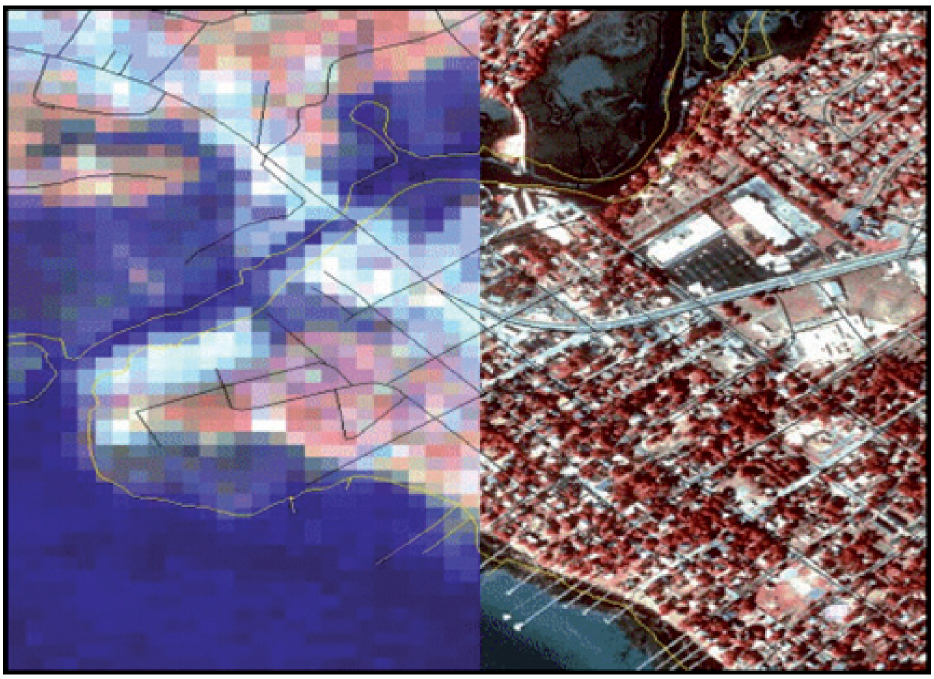
[140] The term "shapefile" is somewhat misleading because, rather than being a single file, it consists of between three and eleven separate files that allow for the representation of 14 different types of geometric shapes.

formats, such as faster drawing speed and edit ability, and the ability to establishrelational intelligence on the fly. Furthermore, the on-the-fly methods determine important relationships that traditional topology fails to capture. Thus, there is often an interest in proximity, direction and distance between disconnected features, and the topology data do not handle this. The same logic that determines direction and distance can also determine all the topological relationships that used to be stored directly in the data. Shapefiles handle single features that overlap or that are non-contiguous. Because shapefiles are compatible with many software programs, they are likely to be a mainstay for GIS software in years to come. For more information on ESRI shapefiles, see ESRI (1998) (available at www.esri.com/library/whitepapers/pdfs/shapefile.pdf) and Theobold (2001) (available at www.esri.com/news/arcuser/0401/topo.html).

5.8.2 Raster data structures

The second method for structuring data used to create digital mapping is through the use of the raster data format. Figure 5.6 shows that the points, lines or polygons representing all spatial entities can also be depicted using a grid of equal size cells (also known as pixels). Grid cells are almost always square, e.g. perhaps representing 1 metre by 1 metre of real world space but, though rarely used, "cells" could theoretically consist of any shape capable of regular tessellation such as rectangles, equilateral triangles or hexagons.[141] Individual cells in a raster structure may be identified by sequential alpha and/or numeric figures for the columns (x axis) and for the rows (y axis). Figure 5.12 illustrates an example of this, with the map showing the same area as was covered in Figure 5.8, and with a similar numerical coding being shown for each axis.[142]

FIGURE 5.9
Examples of raster output for aerial data collected at 30 metre and 1 metre resolutions



Satellite image
(30 metre spatial resolution)

Aerial photograph
(1 metre spatial resolution)

Source: NOAA Coastal Services Center (2012).

[141] Tessellation is the division of space into polygons in such a way that none of the polygons overlap and there are no gaps between them.

[142] Note that on vector georeferencing (see Figure 5.8), the coordinate numbering refers to fixed points along each axis, though with raster alphanumeric coding, the letters or numbers refer to each column or row.

Raster information is built up in "layers" (or separate files), with each file covering a distinct mapping theme (bathymetric depth, land use, water temperature, etc.) Each raster cell is typically allocated a code (or value) based on the predominant feature or class that occupies the majority of the grid cell,[143] or the presence or absence of a feature. For example, elevation rasters contain elevation values, landscape type maps contain landscape category values, and aerial photographs contain reflectance values. The coding given to each cell might be in terms of a numerical value, a weighting, a coding allocated to a colour or a reflectance value. Therefore, rasters do not actually contain the colours illustrated in Figure 5.9; they contain values that the GIS converts to colours when they are displayed.

Some rasters have a single band, or layer (a measure of a single characteristic), of data, while others have multiple bands. Basically, a band is represented by a single matrix of cell values, and a raster with multiple bands contains multiple spatially coincident matrices of cell values representing the same spatial area. An example of a single-band raster data set is a digital elevation model (DEM). Each cell in a DEM contains only one value representing surface elevation. Most satellite images have multiple bands, typically containing reflectance values within a range or band of the electromagnetic spectrum.

There are three main ways to display (render) single-band raster data sets:

- Using two colours. In a binary image, each cell has a value of 0 or 1 and is often displayed using black and white. This type of display is often used for displaying scanned maps with simple line work such as land parcel maps.
- Greyscale. In a greyscale image, each cell has a value from 0 to another number, such as 255 or 65 535. These are often used for black-and-white aerial photographs.
- Colour map. One way to represent colours on an image is with a colour map. A set of values is coded to match a defined set of red, green and blue (RGB) values.

The three main ways to display single-band raster data sets are shown in Figure 5.10 (A).

