Field Techniques Manual:
GIS, GPS and Remote Sensing

- <u>Section A: Introduction</u>

Chapter 3: Geographical Information Systems (GIS)

3 Geographical Information Systems (GIS)

3.1 Introduction

The Chorley Report for the UK Department of the Environment (1987) defined GIS as:

> *"…a system for capturing, storing, checking, integrating, manipulating, analysing and displaying data which are spatially referenced to the Earth."*

The primary type of GIS output is usually a map of some sort, together with a statistical analysis of the features of interest. To obtain the optimal return on time invested in a GIS, new users should examine how integration and analyses were achieved in other GIS applications. This should be followed by reflection on how GIS can be used within a given project.

The development of GIS can be traced back to three types of computer application, developed over the 1960s: computer-aided design (CAD), automated mapping and facilities management. CAD systems are key features of architectural and engineering design; however they are limited to graphics data, rather than map-based information. Automated mapping systems link external files to electronic maps: they are of particular use for mapping and managing utilities and complex facilities, such as airports and hospitals. The Canadian Geographic Information System (CGIS) is widely regarded as the first fully-functional GIS, developed over 1964-1971, to map and analyse land inventory data. CGIS demonstrated how GIS could produce huge savings in time, with the automation of map outputs, digital storage and retrieval, and computer-based analysis of map-linked databases. A comparative analysis of the 6,000 maps in the CGIS archive would have taken 1650 years to carry out manually, but only took a few days using CGIS. As computers have become faster and better at both processing and storing large volumes of data, so GIS software has evolved from slow-running and complex systems, to user-friendly 'desk-top mapping' systems with many users. For more detailed coverage of GIS, the reader is referred to books by Burrough and McDonnell (1998), DeMers (2000) and Heywood *et al.* (1998). Wadsworth and Treweek (1999) focus on GIS for ecologists.

3.2 Data input and organisation

Data organisation and structure is key to successful GIS implementation. Geographical data are available at widely differing scales, so the first decision a user has to make is, *at what scale does the project data need to be captured?* Appropriate scale is determined by various factors including the size of the project area, the size of the features to be analysed and the scale of available data, such as topographic maps.

Data can be input into a GIS by various methods, notably scanning, digitizing and databases, as well as direct importing of pre-existing digital data, such as satellite imagery or commercially produced digital maps. The type of data input will determine which data model is the most appropriate for a particular project. Three data models are widely available: *raster, vector* and *hybrid*. The initial decision on which to use can have profound and far-reaching consequences to a project if the data model is not suitable.

