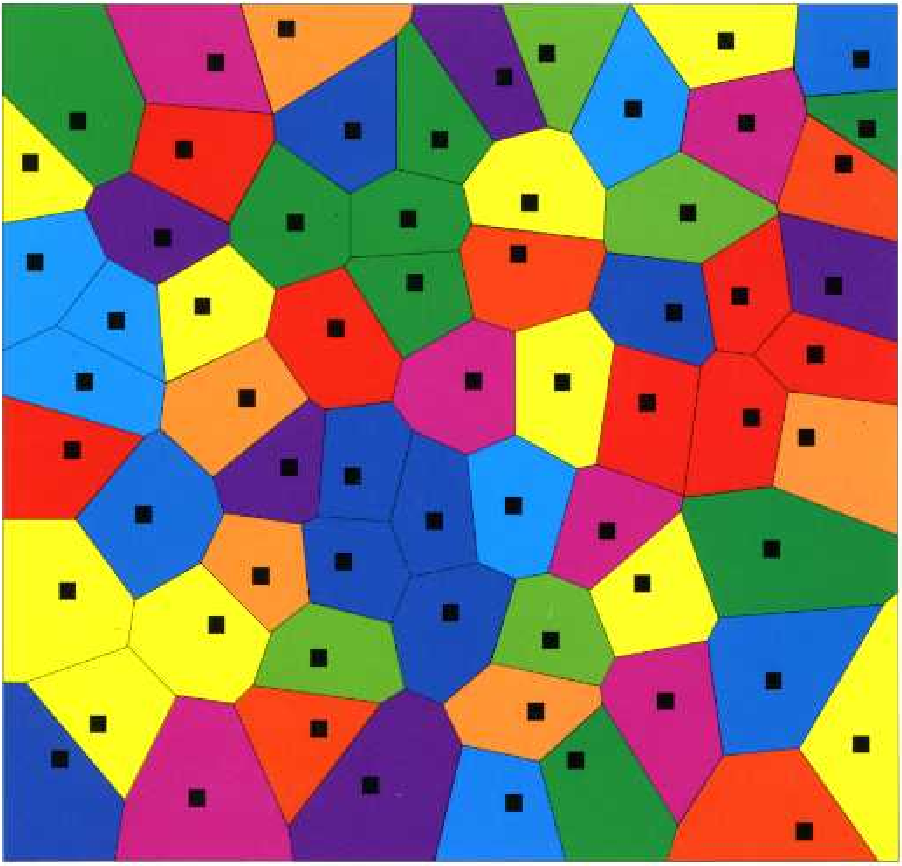
Education Planning and Management, and the use of Geographical Information Systems

John M. Mendelsohn



UNESCO Publishing
International Institute for Educational Planning

Education planning and management, and the use of geographical information systems

Education planning and management, and the use of geographical information systems

by
John M. Mendelsohn

UNESCO Publishing

International Institute for Educational Planning

The publication costs of this volume have been covered through a grant-in-aid offered by UNESCO and by voluntary contributions made by several Member States of UNESCO, the list of which will be found at the end of the report.

Published by the
United Nations Educational, Scientific and Cultural Organization,
7 place de Fontenoy, F 75352 Paris 07 SP, France
Printed in France by Imprimerie Gauthier-Villars, 75018 Paris

Cover design by Blandine Cliquet
ISBN 92-803-1158-1
 IIEP/jn

Preface

Space is one of the key dimensions of our lives. In some places, space is rigidly controlled by such clear features as fences, political borders, walls, and school districts; and people arrange what they do and where they go according to these barriers. These kinds of boundaries may not have been erected in other places, but space remains equally important. Where can water be obtained, where can crops be planted, and how far must a child walk to school? Space is both an important resource and constraint for everyone. It also has a substantial impact on education.

What does someone working as an education planner or manager do about considering the effects of space on education? How does this person quantify, analyse and display the relationships between geography and education? How does he or she determine just *where* education should happen? These are not easy questions to answer, but use of the tools described in this book will make it easier to obtain answers.

A Geographical Information System (GIS) comprises a set of tools useful for processing spatial information. As is true of any set of tools, much of its effectiveness depends on the paradigm in which it is used. Education planners and managers with a strong interest in the details of education, in local conditions and local solutions, and in having their efforts translated into tangible products, will make good use of a GIS. Those with broader interests in policies, processes and central control will not use a GIS to the same extent. This book has been designed to find something of a middle road between those with a strong interest in education and those with a greater interest in the technology of GISs. The book seeks to bridge that gap by showing how modern computer technology can be 'user-friendly' in a discipline that often avoids the use of technology in looking for solutions. Hopefully, it will convince educationalists who wonder whether a GIS can be useful that there is much to be gained. For those embarking on the actual development of a GIS, this book provides introductory material to make clear the methodological principles and broad products to be expected from such a system. And, for those who know a good deal about GIS, but wonder how to apply the technology in an education setting, examples of spatial issues and questions in education should illustrate the potential value of the applications.

This book describes the planning and development of a geographical information system, explains the components necessary for a GIS and presents a variety of examples of how a GIS map can be used in educational planning and management.

Parts of this book will be published in a series of modules on school mapping and microplanning. But given its wider perspective it was felt necessary to make it available to a wider audience.

While this book is intended for those who might wish to develop and use a GIS for education planning and management, it should also be of interest to people working in other social sectors. Many of the principles important to education are relevant to other services and processes involving people.

Jacques Hallak
Assistant Director-General, UNESCO
Director, IIEP

Contents

Chapter 1

Introduction

Education depends on a variety of resources. Some are tangible and obvious: teachers, classroom facilities, textbooks, pupils, and money. Yet others are less easy to define: political will, time in which to learn, and parental support. Without these and other resources, people find it hard to get the education they need or desire, and education managers and planners find it hard to ensure that education is effective.

This book focuses on a special resource – space – and looks at methods of studying the relationship between education and spatial issues. As with the other ingredients of education, some aspects of space are more obvious than others. All, however, are associated with questions containing the word *where*. The following list of questions illustrates some common spatial aspects to education:

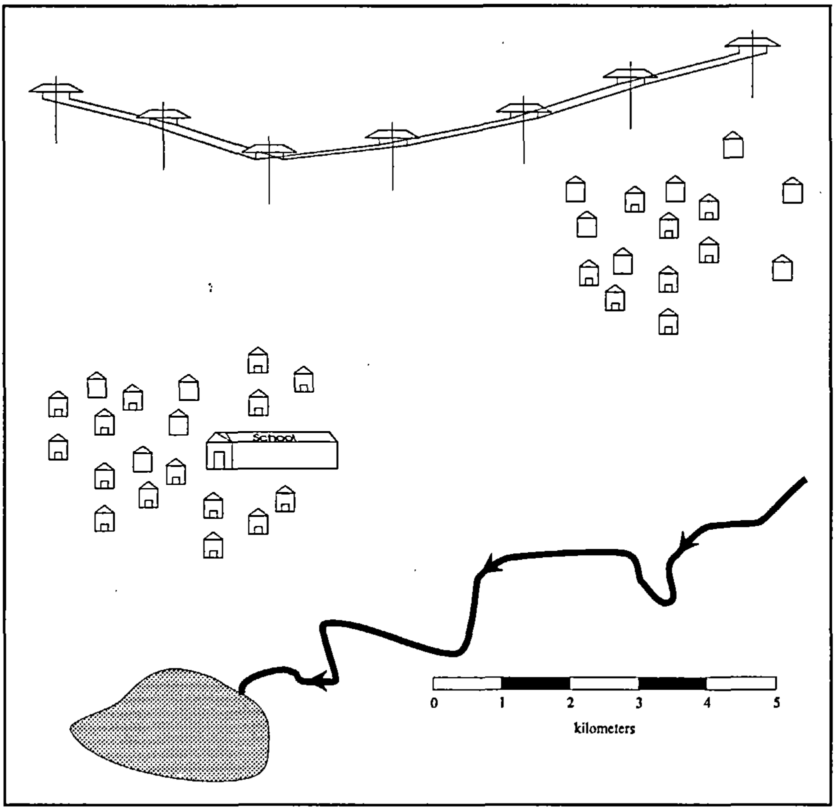
- Where are new schools needed?
- To where can pupils move after completing their primary schooling?
- Where are school facilities duplicated?
- Where are the best bus routes for transporting pupils to school?
- Where are new teachers needed, or where is in-service training required most urgently?
- Where should inspection circuits be created?
- Where do we expect population growth rates to be highest?
- Where is the demand for education going to decline most severely?
- Where is there affordable land, with adequate water supplies and electricity, on which to build schools within easy reach of pupils' homes?
- Where can a teachers' resource centre be built to be within reach of the maximum number of unqualified teachers?

These and many similar questions confront education managers and planners almost daily. Unfortunately, such questions cannot be answered as routinely, so answers are often only available after exhaustive and expensive research, or the decisions that are made are often less than optimal. A major consequence of this is that educationalists seldom view their activities through geographical or spatial spectacles. Schools are built where they may not really be needed, in-service training activities could be offered elsewhere to greater advantage, and the chances of pupils moving smoothly from one phase of education to another is reduced. In short, scarce resources needed for education are not always placed *where* they can be used most profitably, and people *elsewhere* are deprived of these resources as a consequence.

Use of the tools, philosophy and techniques described in this book should help improve the frequency with which better decisions are made on where education is placed. What is certain is that these tools help to make geographical perspectives to education more readily available. They also help to balance other perspectives. For example, politicians (having strong and obvious political outlooks on education) may debate forcefully in favour of new school services being provided to their constituents. The introduction of geographical or spatial issues provides such debates with a more rational or objective input, allowing political stances to be countered where needed. Similarly, a strong social perspective on education could suggest that boarding accommodation at schools should be abolished because of the social costs incurred by pupils living away from their parents. Looking at the issue geographically may, however, add different views by showing that it is not possible for a majority of pupils to live at home and attend schools *where* they are currently placed. As a last example, a principal may request authority to upgrade his or her school such that it provides a science curriculum. Looking at *where* this request comes from could show, however, that there are several nearby schools offering science courses and that these facilities are underutilized. The same exercise might also show that there are several industries in that general area but that vocational training opportunities are in short supply. It might then be more beneficial for the school to turn its energies to providing vocational education.

Geographical perspectives are made possible only by looking at the location of features and by knowing something about those features. Maps or visits to areas show us where features are situated and how their locations relate to the position of other features. Information about features tells us what they are and what they do: a primary school, a school hostel, a school district, a road, a home, mountain, electricity line, or a political region. The following map (*Figure 1*) shows why it is necessary to simultaneously consider information about features (their attributes) and their positions or spatial locations. The dotted line on the map represents a river, while the arrows shows the direction of flow. This line and the arrows immediately tell us that land to the right of the map is at a higher elevation than land to the left or west. We will also know that land on either side of the river is above the level of the river itself. Now, if we didn't know that this line was a river, it would be impossible to know something about the area around that river.

Figure 1. Map of a school, river, lake, villages and an electricity line



Likewise, we would have no way of knowing where the water in the lake came from without seeing that the river flows into it.

The same rationale can be applied to the school in the map. We can guess that children from the homes around the school have ready access to education; those in homes further away must attend another school, or spend much longer getting to school, or perhaps they can't get to school at all. Our interpretation of these circumstances may be guided by the distances involved, as reflected on the scale of the map. Finally, we can guess that better planning could have made teaching at the school easier by allowing for the provision of electricity. Assuming that the school was built first, whoever erected electricity lines could have done a better job by simply routing the lines past the school and allowing a short connection to be made. The same could be said of the supply of electricity to the houses. In these examples, our geographic perspective is based on the nature of the feature, its position and that of other features in the area. All of this allows us to interpret for ourselves a set of circumstances or processes: the river must be at a lower elevation, the river must feed the lake, elevations are higher in the east than in the west, the school does not have electricity, some children don't have ready access to the school. Our interpretations and perspectives through the map are retrospective here – we are looking at how things have come to be the way they are: how the river carved its way through the landscape, and how someone built a school some distance away from a large group of children. This may be an interesting exercise in itself, but it would be useful to use the same approach in looking forward and applying the same philosophy to planning for the provision of education. It is easy to see that if a geographical perspective had been taken in planning the supply of electricity, the school and its pupils would have been better off.

While the historical scenarios or planning about conditions in *Figure 1* are simple, they depend on four critical requirements. All four have to be place. The first is the map itself, assuming that a visit to the area could not be made. Second, is the information about the features on the map – that the big building is a school, that the small ones are homes, that the dotted line is a river, and so on. Thirdly, we have to know the correct positions of the features, such that the positions correctly relate to those of other features. For example, our conclusions about the lake would be quite different if the river flowed past it. Finally we have to process all this information – take it in through our eyes, and turn it over in our minds to provide interpretations that are both logical and, for planning purposes, useful.

For our purpose in this book, the four requirements are in place – the map in *Figure 1* and the narrative in the last few paragraphs. However, more often than not these components are absent; we don't have a map, we don't know where schools are located in relation to villages or housing communities, and we don't have the time or interest to interpret the relationships between all these features. Geographical Information Systems (GISs) are about having these four components in place and working together to provide useful geographical perspectives.

What are Geographical Information Systems (GISs)?

Geographical Information Systems are computer programmes designed over the past 20 years or so to store and process geographical information. More specifically, GISs provide methods to capture and collect such information, to store that information, to process it statistically and spatially, and finally to make that information available according to the needs of whoever requires it.

The capturing and collecting of information means that digital descriptions are provided of what is on the ground and that the digital information is then stored in a systematic way in a database. Common methods entail the copying of maps on a digitising table or tablet, translating aerial photographs or satellite images, and measuring the positions of features on the ground and then entering those records into a database.

Storage of geographical information is done in a database. In the sense used here, a database is a collection of information stored in a structured fashion so that the information can be retrieved and selected according to criteria determined by a user. As in a filing system, the user can retrieve selected bits of information or documents, and these can be retrieved most easily if the filing system has a systematic structure to it. There are two important kinds of databases. The so-called flat databases are those which exist alone, such that all information about an object is stored in the same file. By contrast, relational databases are those that provide information about the same object by drawing data from different files. For a school, we could have one flat file holding all information about all the staff, buildings, and pupils, or we could have different files for each of these and allow the computer to selectively draw out relevant pieces from each file.