When there are multiple bands, every cell location has more than one value associated with it, as presented in Figure 5.10 (B). With multiple bands, each band usually represents a segment of the electromagnetic spectrum collected by a satellite sensor. Bands can represent any portion of the electromagnetic spectrum, including ranges not visible to the eye, such as the infrared or ultraviolet sections.

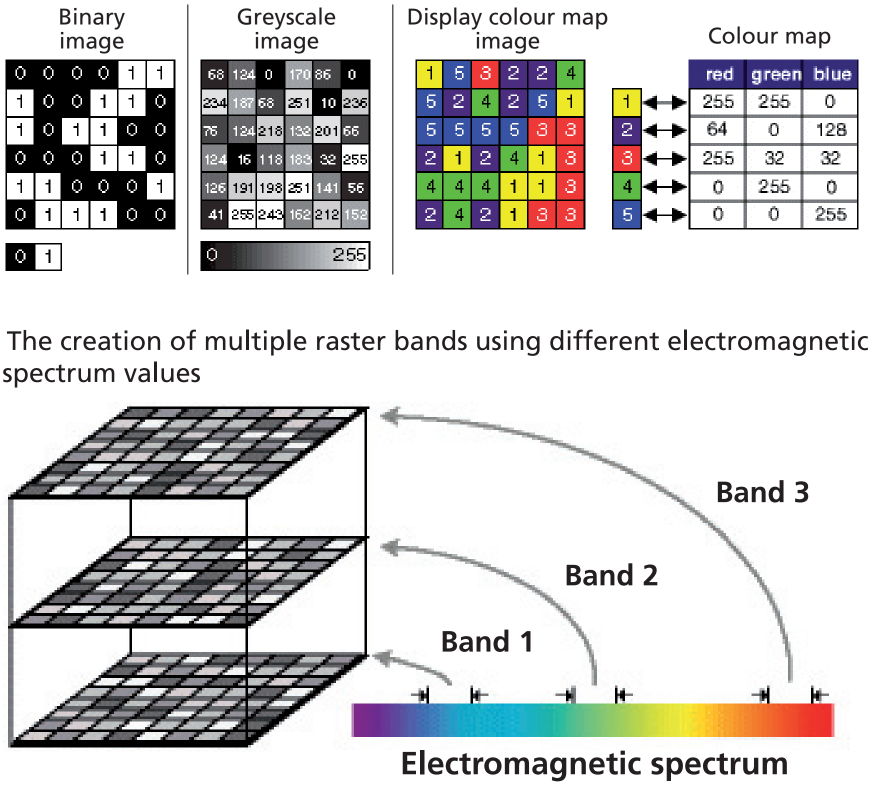
The dimension of the pixels, known as the spatial resolution (not to be confused with cartographic resolution), is of particular importance to the use of the raster structure. Figure 5.11 illustrates the importance of spatial (cell) resolution. Here, a simple 71 m^2 vector-based polygon is compared with its raster equivalent at three different resolutions. The smallest resolution (in this case 1 m cells) corresponds to a fairly good representation of the original, whereas the larger 4 m cells create only a crude replica of the original polygon shape. The figure also lists advantages and disadvantages of using different resolutions, e.g. in this case nearly 16 times as much data must be stored for the smaller cell size resolution.[144] As computer storage capacity has greatly increased relative to storage costs, raster data sets having very small cell (pixel) sizes are increasingly common and they may typically be much larger than vector data sets. In practice, this may mean that many mapped lines are created from rasters whose pixels are so small that the map user may not be able to discriminate the line from a vector-drawn line. Figure 5.9 illustrates an actual example showing raster output at two widely different resolutions; in the left hand image, there is a huge loss of feature clarity causing features to become pixilated in appearance at the increased cell size. Thus, when users collect or create data for a specific project, they must decide the scale at which to work and the cell size that will provide adequate information for successful project completion. Most GIS packages

143 Sometimes an allocated cell value may be that pertaining at the central point of the cell or at a cell corner.

144 It is worth recording that one complete Landsat Thematic Mapper image (see Chapter 6) may contain about 35 million pixels – all of whose values must be stored.

FIGURE 5.10
Types of raster bands

(A) Three main ways to display single-band raster data sets



Source: ESRI (2011c).

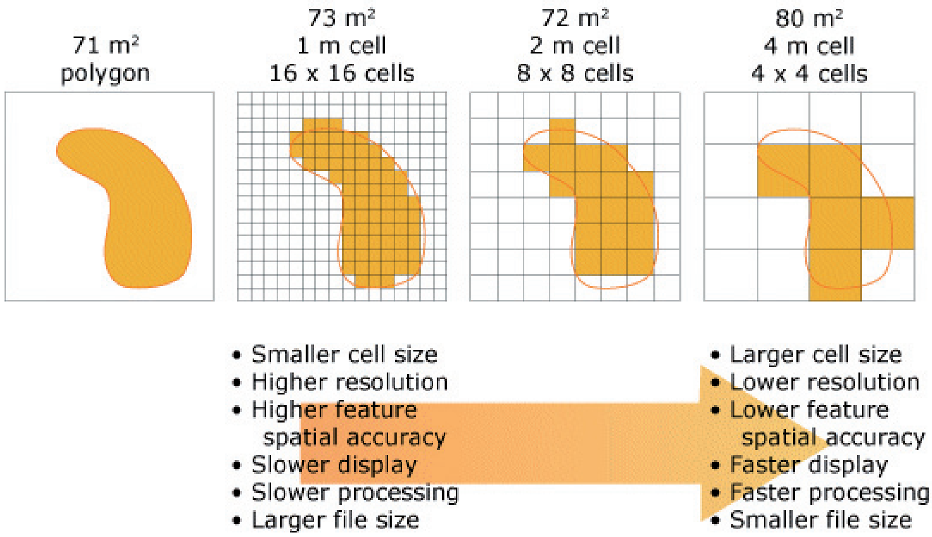
provide the ability to resample raster data[145] so that the data can be changed to a suitable cell size, but it should be noted that any new raster data being created can be of no greater accuracy than that of the original data, e.g. it would make no sense to use the 30 m^2 pixels in Figure 5.9 as the basis for creating the 1 m^2 cell information.