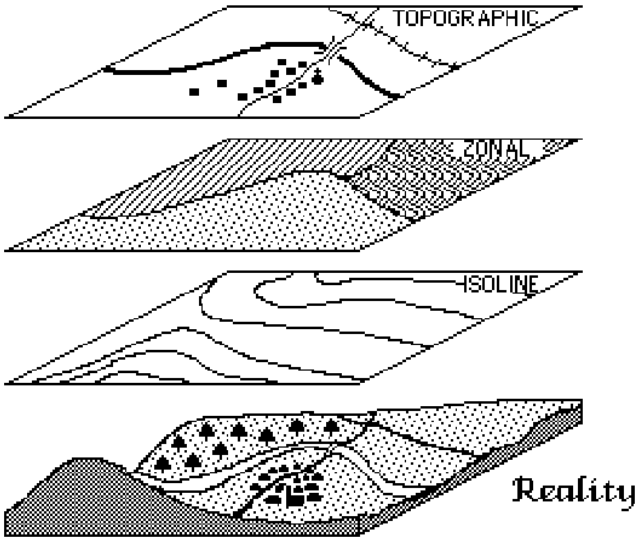


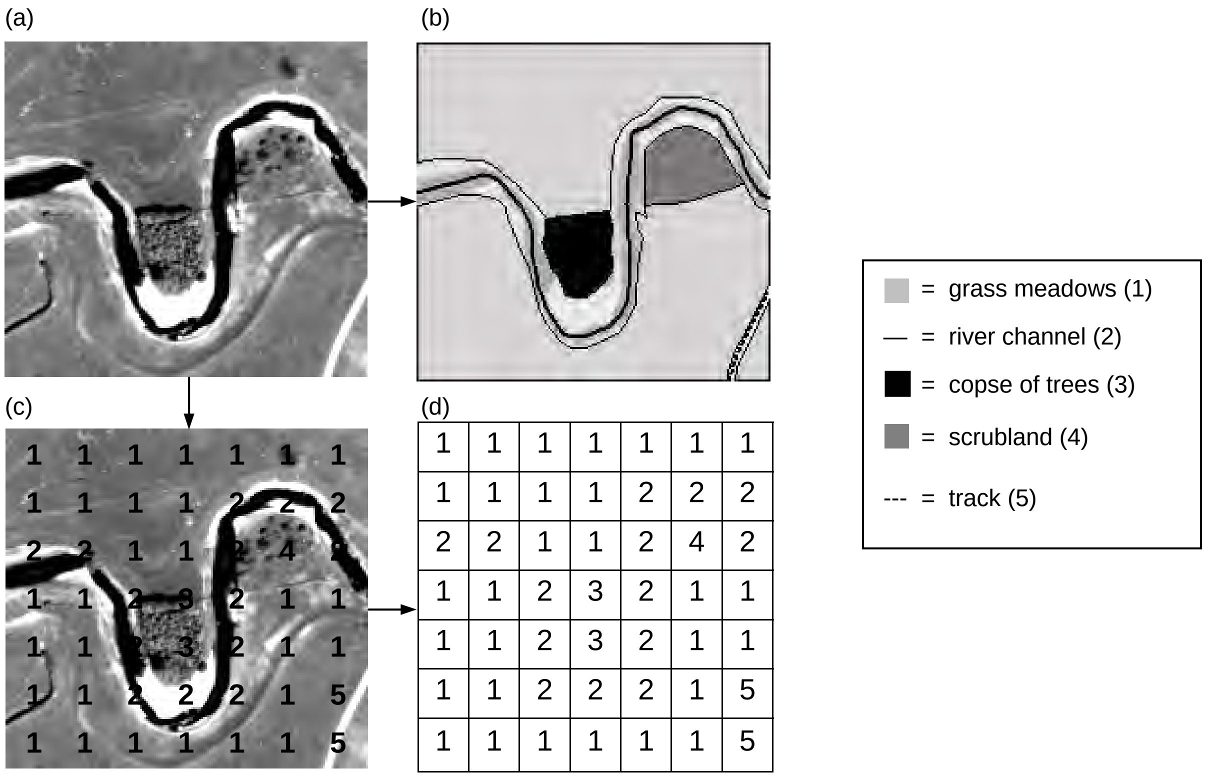
Figure 3-1 An example of GIS data layers (source: Jonathan Raper).

3.2.1 Raster data

The raster model uses a large number of (usually) square 'cells' to represent a geographical surface. These cells are also referred to as *pixels* (picture elements). Each pixel covers a set proportion of the Earth's surface, the size of which is determined by the operator when creating the dataset, or by the data input system (e.g. the spatial resolution of a satellite scanner). Each cell will contain a single value: this could represent surface reflectance (e.g. of a given Landsat band); class data (1 = vegetation, 2 = urban, etc); the average elevation of the cell (forming a digital elevation model); or any other feature of the Earth's surface within that cell. Multiple layers can be used to represent different features, classes, concentrations, etc. Analysis can then take place using arithmetic and mathematical manipulation of the numbers in the various layers and cells.

3.2.2 Vector data

The vector model represents features of the Earth's surface using points, lines and polygons. Points are features that can be represented by a single co-ordinate pair. Points have no length or area, position being their only parameter. Lines can represent linear features such as roads and footpaths that can be represented by a series of connected points. Lines have position and length, but no area. Polygons are features that can be represented by closed lines where the line start and end points are at the same position. Information on the feature is associated with the enclosed area, polygons having position, length (perimeter) and area parameters. Most data types can be represented by a single category, although there may be exceptions to this: elevation data can sometimes consist of contours (lines) and spot heights (points), requiring a combination to fully represent the surface. As with the raster system, multiple layers can be used to build up a map of the Earth's surface and the specific features of interest, furthermore, analysis between the layers can be undertaken. Each created feature can have a significant amount of information attached to it and will have an entry in a data table, called an attribute table. This approach to storing data makes analysis a complex, but powerful, process and also allows easy data retrieval for individual or multiple features.