The GISs provide a variety of options to process geographical information. Most simply and fundamentally, they allow the user to retrieve or enquire about information in a particular area. That information may then be related to information about areas elsewhere, for example, to determine the area from which a school draws its pupils, or to find the nearest source of water for a school. Alternatively, you can use a GIS to quickly count all the houses in a catchment area around a school, apply a figure that describes the average number of infants living in each house, calculate the density of infants living in that area, and provide projected enrolments for that school.

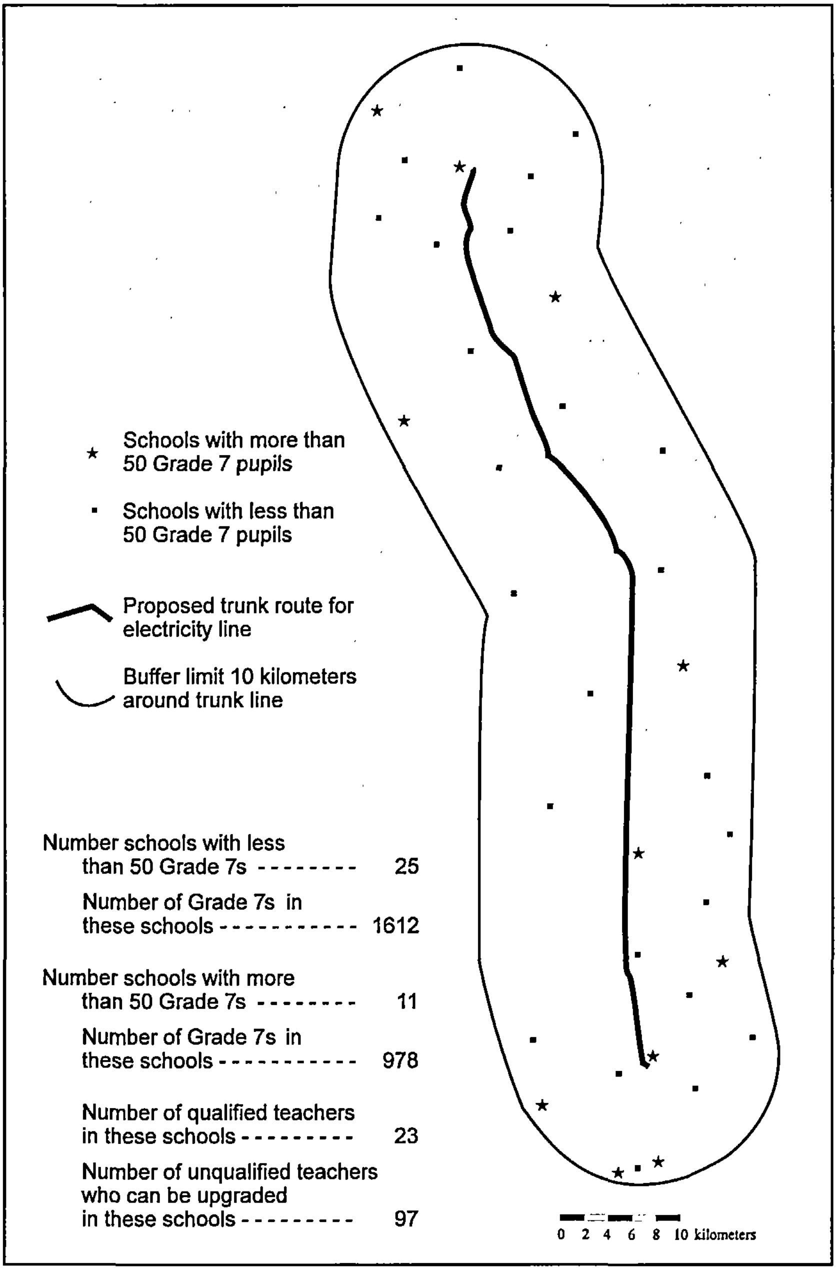
Having collected, stored and processed geographical information, how does a GIS make useful information available? The quickest and simplest is to have the computer draw a map or display the results on a computer screen. Having done this, it is also easy to print the results as a map, or as a set of tables or graphs. It is perhaps the ease with which maps and other products can be printed that really makes a GIS so useful. Maps can thus be printed of all schools offering a particular kind of curriculum, or areas in which children don't go to school, or land that is best suited for a variety of reasons for building new schools. Likewise, lists of schools for a particular area can be generated and printed without having to scan maps and tables trying to see which schools fall in the desired area and which are elsewhere.

Consider the following problem facing a group of regional educational planners, development planners and consulting engineers. A development plan for a region calls for the installation of an electricity trunk line from one point to another, with schools along that line being given the opportunity of being linked to that power supply. All schools in the area now offer only primary grades (Grades 1 to 7) and having electricity would make it possible for some schools to offer Grade 8 and other secondary grades where teaching aids using electricity are required. Among the questions to be answered in this scenario: what schools within a given distance of the proposed power line can be upgraded, how far are these schools from the main power line? How many teachers in those schools are qualified to teach secondary grades? Finally, How many pupils could be enrolled in higher grades? *Figure 2* shows some of the steps taken in answering these questions. First, you could create a buffer area along the planned electricity line. This would be an area either side of the trunk line within which connections to the trunk line could reasonably be made, let's say 10 kilometres. In *Figure 2*, the buffer area extends this distance around the trunk. Next, you would overlay the locations of schools to find those that are within this 10 kilometre buffer area, in other words those that are near enough to benefit from the development. The second step involves using enrolments to identify schools that can be upgraded, those which already have high numbers in the final primary grade (more than 50 pupils, in this example). Eleven schools meet this requirement, and we can then call up a database of teachers to count how many of the teachers in those 11 schools are qualified to teach secondary grades. The result shows there to be 23 such teachers. Another query shows that there are 978 pupils in Grade 7 in those 11 schools with more than 50 Grade 7 pupils. If the 978 pupils all go on to Grade 8 in the next year when a secondary curriculum is introduced, and the curriculum policy is such that the pupil teacher ratio is 30:1, a total (978/30) 33 qualified teachers will be needed. This means that an additional 10 teachers must be found or 10 existing teachers must be given in-service training to teach secondary grades. The database of teachers' qualifications shows there to be 97 teachers who can be readily upgraded to teach a secondary curriculum. Of course, the number of qualified teachers will have to be expanded year-by-year as the first cohort of secondary learners moves up from Grade 8 to Grade 9 and so on.

Additional expansion will also be needed if the 1,612 pupils in the schools with less than 50 Grade 7s move to enrol at the upgraded schools.

All of this information could be retrieved through your GIS, without leaving the map and commands of your screen. The information could also be dispatched to a printer, giving you a map of the area indicating schools for potential upgrading, and printed tables for teachers' qualifications and pupil enrolments. Finally, measurements can be taken on the GIS map to add up the distances over which connecting lines would have to be run to the 11 schools. This provides the consulting engineers with preliminary information on the materials needed and the costs involved. All this would have been a difficult and expensive task in the absence of geographic and database information queried through a GIS.