While rasters with small cell sizes tend to take longer to display, a GIS can often accelerate the process with the use of "pyramids". Pyramids are simply lower-resolution versions of the data used only for display purposes. If a large raster image might take minutes to draw on the screen, a pyramid version will draw it in seconds and the appearance will be indistinguishable to the viewer. Pyramids do not speed up raster analysis, but they can make viewing rasters much easier. Detailed information on the use of pyramids can be found at ArcGIS Resource Center (http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//009t00000019000000) or Lowe (2004).

Rasters can require large amounts of storage space, and higher resolutions require exponentially more storage space than lower resolutions. Because of this, a variety of data compression techniques have been developed to reduce the raster file size. There are two broad categories of data compression methods: "lossless" methods preserve all the original data, and are preferred when the data are intended for precise analysis. Popular examples are: LZ77 (Lempel-Ziv 77); LZW (Lempel-Ziv-Welch); RLE (run-length encoding); and quadtree compression. "Lossy" methods alter the data and are generally

[145] Resampling raster data essentially means that various facets relating to the captured data can be changed, e.g. brightness of the image, in order to better discriminate between objects on the ground. For more details, see ArcGIS Resource Center (http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//009t00000082000000.htm).

FIGURE 5.11
Demonstration of different pixel resolutions



Source: ESRI (2007).

used for mapping or aesthetic purposes. The most popular example of this type is JPEG (Joint Photographic Experts Group). A recent advancement is JPEG2000, which uses complex wavelet transform mathematical techniques to produce highly compressed images, and which can be either lossless or lossy depending on the user preference[146].

This section explains how two of the simpler compression methods – RLE and quadtree compression – work. First, Figure 5.12 shows how RLE has been applied to the raster data set displayed.[147] Because many adjacent cells share the same value, for each row it is only necessary to specify a cell value and column number where that value begins and ends. There are several variations of RLE. Further details on raster data compression techniques can be found at Geographic Information Technology Training Alliance (www.gitta.info/DataCompress/en/html/rastercomp_chain.html), or Heywood, Cornelius and Carver (2006).

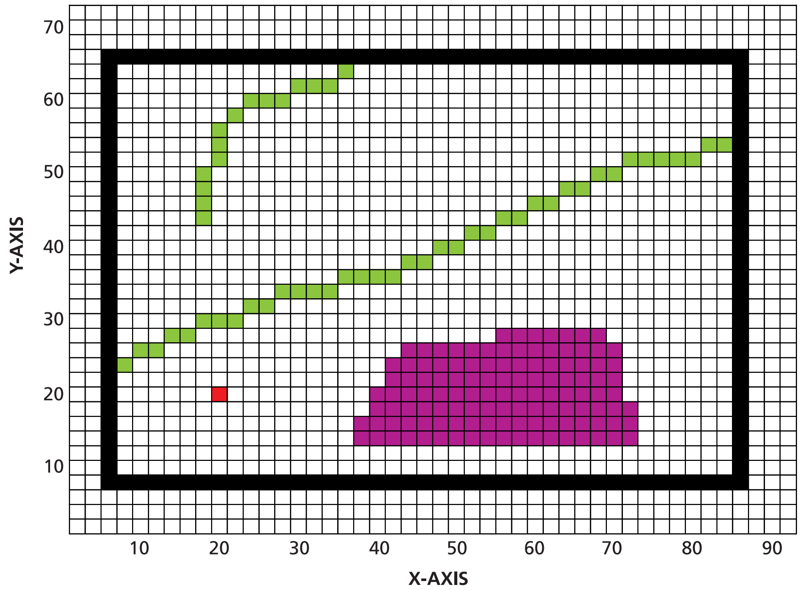
Second, the quadtree technique is based on the successive subdivision of the area under study into smaller and smaller quadrants, i.e. according to whether or not areas of a similar value are wholly in the quadrant or are not. The lowest level of subdivision is a single pixel. Figure 5.13 shows a raster structure[148] created to map areas having three different "values" (A), with areas being successively subdivided into the smallest quadrats containing a single value. In (B), a quadtree has been constructed, which shows that there are links and nodes in the tree starting from the root node (represented by the blank circle at the top of the tree, i.e. the whole mapped area).

From each node, there are four so-called "edges" representing the four quadrants of northwest (NW), northeast (NE), southwest (SW) and southeast (SE). Subquadrants only emanate from nodes that are shown as being subdivided in the raster map, i.e. it can be seen from (B) that three of the main quadrats (NW, SW and SE) are further subdivided, but that the NE quadrat (numbered 8) is not further subdivided, so it ends at the second level down. For the SE quadrat, no further subdivision occurs after the third level, so the tree ends here (at cell 16). On the other hand, quadrats

[146] For more detailed information on "lossless" and "lossy" data compression, see ArcGIS Resource Center (http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//009t00000021000000).
[147] For clarity, only alternate pixels have been drawn and numbered in Figure 5.12.
[148] Only the main raster cells are depicted in this figure.