1	1	1	1	1	1	1
1	1	1	1	2	2	2
2	2	1	1	2	4	2
1	1	2	3	2	1	1
1	1	2	3	2	1	1
1	1	2	2	2	1	5
1	1	1	1	1	1	5

Figure 3-2 Vector and raster data models. Figure (a) is a subset of an aerial photograph showing river meanders. Figure (b) illustrates how this area could be represented as a vector layer. Figures (c) and (d) illustrate how this area could be represented as a raster layer.

3.2.3 Hybrid systems

The hybrid model combines both raster and vector models and allows relatively easy transfer of data, and even combined analysis, in the two formats. The advantage with using a hybrid model is that all potentially useful data can be utilised easily, regardless of the model in which it was encoded. For larger projects with ambitious goals this is invaluable. The disadvantage for the smaller project is the cost of the software and the complexity of the system. Hybrid systems are usually at the higher end of the cost scale for GIS software and are inherently more complex than single model systems.

Once the data types and their respective models have been ascertained, data input can be undertaken. Most current GIS software can import data from various existing sources, and all will have features allowing data to be manually entered. Datasets in a GIS are configured in layers, usually containing a single feature-type, perhaps divided into a number of classes. The feature-type can be broad or narrow, depending on its intended use. A land use layer, for example, may typically contain thirty classes, e.g. from a suburban class to an upland moorland class; whereas a water layer may only contain two classes: ‘water’ and ‘not water’. Speed, accuracy and intended further processing will determine which features should be input into which layers. Deciding which layers are necessary for a project, deciding how they should be input (scale, detail, etc.) and deciding what level of accuracy is necessary, are all factors that should be determined *before* starting a project. Once the project is underway, those parameters may be difficult - if not impossible - to modify. Some idea of how the data can be analysed is therefore necessary: *what can a GIS do to get the most from your ‘raw’ dataset?*

3.3 How can GIS help?

Geographical Information Systems are designed to answer questions about *where* something is, and *what* that feature is (its shape, size, name, etc). Frequently-used simple GIS queries include: “*Where is / Where are…”, “How long / How big…”*. Questions such as “Where is the scrubland on the aerial photo in Figure 3-2?” would fall in this category. The answer is that any pixel in category 4 (raster) or coloured dark grey (vector) highlights where the scrubland is. More complex queries would be:

“What patterns exist…” and *“Where have changes occurred…”*. The ability of a GIS to rapidly sift through numerous datasets, select relevant information, merge datasets, carry out required analyses, and produce colour summary maps, has been a boon to decision makers: they can much more effectively consider a range of *“What if…”* scenarios.

A GIS allows users to highlight answers to questions about the location and distribution of a particular feature or class. Information about the feature/class can be extracted by the GIS, provided each data-layer shares the same geographical referencing system. This ‘geo-referencing’ is usually undertaken at the data input stage (see Chapter 2 for more details). As the co-ordinates are known, it is easy to derive statistical information about a class from the data layer through simple arithmetic. Thus when we get the answer to our question ‘*Where is the scrubland?*’ we can also answer the question ‘*How much scrubland is there?’*. With the vector model, trigonometry provides the answer. With the raster model, the pixel size can be determined, and simple arithmetic can provide the areas covered by the relevant pixels.

Inter-relationships between layers can be determined within a GIS by breaking down each question into its component parts. Each factor can then be assessed individually and the results combined to produce an answer. A complex question that a suitably constructed GIS could answer may be ‘*Where is the optimum site to build a new airport?*’ This can be broken down into specific criteria (or contributing factors), each of which can be addressed by manipulation and querying of a single layer. One criterion which might have to be met would be ‘*The airport should be more than five kilometres from the nearest town’*. By determining urban areas from a land use map, then creating a new layer where all urban perimeters are ‘grown’ by five kilometres, the new layer will contain areas that meet the specific criteria. The extraction of urban areas is very straight-forward, simply a case of deriving a new layer where all non-urban areas are assigned to a ‘non-urban’ class and all urban areas are retained. The ‘distance from’ operation is usually called a ‘buffering’ operation, where the operator assigns a new value to anything within a specific distance from a feature or features. The data-layer that results from this manipulation therefore only contains areas that do, or do not, meet the criterion.