Figure 2. Buffer area 10 kilometers wide around a proposed electricity line, and information about schools, teachers and pupils within that area



As is true of so many computer systems, a great variety of different programmes are available and many changes have been made to these programmes over the years. GISs are now cheaper, easier to run, more diverse, and more powerful than they were 10 years ago. In addition, geographical data in computer or digital format are often available from various sources, so that a new GIS user may not have to spend great amounts of time translating maps into digital formats to be stored and processed in a computer.

School mapping

As this book will show, GISs have their greatest uses in the fields of education management and planning. Both fields have made surprisingly little use of geographical information. One major exception is school mapping, a set of techniques and approaches that has a geographical perspective of education at its core. School mapping has been strongly advocated by the *International Institute for Educational Planning* (IIEP), Paris, and has a similar life span to that of GIS. It has developed over the past 20-30 years from its origins in France where efforts were then needed to provide additional secondary schools on a rational, accessible and cost-effective basis.

Often also called micro-planning, school mapping takes as its departure point the philosophy of a need for detailed information and planning at the local level, this being a more viable framework than national or macro-planning. The approach seeks to probe beneath the often, mythical regional average, by documenting micro-level variation and planning accordingly. Its focus is on local conditions, local constraints, and local solutions. Three phases to school mapping are required: a diagnosis of local schooling conditions, projections of needs, and proposals for reorganisation. Regional, or even national, norms are required against which to make proposals for reorganisation.

The most important goal of school mapping is to rationalize school resources by using geographical units of analysis. Homogeneous zones are identified, schooling problems and needs are analysed in each area, and proposals are made for rationalization and reorganization in each mapping area. These areas may vary in size according to the phase of education being considered. School mapping areas for primary schools are smaller than those for secondary schools, while mapping areas for tertiary institutions can be very large. In all cases, however, the medium of enquiry and analysis is through a map, to show areas and schools where changes are needed, be they new school buildings, more teachers or, possibly, even the closure of redundant schools. GISs have an obvious role to play in school mapping by providing a geographical frame for the work, by allowing large volumes of analytical information to be available for processing, and by making the results readily available through the medium of printed maps.

While school mapping and GISs share the same philosophy of looking at education through geographical spectacles, the area of focus for school mapping is at the micro-level. GISs, however, enable the user to work at any level, from the country as a whole, right down to individual schools, and then right up again to any chosen level of geographical aggregation. This movement, up and down, is possible *at the press of a button* because the programmes allow you to *zoom in* and *zoom out* to any area within the country. In fact, it would be quite feasible to have a GIS for all schools in the world. You could focus in on one school and the streets around it, and then zoom out to a map of all schools on the continent or on the surface of the earth itself. National and even large regional perspectives are easy to grasp because there are so few units to remember or analyse in our minds, but we often avoid greater detail because there is too much to comprehend. The ability of GIS to focus on any level means that we can immediately get to the details of education – to the individual schools – without being overwhelmed by their numbers and complexity. The ability to view education at so many levels should allow us to manage and plan education to the greater advantage of all concerned.

Chapter 2

The nature of Geographical Information Systems

This chapter explores the basic features of geographical information systems: how they operate and what goes into them. There is a substantial variety of GIS programmes available on the commercial market: some are large and expensive, others are small and simpler to use. The programmes are also run in several different computer environments and operating systems. In spite of all this variety, all GIS programmes are organized according to certain rules and principles. Learning how to use any of these programmes requires a clear understanding of these basic features.

In reality, a GIS is simply a method of producing maps and processing geographic information. All that has happened in the modern computer age is that new methods of processing map information are available. Perhaps this is best explained by pulling apart what goes into a map, what is on a map, and what is done with maps.

What are maps?

Maps are usually flat pieces of paper on which objects are drawn to show their location in the world. Assuming that the map is an accurate representation of the world, the positions of features are correctly shown in relation to each other. This is achieved by using units of measurement or co-ordinates to place the features on the paper or map in their true relative position. So people who compile maps (surveyors, cartographers) go out into the real world, measure where geographical features are placed and then translate these recordings into relative positions on maps back in their drawing offices. Exactly the same has to be done in compiling the sets of data that go into a GIS. Instead of positioning features on maps according to their co-ordinates, the co-ordinates are stored as numerical values in a GIS's spatial database. You cannot have a GIS unless you have geographical or spatial data. Different kinds of spatial data are described later in this chapter and *Chapter 3* provides a description of how spatial data are collected.

In addition to the spatial information that is collected, recorded and available on maps, maps also provide attribute or statistical information. Features are drawn in such a way as they are immediately recognized –

a river, a road, a school. Most usually, symbols are used to denote their attributes, such that a river is shown as a blue line, a road is symbolized by a red line, and a school as a small block with the letter 'S' next to it. The meaning of the symbols is given in a legend to the map.

These are attributes of the simplest kind. More detailed maps often give us additional information on the properties of features. The width of the blue line might be drawn in direct proportion to the volume of water carried by the river, the name of the road might be printed next to it, and a small graph or table next to the school may show how enrolments have changed over the years. In principle, any kind of attribute information about features on a map can be collected, stored and be made available on a map. GISs do exactly the same. Attribute information is collected, stored in the GIS and processed so that it is made available according to the needs of the user. So in addition to storing the co-ordinates of features, GISs have ways of storing attribute information about those features.

Once all the information required for a map has been assembled, it has to be drawn. That step may seem so logical and obvious that many of us forget about the enormous amount of work, both of a technical and artistic nature, that goes into the actual drawing. All the tiny contour boundaries, the shading of different vegetation and soil types, and all the annotation has to be drawn on to the map. Good maps provide lots of information, they are attractive to look at, the information is clearly presented, and the lines and colours have to be sufficiently clear to be reproduced by printers. Moreover, the maps should be as current or up-to-date as possible. In essence, the better the map, the greater the effort that has gone into its production.

Until recently, a draftsman would have to spend weeks and months compiling a single map drawing. GISs fortunately make comparatively short work of drawing maps. The computers that operate GISs are linked to printers that can produce a great variety of colours, shading, and lettering in any number of fonts. Depending on the printer, maps can be on almost any size of paper. The GIS user certainly has to do a lot of work in extracting the necessary data, making sure that it displays clearly to meet the same requirements of maps drawn using conventional methods. But the amount of time required to do all of that work is very much reduced. This is true for two important reasons. First the GIS takes care of plotting all the geographic features in their correct positions, so that there is no need to use rulers and set squares to translate co-ordinates on to a piece of paper. Second, and of greatest importance, new maps can be drawn whenever new information becomes available. The fact that most maps are out of date or were produced a good number of years ago is often due to the volumes of work, and hence expense, that would have to go into producing more current versions.

For the education planner and manager who uses a GIS, maps can be printed every year showing the distribution of schools, and custom maps showing very specific sets of information can be produced quickly.

For example, there may be a need to have a map showing the distribution of schools that offer particular subjects, or schools that have special problems, or schools where building projects are currently underway. All of this means that maps showing features of education can be printed on a regular basis or as demands arise. Maps can then become routine tools used in administering and planning education. In the absence of GISs, very few educational authorities would ever bother to even think of employing cartographers and draftsmen to produce such specific maps.