FIGURE 5.12

A simple raster map plus the run-length encoding structure used for data storage



RASTER-RUN LENGTH CODE STRUCTURE

Row	Run-length encoding
72	0, 92, 0
70	0, 92, 0
68	0, 92, 0
66	0, 4, 0; 6, 86, 1; 88, 92, 0
64	0, 4, 0; 6, 1; 8, 34, 0; 36, 2; 38, 84, 0; 86, 1; 88, 92, 0
62	0, 4, 0; 6, 1; 8, 28, 0; 30, 34, 2; 36, 84, 0; 86, 1; 88, 92, 0
60	0, 4, 0; 6, 1; 8, 22, 0; 24, 28, 2; 30, 84, 0; 86, 1; 88, 92, 0
58	0, 4, 0; 6, 1; 8, 20, 0; 22, 2; 24, 84, 0; 86, 1; 88, 92, 0
56	0, 4, 0; 6, 1; 8, 18, 0; 20, 2; 22, 84, 0; 86, 1; 88, 92, 0
54	0, 4, 0; 6, 1; 8, 18, 0; 20, 2; 22, 80, 0; 82, 84, 2; 86, 1; 88, 92, 0
52	0, 4, 0; 6, 1; 8, 18, 0; 20, 2; 22, 70, 0; 72, 80, 2; 82, 84, 0; 86, 1; 88, 92, 0
50	0, 4, 0; 6, 1; 8, 16, 0; 18, 1; 20, 66, 0; 68, 70, 2; 72, 84, 0; 86, 1; 88, 92, 0
48	0, 4, 0; 6, 1; 8, 16, 0; 18, 1; 20, 62, 0; 64, 66, 2; 68, 84, 0; 86, 1; 88, 92, 0
46	0, 4, 0; 6, 1; 8, 16, 0; 18, 1; 20, 58, 0; 60, 62, 2; 64, 84, 0; 86, 1; 88, 92, 0
44	0, 4, 0; 6, 1; 8, 16, 0; 18, 1; 20, 54, 0; 56, 58, 2; 60, 84, 0; 86, 1; 88, 92, 0
42	0, 4, 0; 6, 1; 8, 50, 0; 52, 54, 2; 56, 84, 0; 86, 1; 88, 92, 0
40	0, 4, 0; 6, 1; 8, 46, 0; 48, 50, 2; 52, 84, 0; 86, 1; 88, 92, 0
38	0, 4, 0; 6, 1; 8, 42, 0; 44, 46, 2; 48, 84, 0; 86, 1; 88, 92, 0
36	0, 4, 0; 6, 1; 8, 34, 0; 36, 42, 2; 44, 84, 0; 86, 1; 88, 92, 0
34	0, 4, 0; 6, 1; 8, 26, 0; 28, 34, 2; 36, 84, 0; 86, 1; 88, 92, 0
32	0, 4, 0; 6, 1; 8, 22, 0; 24, 26, 2; 28, 84, 0; 86, 1; 88, 92, 0
30	0, 4, 0, 6, 1; 8, 16, 0; 18, 22, 2; 24, 84, 0; 86, 1; 88, 92, 0
28	0, 4, 0; 6, 1; 8, 12, 0; 14, 16, 2; 18, 54, 0; 56, 68, 3; 70, 84, 0; 86, 1; 88,92, 0
26	0, 4, 0; 6, 1; 8, 0; 10, 12, 2; 14, 42, 0; 44, 70, 3; 72, 84, 0; 86, 1; 88, 92, 0
24	0, 4, 0; 6, 1; 8, 2; 10, 40, 0; 42, 70, 3; 72, 84, 0; 86, 1; 88, 92, 0
22	0, 4, 0; 6, 1; 8, 40, 0; 42, 70, 3; 72, 84, 0; 86, 1; 88, 92, 0
20	0, 4, 0; 6, 1; 8, 18, 0; 20, 4; 22, 38, 0; 40, 70, 3; 72, 84, 0; 86, 1; 88, 92, 0
18	0, 4, 0; 6, 1; 8, 38, 0; 40, 72, 3; 74, 84, 0; 86, 1; 88, 92, 0
16	0, 4, 0; 6, 1; 8, 36, 0; 38, 72, 3; 74, 84, 0; 86, 1; 88, 92, 0
14	0, 4, 0; 6, 1; 8, 36, 0; 38, 72, 3; 74, 84, 0; 86, 1; 88, 92, 0
12	0, 4, 0; 6, 1; 8, 84, 0; 86, 1; 88, 92, 0
10	0, 4, 0; 6, 1; 8, 84, 0; 86, 1; 88, 92, 0
8	0, 4, 0; 6, 86, 1; 88, 92, 0
6	0, 92, 0
4	0, 92, 0
2	0, 92, 0

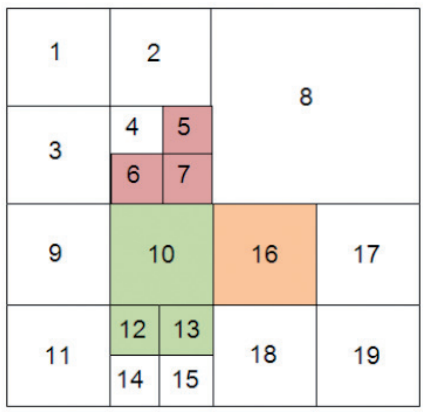
Source: Adapted from Meaden and Do Chi (1996).

NW and SW can be further subdivided to another level. Where edges can be further subdivided, an open (white) circle (node) is drawn; where cells cannot be further divided but there is none of the mapped (white area) showing, a blank square is shown; and where cells cannot be further subdivided and there is mapped area showing, a colour-shaded square is shown. The software encodes this map by only recording georeferencing data for the coloured squares. Savings of at least 50 percent in data volumes are possible using quadtrees. Longley *et al.* (2005b) note that quadtrees offer various advantages for GIS-based work over other methods of data compression.

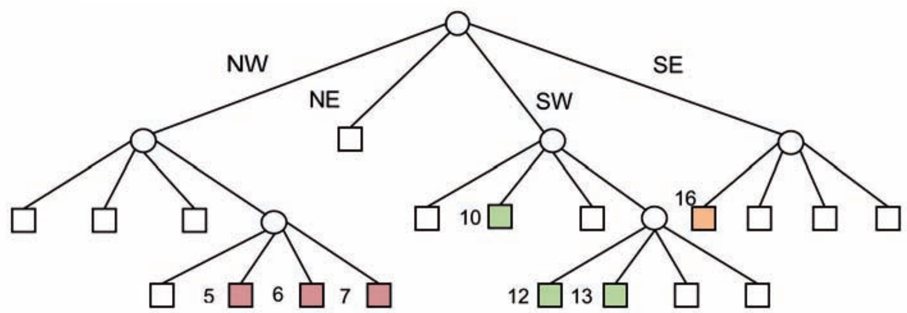
In the past, GIS packages might have been designed to work with one or the other type of data, raster or vector. Now most GIS packages can integrate both raster and vector using additional "add-on" software or extensions. Even though the raster and vector structures are both valid ways of representing real world features, preferences for either can come down to the user's perceived preference of one format over another for use in specific applications. The advantages and disadvantages of each data format have been extensively discussed and Box 5.6 briefly lists these qualities.