Once a complex question has been broken down into specific criteria and data-layers have been produced which define those criteria, simple mathematics can be used to produce a final data-layer which answers the original question. This approach to combining data-layers, based on ‘Boolean logic’, is central to GIS analysis, and is shown schematically in Figure 3-3; further details are given in Chapter 7. Burrough and MacDonnell (1998), DeMers (2000) and Heywood *et al.* (1998) provide readable, yet comprehensive, descriptions of more complex GIS analysis techniques.

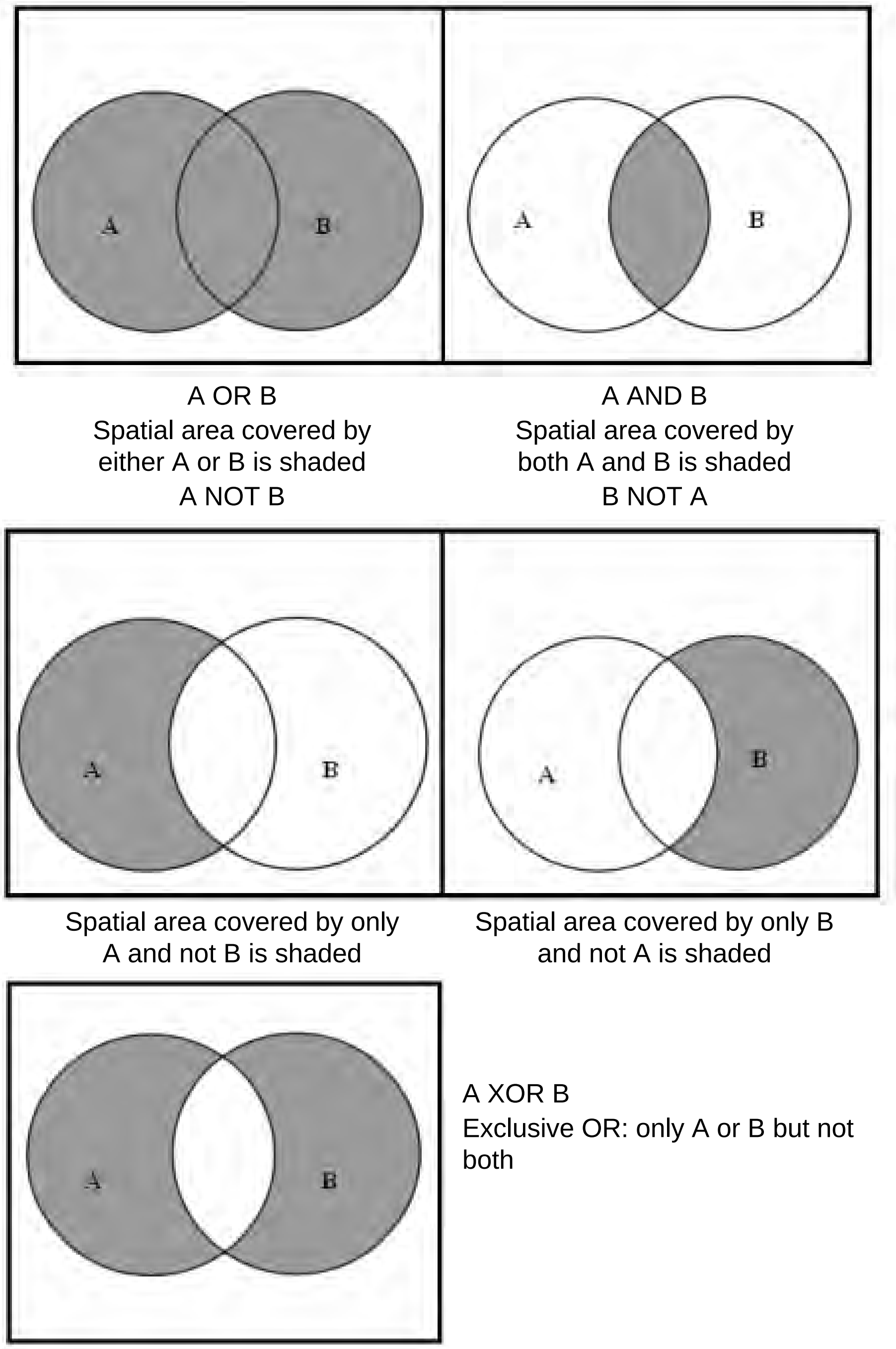


Figure 3-3 Logical operators in spatial queries and operations.

3.4 GIS and fieldwork

Empirical approaches to complex and dynamic systems necessitate the handling of very large sets of data. Multivariate analyses are often used with GIS, either to pre-treat data before GIS input or to process data after GIS analysis. Comparison of remotely sensed data from two or more points in time, using archive aerial photography or multispectral imaging, allows the quantification of dynamic processes, such as floodplain erosion (Winterbottom & Gilvaer 1997, Mount *et al.* 2003). Jorge & Garcia (1997) used GIS and Landsat-derived land cover maps to assess the landscape ecology of a floodplain forest-savanna region. The degree of habitat forest fragmentation was derived from (1) the mean patch area and perimeter, (2) the patch number and density, (3) the perimeter-area ratio, fractal dimension (D), and shape diversity index (SI), and (4) the distance between patches and dispersion index (R). Other types of application are listed in Table 3.1.

Table 3-1 Possible uses of GIS for expeditions and fieldwork.

- Expedition planning: site selection, sampling design, navigation
- Simple maps: data collection sites, roads, contours,
- Data recording and verification:
- Thematic maps: vegetation, animal distribution, soils, land use
- Map layers: overlay and compare different data sets
- Spatial operations: intersecting/overlaying layers, distance buffering
- Spatial analysis: correlation, interpolation
- Modelling: using some spatial data sets to predict another e.g. plant distribution
- Monitoring: record, map and analyze spatial phenomena at time intervals: change analysis