The last major point about maps concerns how they are used. The geographical features and attribute information shown on a map depend on what the map was intended for in the first place. People read or study maps to find their way, to find out where something is, or to gain a general understanding of the character and environment of an area. Cartographers and draftsmen, accordingly, produce maps with simple and standard reference information to provide for such uses. GISs can do exactly the same. However, maps can also be used to provide information on *relationships* between different features. Simple examples of this were given in *Chapter 1* where it could be seen from a map that a river fed into a lake, such that the river was the source of water for that lake, and that certain schools could benefit from being close to a proposed electricity line. The relationship between the river and lake is a simple one. But what about asking whether the water in that lake could be used to supply the school with water? To answer that question, it would be useful to have a map showing possible routes along which a pipe could be laid to the school. It may be logical to choose a direct line for such a pipe, but that option could be more expensive if it were discovered that the trench for the pipe would have to be excavated through very hard types of rock. Given this scenario, a map could be asked for showing costs of laying a pipe along different routes, some from further upstream, others downstream, some over high or low ground, some through hard rocks, others through softer soils. All of these alternatives can only be determined if the *relationships* could be mapped between the length of pipe needed, the kinds of rock, and the heights of the ground over which the pipe would be laid. All of this can be done manually, but thousands upon thousands of calculations would be needed. It would also be very hard to draw all the results on to a map in a way that would allow the engineers and school administrators to make the best decisions. The analysis of these kinds of relationships is a particular strength of GISs. The programmes can overlay a variety of spatial and attribute information to calculate factors that reflect relationships between rock types, the distance or length of pipe from the river and to the school, and the ground heights. This kind of GIS processing involves analysing the relationships between different spatial objects.

In summary, maps provide us with information on the location of spatial features or objects, and give us information about their attributes. To do so, the maps have to be drawn and printed and, of course, they have to be used in a way which provides the information needed. GISs do all of these things, and it is as well to remember the similarity between maps or mapping and the development and use of a GIS. Too often people get so involved in the technical aspects of databases, programmes, processing options, computer environments and the like that they fail to recognize that a GIS *is simply a tool.* Under those circumstances, a GIS is unlikely to give you the mapping information, processing and maps you needed or wanted in the first place!

Spatial information

It is hoped that it is clear by now that there are two distinct kinds of information or data that go into a map and thus into a GIS: spatial data and attribute data. GISs always hold spatial and attribute data separate in one sense or another, even though the user may not always see them as separate. People who develop GISs, especially in developing countries and for purposes of planning and managing education, are likely to spend a good deal of their time initially assembling spatial data. This would include the co-ordinates of schools, the boundaries of census blocks or enumeration areas, political and administrative boundaries, the locations of water and electricity supplies, and the locations of roads.

Spatial databases consist of any number of spatial objects. For example, a spatial database for school planning might consist of all the schools in a region, roads, electricity lines, telephone lines, districts, school catchment areas, census areas, water points and school hostels. Each school, hostel, road, or school catchment area is treated and stored as a separate spatial object, and each object is labelled with a unique label or *geocode.*

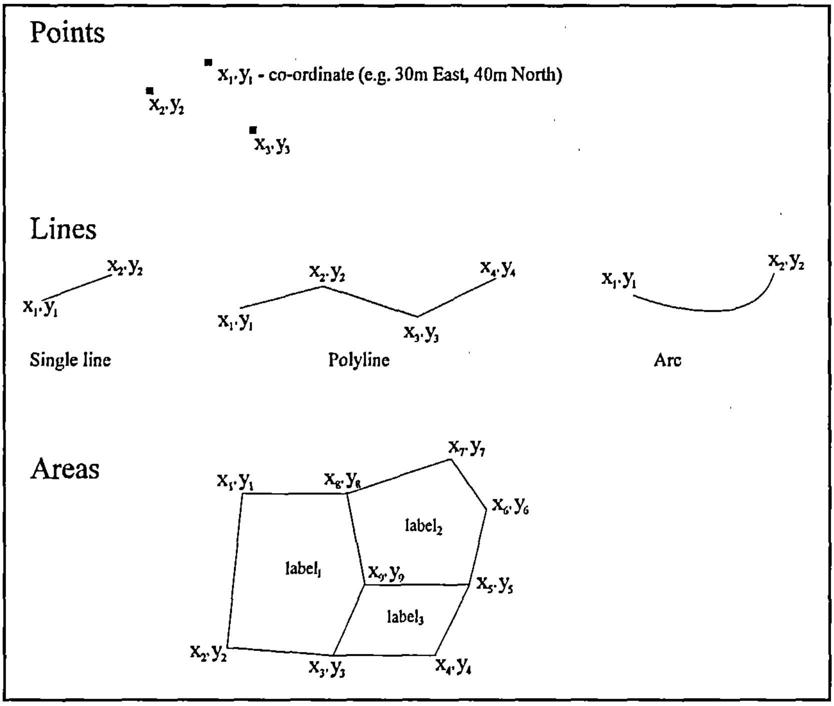
Geographic information is recorded using co-ordinates. These are units of distance which determine where a feature is found. A variety of *co-ordinate systems* are used by geographers, the choice of system varying from country to country and according to the nature of the area being mapped. In addition to recording the position of an object, co-ordinate systems provide methods of translating the positions of features as they exist on the curved surface of the earth onto a flat piece of paper. The mathematical rules for making these translations are embodied in *map projections.* No single projection can accurately portray features from a curved surface into two dimensions without causing some kind of distortion, be it distortion in shape, distance, area or direction. Different kinds of map projections and co-ordinate systems have been designed to best map different kinds of areas. GIS programmes can handle most standard co-ordinate systems and it is advisable that the developer of a spatial database adopt some well-known system at the outset, preferably one used by others in the same country. This makes it much easier to exchange data so that different sets of map information can be used together.

The GISs store spatial objects in one of three ways: as points, lines, or areas. The co-ordinates used to describe these three kinds of objects differ. *Point* objects are described by a single set of co-ordinates, one giving a unit of measurement along the x-axis and the other along the y-axis (*Figure* 3). A point therefore has no dimension or size, it is simply a point on the surface of the earth or on a map. Typically, features such as schools, towns, villages, wells, trees, etc., are recorded as points in a spatial database. This would be the case for spatial data covering a large area where there is much open space between towns or schools. However, points would not be useful in detailed maps showing the lay-out of a town or individual school buildings. In this case, areas and lines would be used to describe the streets, buildings and other components of a town or school. Points are the simplest kind of spatial object in a GIS. They also require the least storage space in the database, since only one set of co-ordinates and a label identifying the point as a unique object has to be recorded.

A *line* is defined by two sets of co-ordinates, one determining the origin and the other its end point. By connecting these pairs of co-ordinates, a line tells us the distance between the two points and the angular direction of one point in relation to the other. In many applications, it is useful to have line segments linked to one another to describe linear objects, for example, a road, river, or telephone line. Such a series of linked line segments is usually called a *polyline*, although other names are used, depending on the GIS programme. The joint between one line and another is called a *node*. In building or compiling a spatial database, you have to be absolutely sure that the line segments are actually joined if they are to be recognized as being linked and thus part of the same linear object. In other words, the two line segments have to share the same node. Each line segment can have its own unique label, or all the lines that form part of one polyline can share the same unique code number or name. An *arc* is a special kind of line segment or polyline where the lines between nodes may be curved according to a mathematical function determined by the user and stored in the GIS database.