FIGURE 5.13
The quadtree raster data compression method

(A) Raster map



(B) The quadtree structure



BOX 5.6
Comparison of raster and vector data structures

Raster data format

Advantages

- simple data structure;
- numerous data sets have been generated through remote sensing and scanning technologies;
- spatial analysis procedures are simple (overlays, area analysis, merging, etc.) and are often faster;
- the inherent nature of raster maps is ideally suited for mathematical modelling and quantitative analysis.

Disadvantages

- large data storage requirements;
- topological relationships can be difficult to represent;
- without very small cell size, detail can be lost;
- map products can be relatively crude if cell size (resolution) is large;
- linear analysis is more difficult and less accurate.

Vector data format

Advantages

- compact data structure takes up less storage space;
- features can be accurately located;
- topological relationships can be established;
- very small features can be shown and all features can be accurately drawn;
- data about individual features can be easily retrieved and updated;
- linear analysis is easily performed;
- map products can be more aesthetically pleasing resembling hand-drawn analogue maps.

Disadvantages

- the data structure is more complex;
- overlay procedures can be more time consuming;
- data capture can be slow and expensive;
- some spatial analysis procedures can be difficult (e.g. area analysis);
- it is not compatible with remotely sensed data.

5.8.3 Terrain (surface) data models

The mapping of terrain is a common task for cartographers, and various techniques have been devised to facilitate this process. Most of these techniques can be easily transferred to a submarine environment where there is a varied interest in reproducing bathymetry[149] and bathymetric modelling. The breadth of interest in, and uses for, bathymetric modelling using GIS is provided in Breman (2010). It should be mentioned that the mapping of terrain or bathymetry is considered as 2.5D (two-and-a-half dimensional) mapping. This means that the vertical dimension of the ground surface, seafloor or lake-bed area being mapped, measured or modelled is only considered in terms of the altitude and/or depth of the "surface" level above or below a datum line,[150] i.e. no consideration is made of objects that are either in the air or in the waterbody. The datum used may be a locally agreed mean tidal level, or it is increasingly measured

[149] Bathymetry is the measurement of the depth of oceans, seas or other large bodies of water.

[150] Most countries that have a marine coast line will establish a "datum line". This is typically an accurately measured and demarcated small section of the coast lying midway between and high and low water tidal marks. Once this line has been designated, it serves as a baseline above (or below), which all altitudinal (or bathymetric) heights (or depths) are measured.

using an internationally agreed datum such as that based on WGS 84.[151] True 3D modelling is also possible using GIS, but this modelling is not discussed here because it has rarely been used for fisheries or aquaculture purposes,[152] i.e. it is mainly used for geological purposes or for volumetric calculations.[153] Data for bathymetric mapping and modelling are commonly gathered from field surveys (mostly using GPS), archived maps containing depth soundings, various photogrammetric techniques[154] or, more recently, from light detection and ranging (LIDAR) and acoustic single or multibeam echosounding surveys (Kearns and Breman, 2010). Terrain data modelling is usually called "digital terrain modelling" (DTM), and within this modelling there are two common data formats: (i) a vector format that is used to create triangulated irregular networks (TINs); and (ii) a raster format to create digital elevation models (DEMs).[155] Most modern GIS packages are capable of viewing and analysing terrain models. DTM can be used for a variety of GIS-based analyses, as follows:

- drawing of block diagrams;
- estimations of water volumes or materials in civil construction projects;
- interpolation and contour mapping;
- line of sight (or intervisibility) mapping;
- calculations and mapping related to aspect, hill shading or slope (gradient);
- catchment estimation;
- defining river networks;
- hydrologic modelling of water flow potential.

Some of these analyses procedures are explained in Chapter 7, where GIS functionality based on TINs and DEMs are examined.

Additional detail on terrain modelling can be obtained from Wilson and Gallant (2000) or Lo and Yeung (2002), who also outline a range of stand-alone terrain analysis and contouring packages, and there is now a searchable DEM database being developed for tsunami inundations and described in the National Geophysical Data Center (www.ngdc.noaa.gov/mgg/inundation). Other sources of DEMs include: ASTER Global Digital Elevation Model, which provides free DEM data to various categories of user (see ASTER Global Digital Elevation Model: www.ersdac.or.jp/GDEM/E/3.html); GEBCO global bathymetry data for oceanic areas (see GEBCO: www.gebco.net/data_and_products/gridded_bathymetry_data); or SRTM (Shuttle Radar Topography Mission elevation data (see USGS: http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/Elevation_Products).

Triangulated irregular networks

In its simplest form, raster-based DTMs or DEMs record spot height values on a regular grid pattern, and typically the values would be recorded at each grid intersection (Figure 5.14). However, the use of regular grids often causes a loss of detail because actual data on highest or lowest points are often omitted and thus only a generalized terrain surface can be created. A better quality terrain model is created through the use of vector-based triangulated irregular networks (TINs). The TIN approach to

---

151 World Geodetic System 1984 (WGS 84) is a model of the earth's surface based on an oblate spheroid with an equatorial radius of 6 378 137 m and a polar radius of 6 356 752 m, with an origin approximately at the centre of the earth's mass and a zero-longitude set at the International Earth Rotation and Reference Systems Service (IERS) reference meridian. This datum is designed for positioning anywhere on earth.