- Training: team members from all countries learning GISci field and lab techniques
- Education: data collection and maps provide a good way to communicate and involve people
- Results: maps are an effective means of presenting results

Apart from the increasing presence on the Internet, the next step in the evolution of GIS is to merge 3-D GIS applications with map and remote sensing archives, allowing the visualisation of changes over time. These techniques offer opportunities for animations, showing complex environmental processes, such as landscape dynamics or nutrient flows, varying in space and time (Reichhardt 1996, Mount *et al.* 2003).

Perhaps the most challenging advance lies in GIS-based modelling, offering a means to test hypotheses, through experimental designs, at different spatial and temporal scales. Models allow scientists to study the response of simplified systems (reduced to a set of variables deemed to be most important) to a restricted set of variations. GIS provides effective tools to run spatially explicit models and offers opportunities to extend models with cartographic input and output (Burrough & McDonnell 1998, Wadsworth & Treeweek 1999). GIS-based models have been used for fundamental research (Carpenter *et al.* 1999, Ormerod & Watkinson 2000, Nemani & Running 1996). Several GIS-based models for mapping the Habitat Suitability Index (HSI) of various species have recently been developed. GIS databases provide a source of habitat information for developing spatial HSI models. By extending the use of HSI models from simple habitat mapping to scenario testing, GIS can assist in the prediction of the outcomes of alternative resource-use strategies in ecosystem studies (Kliskey *et al.* 1999, Jochem *et al*. 2002, Leuven *et al*. 2000).

Table 3-2 gives an example of how several geographical datasets have been used in a GIS to predict the likely distributions of large mammals in a Tanzanian Game Reserve. Data were assembled from several sources: conducting regular surveys during an expedition, digitizing from existing maps, and making use of satellite imagery. The results of this study provided information for Mkomazi's management plan; expeditions using GIS can have a useful role in supporting environmental decision-makers and policy-makers.