Areas are often called polygons in GIS terminology. They are units that circumscribe distinct pieces of ground, for example: districts, regions, census blocks or enumeration areas, land parcels, and catchment areas. Polygons often lie adjacent to one another and share boundaries, such that two districts are next to each other. A series of lines are used to describe the boundaries of a polygon, and the positions of the line segments are stored using the same kinds of co-ordinates and nodes used for polylines. Each line therefore has to connect with the next and, to form a polygon, the last line or segment has to join up with the origin of the first line (*Figure 3*).

Figure 3. The main kinds of spatial objects: points, lines and areas (polygons)

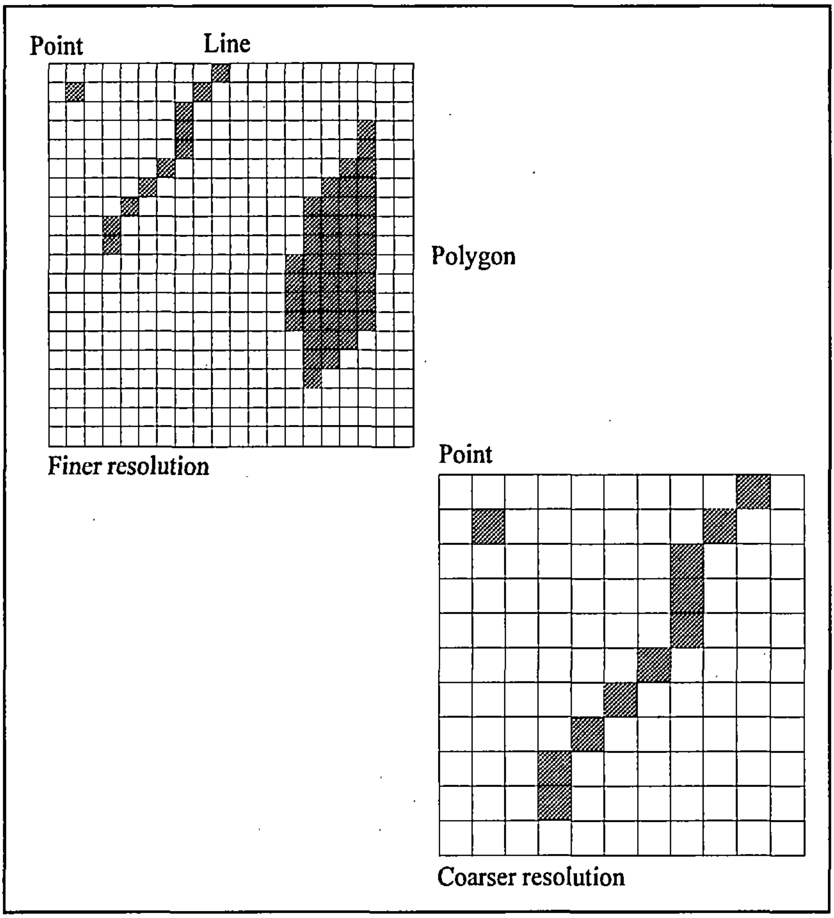


Note: The positions and dimensions of each are governed by series of co-ordinates (x,y).

By doing so and thus enclosing an area, the GIS is able to 'recognize' the polygon as a distinct area and give it a unique label. In addition, the relationship between two adjacent polygons is established by having the GIS treat the border between them as a shared border. One useful feature is that GISs generally calculate and make available to the user the area of a polygon for calculations requiring estimates of density, land value etc.

Most GISs now store and process spatial data in terms of co-ordinates and associated rules for connecting them as lines and areas, as described here. This is called *vector* data. Some programmes, however, treat spatial objects as *raster* data. The area covered by the map or spatial database is then divided into a number of grid cells, the position of each grid cell being controlled by a set of co-ordinates identifying its location. The positions of spatial objects are determined by whether they are in or out of a grid cell. A point object would comprise of one grid cell, while a line would consist of a number of adjacent grid cells. Areas or polygons make up a number of grid cells covering the desired area (*Figure* 4). This approach is identical to the method used in printing photographs where each photograph is made up of thousands of tiny blocks. The image in the photograph is created by having some blocks dark and others light, as is most easily seen in newspapers. Photographs can be printed at various resolutions and different resolutions can similarly be set for a spatial database.

Figure 4. Points, lines and polygons stored as raster or grid information



Note: The point and line in the figure on the left are shown expanded using a coarser resolution in the right figure.

Figure 4 shows how a point and line appear in grids using different resolutions. A high resolution provides greater accuracy since the area is divided into many cells, allowing objects to be more correctly placed in their true positions. Raster databases are often very large, especially if high resolutions are used. Lots of computer memory is then required and the processing of data can take a long time. Satellite images and aerial photographs are often used as sources of data for GISs, and procedures are then required for translating the raster information in these pictures into a vector format (see *Chapter 3*). The individual cells or grid blocks that make up these images are called *pixels*.

Even vector spatial databases can be very large if they hold substantial volumes of detailed data. Of course, the more data you have and the greater the detail, the more useful and accurate a set of data may be. But these data requirements may stretch a computer's performance so that it grinds along, taking long and frustrating periods to obtain results. In principle, it is always good to be economical, only storing and processing such data in a GIS database as are required for the task at hand. Additional data can always be held in reserve, ready for use when necessary.

When building databases of polygons and lines it is always useful to consider the number of line segments that are required to draw boundaries. This is especially true for large areas, such as the provinces in a country, where the boundaries are long and convoluted. The same holds true for rivers. If you were to draw all the tiny bends in the river courses, you would have a very accurate database, with thousands or millions of nodes. If you then combined or over-laid this data with, say, a database of roads along the rivers, you would have very reliable information on how the positions of roads relate to those of the rivers. But you may suffer because these data take a long time to be processed and printed. A compromise, giving you sufficiently accurate information on the position of a river and boundary, would entail smoothing or generalizing some of the minor curves so that fewer lines and nodes are required in the database. This is another reason for recording boundaries between polygons as being shared (*Figure 3*); the system only has to store and process the co-ordinates of one set of lines.

The GIS programmes use a variety of techniques to store spatial data, and for most users the details of how and where data are stored are irrelevant. However, one general feature common to many programmes is that they organize the data into layers. A good way to think of these is to imagine a series of transparencies, with one set of map information being printed on each sheet: one for all the roads, one for districts, another for schools, and one for rivers. If you were to put all the transparencies together, lined up correctly so that each one covers exactly the same area of the earth, you could easily count the numbers of schools in each district, and see all the points where bridges might be needed for roads cutting across rivers. Many GISs process data in an analogous fashion, especially when analysing the relationships between different spatial objects. In addition, the analogy of layers is helpful to the user in thinking about how spatial data should be organized. Some programmes call different sets of spatial data 'layers', but many others refer to them by other names.