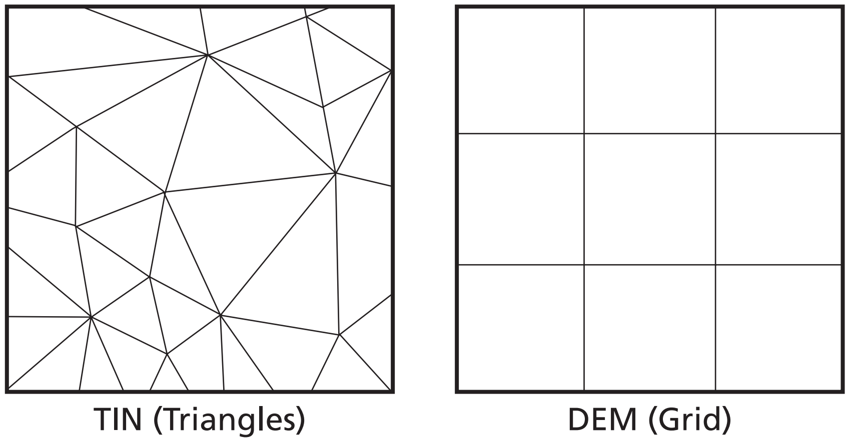
152 One of the few studies on 3D is a study by Moreno Navas (2010) to explore the use of a 3D-hydrodynamic model coupled to a particle tracking model to study the circulation patterns, dispersion processes and residence time in the Irish loch for marine cage culture.

153 Note that 2.5D modelling can only have one "z" (vertical) value per x,y location, whereas 3D modelling can utilize multiple z values per x,y location.

154 Photogrammetric techniques are the processes of making maps or scale drawings from photographs, especially aerial photographs and satellite remotely sensed images.

155 Sometimes there is confusion with the terms DEMs and DTMs. A DEM is a DTM, but a DTM is not necessarily a DEM.

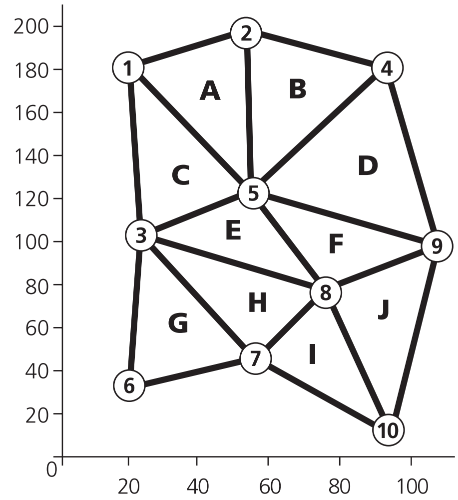
FIGURE 5.14
Vector-based approaches to digital terrain modelling



digital terrain modelling relies on the recording of height or depth values at sampled (surveyed) georeferenced points across an area to be mapped. Here, more purposefully placed height or depth data are recorded in a non-regular way, e.g. in very flat areas only a few values need to be recorded, but areas with rugged topography require frequent height or depth values. The TIN method joins the height observations at the nodes by straight lines to create a mosaic of irregular triangles (Figure 5.14).

Each triangle in a TIN defines a relatively flat surface covering a small part of the total topographic surface. The vertices (or nodes) of the triangles store x, y and z information, i.e. location and elevation values. Given this nodal information, the surfaces of each triangle can provide information on length of slope, area, slope gradient and orientation (aspect), and this information can be stored as attribute data for each triangle. Each GIS package may have its own algorithm for determining the basis for how the triangles are apportioned, i.e. what nodes are chosen for each link (edge).

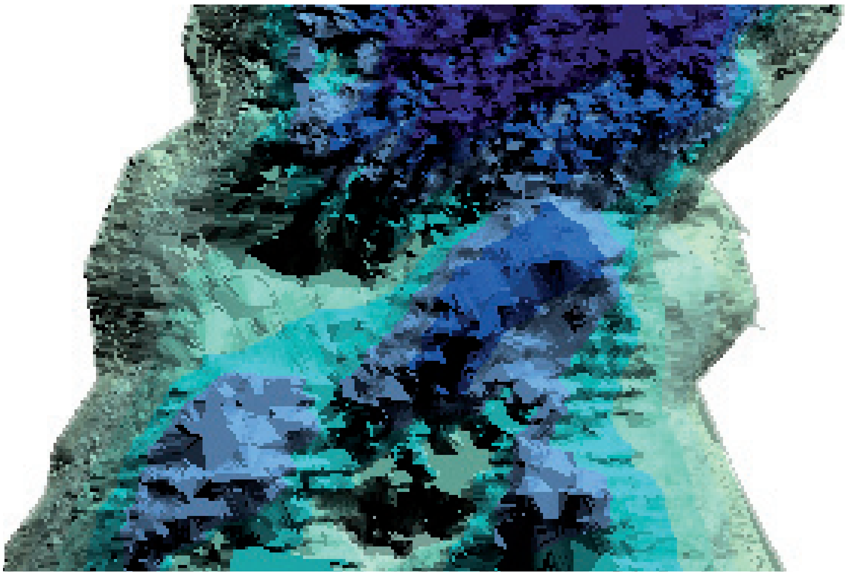
FIGURE 5.15
The topology of a TIN



Node	x - y coordinate	z elevation
1	21, 180	89
2	58, 198	91
3	23, 103	69
4	96, 180	83
5	58, 122	69
6	20, 37	54
7	58, 46	52
8	77, 78	58
9	110, 99	60
10	93, 13	46

Triangle	Node list	Neighbours
A	1, 2, 5	-, B, C
B	2, 4, 5	-, D, A
C	1, 5, 3	-, A, E
D	4, 9, 5	-, F, B
E	3, 5, 8	C, F, H
F	5, 9, 8	D, J, E
G	3, 7, 6	-, H, -
H	3, 8, 7	E, I, G
I	8, 10, 7	H, J, -
J	8, 9, 10	-, I, F

FIGURE 5.16
Triangulated irregular network showing bathymetric data for Lake Michigan, United States of America



Source: University of Michigan (2012).