Table 3-2 Geographical data inputs, analyses and outputs used in a study in Mkomazi Game Reserve, Tanzania. The aim was to model the distribution of large mammals in relation to their observed habitat preferences.

data requirements ↓	***mammal distribution in sample areas***	***habitat preferences, indicated by environmental variables across the entire study area***			
		seasonal changes in vegetation cover	vegetation categories	proximity to open water	terrain steepness
date source ↓	regular field surveys using GPS over 1 year	satellite imagery: AVHRR, 1km resolution, 1- day composites	satellite imagery: Landsat, 30m resolution	waterhole locations, from GPS survey	digitized 1:50,000 contour map
data processing ↓	create raster grid showing presence / absence of each species	temporal Fourier processing to give indices of seasonality	classification of images to give categories	calculate distance to closest water hole for each cell in a taster grid	convert to raster elevation model and calculate slope angles
analysis	mathematical model relating locations of mammal observations to environmental variables (using logistic regression)				
output	probability of each species' occurrence in each 1km grid cell in the study area -> map of the predicted distribution of each species (elephant, giraffe, impala, etc.)				

3.5 Limitations of GIS

Table 3-3 summarises common drawbacks when attempting to use GIS. With the multiplicity of GIS software, problems frequently arise from format compatibility when importing data. For GIS databases to be comparable or compatible, a preliminary standardisation of the data (e.g. time and spatial scale, variable definition) must be operated. A GIS is useless without good data ('GIGO': garbage in, garbage out). With reliable data, there is still a risk that inappropriate data processing will produce unreliable results, even though the GIS-generated output might look very impressive. Moreover, there is always the danger of over-interpreting GIS results. Users need to know the assumptions that went into remotely sensed data and creating a map or GIS-based model, including the scale and quality of the original data (i.e. meta-data records) (Reichhardt 1996). Understanding the errors and uncertainties in the input and output data are central concerns. Table 3-4 gives some general advice to anyone considering a GIS-based approach.

Table 3-3 Common drawbacks when attempting to use GIS (Wadsworth & Treeweek 1999).

- Lack of digital data
- Lack of time for data collection and entry
- Lack of experience and familiarity with software
- False precision (obscuring sources of error)
- A technology-led approach
- Over-investment in data irrelevant to research goals or decision-making

Table 3-4 General advice to anyone considering a GIS-based approach (Johnston 1998).

- Keep it simple
- Ask whether a GIS is necessary to tackle key questions
- Use existing data where possible;
- Plan ahead or conceptualise (for example, use data management systems or flow-charts to guide GIS development)
- Keep good records particularly source data and analyses performed at each step in the GIS process
- Always check results to see if the GIS output is logical
- Consult with experienced GIS users for advice on database management, data needs and procedures

At a more fundamental level, several objections have been made to the approach of a ‘sliced environment’ introduced by remote sensing and GIS tools. According to Muller *et al.* (1993), there is no evidence that each discipline can be reduced to a set of layers of spatial information. For most botanists, for example, the quality and hence the value, of a vegetation map is highly dependant on the selected classification system. A vegetation layer in GIS can never show the vegetation in all its aspects.

Over recent decades scientists have battled with the problem of how to present spatial information within the limits of physiology and the psychology of perception (Wadsworth & Treweek 1999). Tufte (1983) described five principles of ‘graphical excellence’:

- Well-designed presentation of interesting data (substance, statistics and design)
- Complex ideas communicated with clarity, precision and efficiency
- Giving the viewer the greatest number of ideas in the shortest time
- Data is nearly always multivariate
- Telling the truth about the data

While there are always going to be limitations as to what can be done with *GISci* technologies, there have been tremendous advances in recent years, with the quality of data going up, as the costs of data and data processing have been falling. Hatton (2003) produced a review of development and possible future applications of geoscience information systems, which included some quotes that highlight how rapidly information technologies have developed. Pause for a moment and reflect on how difficult it would be to carry out fieldwork using *GISci* techniques, if the following expert predictions had come true….

"*Computers in the future may weigh no more than 1.5 tons*." Popular Mechanics, 1949.

"*I have travelled the length and breadth of this country and talked with the best people, and I can assure you that data processing is a fad that won't last out the year*." Senior editor for Prentice Hall publishers, 1957.

"*640K ought to be enough for anybody*." Bill Gates, 1981.