Points, lines and polygons are the main kinds of spatial data found in all GISs. Some GISs, however, treat several other kinds of objects as spatial data. All are really drawing aids, designed to make map production easier by linking various graphical and spatial objects. *Annotation* sets of data determine how and where labels are placed on maps. For example, an annotation record can ensure that a label for a school will always be placed on the map drawing above and to the right of the point marking the school's position. It may also determine the height, font and colour of the label. *Symbols* are graphics used to symbolize what different objects are on a map: a little square building for a school, a dotted blue line for a river, waves for a lake, etc. In drawing a map, the GIS can plot the size of the symbol in proportion to the object's attributes, for example, a larger square building for a school with high enrolments, and a thinner line for a road carrying little traffic.

Many GISs construct index files to go with the spatial databases. As the name suggests, these files contain information which speed up the sorting, finding and processing of spatial data. These files are seldom the direct concern of the user, but it is useful to know that they may be created by the GIS, and what they are called.

In some countries, national committees and other bodies have been set up to establish data standards which help to ensure that spatial data conforms with certain requirements. These usually stipulate that data should be captured according to predetermined levels of precision and that these levels be clearly described in documentation that accompanies the data. Similar issues include the original source of the data (for example, the date and author of the paper map from which the spatial data were copied), the name of the person and organization who digitized the data, what software was used, and the extent to which all data were actually collected (for example, were all roads and tracks digitized or only those of a certain class). Clearly, spatial data accompanied by good documentation will be used more correctly and for the most appropriate purposes.

Attribute information

Attributes of spatial features are as diverse as are the ways in which they can be thought of and then described. For a school, as a point object, attributes may include all the statistics collected about that school, for example, enrolments, pass rates, levels of teachers' qualifications, numbers of classrooms, and the subjects taught at that school. All the records of all the students that ever attended the school could also be attributes. Through a GIS, you could then select that school and search for and retrieve the records of a certain student that attended the school in a given year. Information collected about the surrounding catchment area could also be held as attributes of a school, for example, the levels of education of parents, projected enrolments for the next 15 years, or information on the availability of water.

The same principles hold for polygons and lines. In an education application, you may have a set of polygons representing census enumeration blocks. For each block, attributes could be generated by aggregating information recorded during the last census: numbers and proportions of children at school, levels of income, estimates of fertility and mortality rates, proportions of homes with children at school but not having electricity so that students would find it hard to do homework. Lines around schools often represent roads, some of which might present traffic hazards to children on their way to and from school, while others might be used as bus routes. In the latter example, you may want to know how many students arrive at school along different routes. The number of children being transported along each route would form an attribute for each road stored as a polyline in the GIS.

While the great majority of attribute information is treated and analyzed as numeric and character data which describe a spatial object, other kinds of attribute information can be used in a GIS. For example, a photograph can be stored in a digital format and then be linked to a spatial object. Likewise, architectural plans of buildings could be linked to a school. In fact, anything that can be stored in a digital format (documents, video footage, pieces of sound, drawings) can be treated as attribute information if it is related and linked in a GIS to a spatial object.

Although it is quite easy to think of all the attribute information that can be linked to map objects, it is very much harder to manage all that information. GISs approach this task in essentially two ways. The first and simplest way is to build all the attribute data into the GIS's own databases. Each spatial object would then have a record listing all the attributes provided for that object; for example, total enrolment, type of school, principal's name, etc. To do this, the data have to be entered into the GIS's database, either manually or by importing the data into the system.

Some GISs have their own proprietary ways of storing data, so once the data are in the system it might not be that easy to change or manipulate them. Another, and more flexible approach available in many programmes, is to simply store the data in a standard database programme and then to have the GIS link with and access these data (*Figure 5*). Examples of these standard databases are dBASE, Oracle, Sybase, Informix, Access. This approach allows ready use of data stored for other purposes and managed by other people. For example, an education department may have a separate financial database listing all expenditures at schools. You could then link mapped schools with that database to create a map showing average expenditures on textbooks per pupil. For the same map, it would be very useful to draw on an external database of population census records to analyze and find areas with very high levels of income around schools where government spending on education is also high. Such an analysis could highlight schools where greater cost recovery could be considered. Connecting GIS spatial data with attribute data in external databases is also useful in allowing one to have different sets of spatial data use the same attribute data. By way of example, separate sets of spatial data on schools and hostels may both require access to a database listing all staff employed by the education authority, so that both staff working in hostels and those employed at schools can be retrieved through the GIS.

Figure 5. Linking spatial and attribute information in a GIS

Spatial information						
Code No.	Co-ordinates	Name	No. of teachers	Grades offered	No. of pupils	No. of class-rooms
1001	x_1, y_1	School A	6	1-5	197	6
1002	x_2, y_2	School B	4	1-4	231	6
1003	x_3, y_3	School C	18	8-12	456	17
1004	x_4, y_4	School D	3	1-4	34	3
1005	x_5, y_5	School E	12	1-7	590	9

Attribute information

Facilities

Code No.	No. of classrooms	Library
1001	6	yes
1002	6	no
1003	17	yes
1004	3	no
1005	9	no

Teachers

Code No.	Name	Qualified
1002	Voigts	yes
1003	Nieuwoudt	no
1003	Chung	yes
1003	Tjombonde	yes
1004	Gunning	no

Note: In the top table, attribute information has been merged into a set of data on school locations (x, y co-ordinates).

The two tables above are separate but linked to the spatial information using a unique code for each school. For example, facilities and teachers in School C are linked and highlighted to the spatial information for that school.

The GIS attribute information stored in databases such as dBASE and Oracle have to obey the rules used by these programmes for data formatting and management. Fields have to be established for each variable, the type of variable (character, numeric, date, etc) has to be specified, as well as the length of the field. Another advantage to using these databases is that other programmes can be used to process and manipulate the data before it is linked to spatial data. For example, data can be exported to a statistical programme to perform multi-variate analyses and the results for each object can then be exported back for use in your GIS.

Attribute data are usually managed on a one-to-one basis with spatial objects, each object being linked with one attribute record, or on a one-to-many basis. The example of drawing on a database of staff at a school would be a one-to-many relationship, a selection being made on one school to find the many people working at that school.

Databases listing attribute data are most robust if you have different files holding separate kinds of information. These are relational databases, links being set between the files so that different kinds of related information can be selected from each file. The links are established using common fields or codes. For a school, one file may list details about all the teachers, each line or record listing a name, and his or her age, experience, qualifications and so on. Another file would list all the subjects and the classes in which the subjects are taught. The teacher file would have a code field which carries a unique code number for each teacher, while the file listing classes and subjects would have the same field giving just the code numbers for the teachers who teach the subjects in each class. For example, the code for Teacher x would be recorded in the three records which show him or her giving mathematics in Grades 4, 5 and 6, and the two records which list Geography classes in Grades 7 and 8. Linking these two files, using the codes for each teacher, would allow you to select the qualifications and experience of all the mathematics teachers in the school, or to list all the names of the teachers that teach Grade 5, or to select all those classes and subjects that are taught by under qualified teachers in the school. What is key in relational databases is that there are fields or codes to link the sets of data, a code for the class and subject, and a code for the teacher in our example. While relational databases are often hard to understand, they are much more efficient than so-called flat files where large volumes of data have to be repeated needlessly. In this example, each record for a subject and class would have to list the name of teacher, his or her experience, etc. so that the same information about the same teacher would have to be repeated for each class and subject taught. Relational database structures are also helpful in forcing one to think about how data should be organized into logical groups to serve the purposes of having the data in the first place.