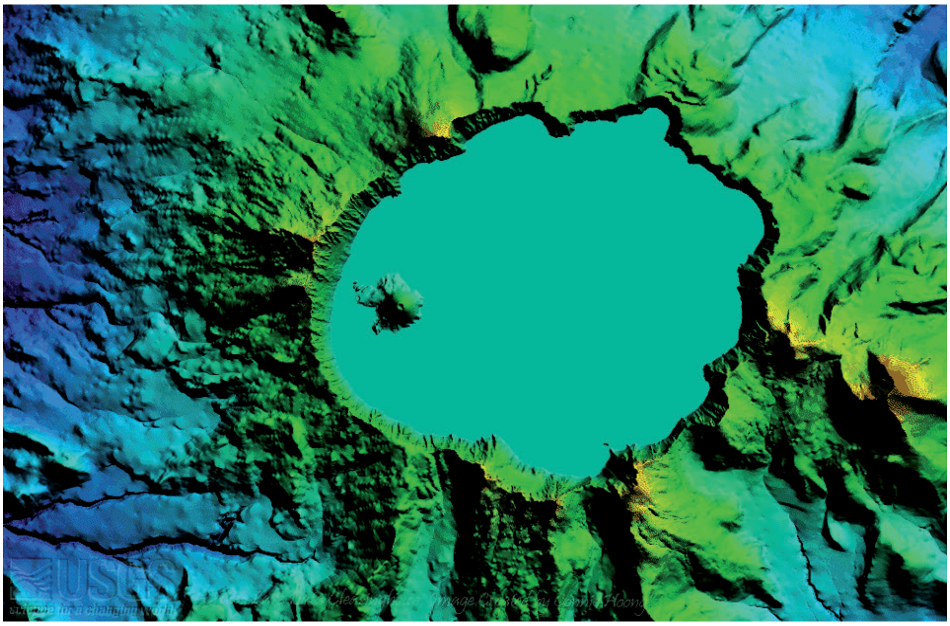
TINs use the topological model vector-based data storage methods as described in Section 5.8.1 and shown in Figure 5.15. The TIN data structure provides information that allows a GIS to create contour maps (using interpolation techniques), including maps showing gradient, aspect or hill shading. Figure 5.16 shows that sophisticated submarine bathymetry can be created based on depth soundings; in this case also showing "depth" shading (deepest areas in dark blue) and bottom gradients across a central portion of Lake Michigan, United States of America.[156]

Digital elevation models

Digital terrain modelling using the raster data format produces digital elevation models (DEMs). It is clear that the raster data structure described in Section 5.8.2 can be utilized to record a single height or depth (z value) for each cell (pixel). The z value can be the height or depth pertaining to the centre of the cell or to a given cell corner. As with other raster-based work, it is clear that the cell resolution will greatly affect the accuracy of the GIS output. Because height or depth data may not be available for every cell, it is usual to use the interpolation functions that are provided in most GIS packages to calculate missing pixel values (see Section 7.5.3). Raster-based DEMs provide a simple structure but one that allows efficient modelling with low computational overheads (fewer calculations) plus easier storage. Figure 5.17 illustrates how a DEM produced map of Crater Lake in Oregon, the United States of America, has been enhanced by altitude and relief shading. The data sets behind DEMs are often available from national mapping agencies at different resolutions, though marine or lacustrine (lake) bathymetric data are unlikely to be widely available. Most GIS packages support raster DEMs and provide for the conversion of x, y and z values to the raster grid cell structure. There are relative advantages offered by each of the two terrain models described and these are summarized in Box 5.7, though it should be mentioned that the relative resolutions adopted can greatly influence the quality of the GIS terrain modelling and, contrary to the apparent evidence in Box 5.7, Longley *et al.* (2011) argue that DEMs are the most useful and accurate form of terrain modelling.

[156] The apparent "flat" areas shaded in grey are in fact islands protruding above the lake's surface.

FIGURE 5.17
Digital elevation model of Crater Lake, Oregon, United States of America enhanced by altitude and relief shading



Source: Rob's maps (2009).

5.8.4 Network data models

A network is an interconnected set of points, lines and polygons that represent possible routes (or "pathways") from one location to another. Although a network is commonly perceived as a human routeway in terms of various forms of transport (airplanes, shipping, vehicles, pedestrians, etc.), it can also represent other built features such as pipelines and cabling routes, plus a range of designated boundaries, e.g. county or zip code boundaries. Besides human features, networks are also represented by natural features such as waterways, vegetation corridors and migration routes. Given this huge array of networks, their modelling is an important task for any GIS. Network data models are usually enacted using the vector data structure because this is more readily associated with the points, lines and polygons that make up networks, although some types of network analysis (least-cost path models in particular), lend themselves better to raster formats. Here, the focus is mainly on vector models.

For GIS functionality, all network data models envisage that networks are comprised of links (arcs or edges) and nodes (points) (Figure 5.18). Links represent the linear element of a network and nodes represent junctions, starting or ending places, stopping places, switching stations, confluences, centres, etc., depending on the type of network being modelled.[157] There are two broad classes of network models[158]:

(i) Looped (or circuit) networks. These are so-called "undirected" models, meaning that movement through the network can proceed in either direction, e.g. most transport routes.

(ii) Geometric (also called radial, tree, directed, oriented or unidirectional) networks. Here, movement can only be in one direction, e.g. one-way streets, electrical cables, storm drainage or rivers.

---

[157] There are an almost infinite number of potential nodes, and these will largely depend on the purpose and/or scale of the mapping or modelling process.

[158] To add some confusion to a complex topic, many references (especially in the field of graph theory) refer to the undirected "looped" models as "directed networks", and to "geometric models" as "oriented networks". These sources also refer to "looped" networks as a special case where a node is only connected to itself without going through any other nodes.

BOX 5.7
Relative advantages of TIN and DEM surface models

Advantages of TINs

- Linear features are more accurately represented.
- Aspect and gradient can be more accurately produced.
- Fewer points are needed to represent topography; therefore, less storage space is required.
- Elevation nodes can be concentrated in areas where topography is more variable or where more detail is required.
- Survey data and known elevations can easily be incorporated into the TIN.
- Some functions cannot be performed using DEMs but are easily accomplished with TINs, for example:
  - creating storage-capacity curves for reservoirs;
  - floodplain delineation.
- TINs can be draped with imagery such as remotely sensed data.
- Field surveying and map digitizing is better suited to producing data for use in TINs.
- TINs can readily be converted to DEMs.
- *Z* values can more accurately be interpolated.
- Cross-sections of a terrain can more accurately be portrayed.
- Choosing an appropriate resolution is incorporated into the methodology.