3.6 GIS data types and sources

This section provides an overview of the sorts of data that might be of use to fieldwork projects where GIS usage is planned. Aspects of remote sensing, photogrammetry and GPS data are briefly covered here, but are dealt with in detail in later chapters. The contact details for data suppliers have been kept to minimal levels, as full details are given in the Appendix and as web links on the Manual's CD; further information can also be readily obtained, in most cases, via Internet searches. Allow adequate time when ordering data: digital data such as satellite images can now be delivered within days, thanks to advances in data compression technology and Internet FTP (File Transfer Protocol) links; but paper maps, and prints of aerial photographs in particular, can take many weeks to arrive.

3.6.1 Types of data for fieldwork

Analogue data are all forms of paper record or 'hard copy': maps, sketches, photographs (graphical representations), plus tables and statistics (quantitative summaries). Graphical representations can be transferred onto a GIS either by digitising, using a vector system; or by scanning, using a raster system. Quantitative analogue data has to be typed into a database or spreadsheet, such as Access or Excel, before it can be loaded onto a GIS. *Digital data* can be tabular (database/spreadsheet) or graphical (digitised or scanned) and are in a computer-coded format, mostly stored on CD-ROMs, ZIP disks or DVDs, all of which can hold 100s of megabytes of data. Satellite remote sensing images, which are of particular value to expeditions looking at poorly-mapped regions, require larger-volume data storage systems. For instance, a Landsat TM scene covering c. 180 km x 180 km in seven parts of the electro-magnetic spectrum takes up c. 260 megabytes.

3.6.2 Analogue maps

Published maps for topography, geology, vegetation cover and soils will probably exist at national and regional levels.....*whether they are still in stock in the host country is another thing.* Try to purchase regional and local maps from a UK map supplier, such as Stanfords or GeoPubs, before going. *Tactical Pilotage Charts* (TPCs) show elevation, topography and basic vegetation cover, at 1:500,000 scale, and are available for most parts of the world. Regional surveys of soils and vegetation cover have been carried out for many countries by the Food & Agriculture Organisation of the United Nations (FAO, Rome). You may be able to view maps in the map collection of the RGS-IBG, or review the coverage of a region using the CARTO-NET database of the British Library. A worldwide directory of national earth-science agencies and related international organisations is published by the US Geological Survey. It gives details of overseas agencies that might hold maps or sets of airphotos. It may also be worthwhile to enquire at the National Cartographic Information Service (NCIS) of the US Geological Survey, or the Cartographic Section of the US National Archive. More detailed maps, at 1:50,000 or 1:25,000, often only cover a small percentage of developing countries and may date from colonial times. If you can obtain detailed maps, think yourself lucky, but bear in mind that the map data may be over 20 years old and accuracy may not be as high as UK Ordnance Survey maps.

3.6.3 Analogue data sources

Maps, reports and data on features in your study area (such as climate, soils and vegetation cover) should be available at the national, and probably regional, level in major encyclopaedias, atlases and textbooks. Useful information for overseas projects might be obtained from the following sources:

- For former UK colonies: the Natural Resources Institute at Greenwich University, the British Geological Survey, the Ordnance Survey, the Natural History Museum, Royal Geographical Society (with IBG) Foyle Reading Room.
- French overseas dependencies: Institut Geographique National (IGN), Bureau de Recherche Geologique et Miniere (BRGM), the Centre Nationale de Recherche Scientifique (CNRS) or the Institut de Recherche pour le Development (IRD, formerly known as ORSTOM).
- Libraries of learned societies (e.g. the RGS-IBG, the Geological Society) or research institutes, such as the Scott Polar Research Institute or the Oxford Forestry Institute.
- University libraries, both in the UK and in host countries, which may have archives of reports and maps.
- The British Council or Association of Commonwealth Universities, which may have records of existing links between UK Universities and the expedition's host country.
- UK aid agencies (e.g. OXFAM, VSO, ITDG, WaterAid).
- The US National Archive and the Smithsonian Institute hold geographical data from around the world that may be of use.
- Relevant ministries/agencies in the host country (agriculture, forestry, land survey).
- Commercial map suppliers, notably Stanfords and Geopubs in the UK, or Omni Resources in the USA.