Advantages of DEMs

- Most analyses are more easily performed.
- Provides a very simple model.
- DEMs are computationally efficient.
- Provides uniform data coverage over an area.
- Can provide very detailed elevational data (with small cell size).
- The raster data format is more readily compatible with other data sets.

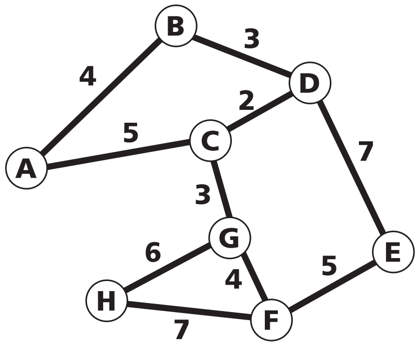
Each of these network models rely upon different algorithms for their functionality. As with some other vector-based models, network data are held as topological tables with every link and node having a unique identifier. For network analysis to function correctly, it is vital for the digital network to be an exact topological representation of the real world network, i.e. meaning that each line segment is "aware" of the lines it is attached to at a specified node. This is far more important than the fact of the network being cartographically correct, i.e. with features being located on the map in the correct locations.[159] Figure 5.18 shows a hypothetical undirected model for a transport routeway, and Figure 5.19 shows a real undirected network model – that of the proposed high-speed rail network for the Kingdom of Spain. The nodes in Figure 5.18 are connected by links whose values show "impedance", e.g. in this case impedance might be showing travel times. So, impedance is a value demonstrating a "cost" (or weighting) of getting from or to adjacent nodes, and the cost can be in terms of fares, fuel, or time taken, etc. Factors that influence impedance include the time of day, whether there are road repairs, how many traffic lights, maximum speed limits, to name a few. This type of undirected model can be used for tasks such as calculating the shortest route between two nodes, working out the quickest route, establishing a route that is the shortest one going through specified nodes,[160] and showing all parts of the mapped area that are within x kilometres or y travelling time of a specified node. It is easy to see how network analysis could be used to optimize the railways for new high-speed trains in the Kingdom of Spain when the whole system is needed. Section 7.5 gives examples of different analyses that may be performed using the network data modelling structure.

---

[159] Many of the world's underground railway maps provide examples of where the topological connections and the lines are all correct but their geo-location on the map may not be correct.

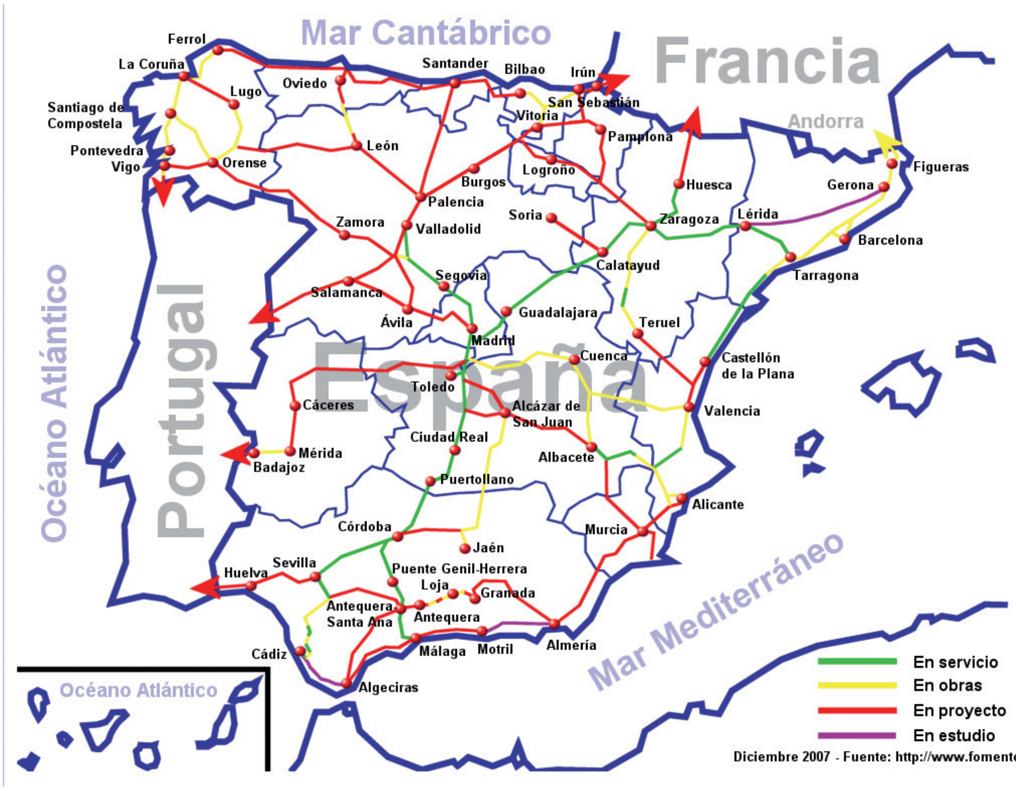
[160] This is known as the "travelling salesman" problem.

FIGURE 5.18
Network data structure showing nodes and impedance plus topological table



Node ID	Connected nodes (+ impedance weighting)
A	B(4), C(5)
B	D(3), A(4)
C	D(2), G(3), A(5)
D	E(7), C(2), B(3)
E	D(7), F(5)
F	H(7), G(4), E(5)
G	C(3), F(4), H(6)
H	G(6), F(7)

FIGURE 5.19
The proposed network of high-speed railway lines in Spain (2007)



Source: Sánchez de León (2007).

Economic feasibility is an indispensable criterion for aquaculture development. In a strategic assessment study on the potential for freshwater fish farming in Latin America, Kapetsky and Nath (1997) conducted an analysis for market potential to calculate the least-cost path from any cell location to the closest town in real distances along the roads. Because oil prices have been rising worldwide, at a rate far in excess of general inflation rates, the transport-related costs become ever more important as a location decision factor, and it is likely that most aquaculture facilities will need to include network analyses as a fundamental input to their location decision.