If your fieldwork is overseas, you should be aiming to carry out your fieldwork jointly with a team from the host country: this will facilitate access to data. Furthermore, this facilitates the transfer of skills to places where they are most needed and helps to reduce bureaucratic and logistical problems.

3.6.4 Sources of digital data

More and more data summarising the features of the world are being digitised and are available over the Internet, from word-processed documents and spreadsheet tables, to scanned airphotos and satellite images of entire continents. Digital map data can be obtained in three ways: inputting your own data, purchasing ‘off-the-shelf’, and down-loading from the Internet.

(a) *Inputting your own data* can be time consuming and tedious, but analogue data is more useful in a GIS-usable format. Data can be input as follows:

- Typed-in tables, spreadsheets or databases, with an object's grid reference followed by its attribute data.
- With a vector-based system data can be digitised in as points, lines and areas, each with an identifying code that can be linked to a database management system (DBMS). The required digitising board and stylus will cost at least US$300.
- Graphical data can be scanned in, saving a lot of time, and provides an excellent means of comparing or merging different maps or airphotos of the study area. However, the scanned image is only a picture made up of cells with varying values, not an inter-related network of points lines and areas: vector overlays of points, lines and area boundaries may have to be created by on-screen digitising using your pc mouse.

(b*) Surfing the web* A vast array of information, including digital maps and images can now be found on the Internet: a summary of useful Internet websites is given in the Appendix. At the expedition planning stage, the sites of the RGS-IBG (*www.rgs.org/mapping*), the US National Geographic Society and the US Geological Survey (USGS) are particularly useful.

(c) *Digital maps* are now easily obtainable at broad continental and national levels, usually on CDs or increasingly for download from websites. These digital maps of developing countries are unlikely to be more detailed than 1:1,000,000, and are therefore only of use at the expedition's preliminary planning stage or where regional studies are important. A summary of some digital map coverage that might fit into an expedition budget is given in Table 3-5.

Table 3-5 Some examples of digital map data sources.

package	datasets	regions	scale	format	approx. cost
Bartholomew Digital Maps	Similar to DCW (*licensed to some universities - check for availability)*	Global Europe	1:1M 1:250K	CD or Internet	£360/year
Digital Chart of the World (DCW)	Altitude, coasts, rivers, vegetation, borders, roads/rail, populated places. Data source: Tactical Pilotage Charts	Global	1:1M	CD / Internet	£245 / FREE
Eurographics	Various, e.g. soils, vegetation, geology, topography, relief.	European nations	various	CD / Internet	enquire for details
Global Land Cover Facility (GLCF),	Online access to archive Landsat data hosted by University of Maryland	Global (ex. Antarctica)	various	CD / Internet	FREE via FTP, c. £30 per CD
GRASS (datasets for this popular GIS software)	Soils, altitude, aspect, vegetation, threatened species, marine biology	Global	1:1M	CD / Internet	FREE
MacArthur Project Cambridge University	Vegetation, land use, soils, admin boundaries, elevation, climate, people	Inner Asia			enquire for details
National Geographic	Various, notably key biodiversity sites, national parks, relief, politics	Global	various	Internet	enquire for details
ODDENS, University of Utrecht, NL	Extensive online access to atlases, catalogues and maps	Global	various	CD / Internet	enquire for details
Shuttle Radar Topography Mission (SRTM)	Digital elevation model (DEM), good horizontal and vertical accuracy	60ºN to 60ºS	90 m cell size	Internet	FREE
United Nations (UNEP / FAO)	Global Resource Information Database (GRID) *NB many GRID datasets (e.g. FAO Africa soils map) are included with IDRISI GIS software*	Many: environment data sets	1:1M, others	CD or Internet	£15 each
US Geological Survey (USGS)	Topography, relief, geology, aerial photography – mostly US coverage Satellite imagery – worldwide cover.	Mostly USA and the Americas	various	CD / Internet	enquire for details
World Conservation Monitoring Centre	Digital atlases of tropical forests, coral reefs, wetlands, protected areas	Global	various	CD / Internet	enquire for details
Vector Map of the World	Supersedes DCW	Global	1:1 M	Internet	Free