Geocodes

The key to organizing data on a relational basis is having distinct codes to join or link different sets of data. This brings us to a point that is critical in any GIS: you have to relate your spatial data with attribute information. These relations are achieved by having each object labelled or known by a unique code which can then be used to link it with attribute information. These codes are often called *geocodes* and may be generated by the GIS programme automatically as the spatial database is compiled. Alternatively, and better, the user can create his or her own unique codes which are likely to be more sensible and easier to link with attribute databases elsewhere. For example, the organization (usually a central statistics office) that conducts a population census will have its own unique code for each census enumeration area. Whoever compiles a spatial database of enumeration areas should use that organization's codes to label and identify each census polygon. Likewise, most education departments that collect statistics from schools process and store those statistics using unique codes for each school. Having these same codes in a set of spatial data on schools allows very easy linkage between the two sets of data.

The most important point about geocodes is that they are unique to each spatial object. Much of the functioning of a GIS depends on this, since the spatial information that describes the location of an object is linked to that code. Index files used by a GIS to find and sort spatial data use these codes, the annotations and symbols used to draw objects are linked to the codes, and of course all attribute data is associated with spatial objects using geocodes. Any effort made to ensure that there are no two objects having the same code will be worth it.

Knowing about data structures

Different GIS programmes store and file their data in different ways. Some have a file for each set of spatial data, one for schools, one for roads, and one for district council areas, etc. Others put these data sets into separate directories, while others store them all in the same database file or in what is called a table space. Several programmes use a data dictionary, this being a file that lists all files, where they are stored and what data sets they contain. The dictionary is used to link different sets of data. The details of how each programme organizes its data files are not important here. What is important is that you make the effort to understand how these structures work in the programme that you choose to work with. You need to find out where different files are stored, and what files have to be associated with each other. Look at the data dictionary if one exists, and make sure all the right files are listed. Because of the nature of their processing, GISs often create temporary files which are deleted at the end of a session. However, these files may not be deleted, especially if an error is encountered. It is then wise to delete unwanted files and to ensure that the computer's memory does not become clogged by redundant files and data.

Chapter 3

Developing a Geographical Information System (GIS)

Planning a GIS

The development of a GIS requires that several steps be taken. Some will be big steps, others will be small. And much will depend on how robust and useful you want the GIS to be. What is most important is that careful thought be given at the outset to what *products* the system must deliver. The delivery of these products needs to be stressed all the time, from the initial planning and right through all the development work. Failure to do so may well result in the system being distracted by the very process of building a GIS, or being wrapped up in the technicalities of what different programmes and computers can and can't do. It is much too easy for people working with data systems to become so engrossed in capturing, processing, manipulating, and converting data that products to be delivered by the data fade from any part of the whole exercise. Likewise, computer system managers may spend so much time comparing, acquiring and configuring new machines and programmes that these tools seldom get to do the job they are supposed to do.

Careful and clear thinking about the products you want will tell you what data are required, what computers and programmes must be bought, what people you will want to involve, how much it is likely to cost, and how long it may take before the first products can be expected. It is widely agreed that many, many more GISs are started (the computers are bought and the programmes are loaded) than ever produce results. This is simply because the products are neglected and too many opportunities are given to the system being hijacked by the process of GIS development. A particular danger in developing countries is that donor and development agencies readily supply ideas, computers, programmes, consultants and even data to build a GIS, but no one is present to conceive of and deliver the products.

Accordingly, some questions you have to ask, and make sure they get answered! Why should you want a GIS? What questions do you want a GIS to help answer? What maps do you want produced? What should be on those maps? When do you want that information? Who will use the information? Let's go through some of the steps in asking and answering these questions. Let me also hasten to say that what follows will not, and cannot, give you the answers. All that is intended is that it may help you to think through the answers you require for yourself.

Why should you want a GIS? Perhaps the best material in this book to help answer this is in *Chapter 5* where various GIS applications are illustrated. Answers will be biased by the overall direction you wish to take, for example: micro-planning, centralized planning, management of school facilities, school building or in-service training programmes. Central to all of these interests is a need to look at education from a geographical perspective. I also dare to say that it requires a real interest in the details of education, with a real interest in asking *where* schools are located in the country? *Were* people live that don't go to school? *Where* schools are that need help? *Where* schools really have to be built? etc. A GIS will be of interest to people who want answers to these questions. It will not be of interest to those who don't ask these questions.

Very many education managers simply want maps of schools to know where schools are located so they can get to them. Being able to produce updated maps will be very satisfying to these users. The maps will also have to present an adequate amount of reference material to help orientate the map reader: major roads, town and village names, rivers and other landmarks. Other users will ask how various education indicators vary from region to region, and they will want maps that highlight regional disparities and differences. These people will want to explore variation in a whole range of indicators: cycle times, unit costs, pupil: teacher ratios, qualifications of teachers, enrolment rates, gender disparities, etc. In a similar vein, central level managers may ask for a geographical analysis of education, an *atlas of education*, to inform and stimulate discussion on how education is provided. People responsible for building schools could ask many questions that a GIS can help answer: Where is there suitable land for a school building? What site will best serve the maximum number of children? What water and electricity sources can be used for the school? Is a new school really necessary in light of enrolment changes and projected population changes? Can another nearby school be expanded rather than building a whole new campus? From where are teachers likely to be recruited for a new school? and what are the costs of building on different potential sites? Curriculum planners may ask for maps showing how the provision of different subjects varies from place to place, how schools can be rationalized so that pupils can move logically from one phase of education to another, or how teachers can be transferred between nearby schools to provide better subject choices.

Having decided what you want a GIS to do for you, the next step is to ask how it can be developed. While it can be assumed here that you have to develop your own system, it will be worthwhile checking whether there are existing GISs in your country or region that can give you the products you desire. Examples of people and organizations that might have useful systems are discussed later when sources of geographical data are examined.

The development of a GIS requires that you obtain data for the system, equipment (one or more computers, printers, a digitizer), software, and people to do the work. These requirements are discussed in *Chapter 4*. What is important is that you are able to plan to put these pieces together at the right time and in the right sequence. Logically, the larger and more complicated the system planned, the longer it will take to develop before useful products are delivered. A large system will need a sophisticated computer system and software which might take several weeks to install and configure, and several months may pass before the operators of the system have been adequately trained to use it effectively. Likewise, large data sets will take long periods to collect and integrate. The most rational method is to develop a system progressively, starting small and expanding as the system proves effective and other sets of data can be built into the system. For example, a GIS could start off by producing simple thematic maps showing how educational indicators vary from one region to another. A small computer system, on which people can be trained readily, with simple sets of spatial data showing the regional boundaries and indicator information would achieve this readily. From this point of proven use, other sets of data can be added progressively: a set of point data on the locations of all schools and attribute information for schools from education censuses, polygons showing the boundaries of other administrative units, geographical and attribute information from the most recent population census, information on water supply and other services, etc. Another approach would be to build a comprehensive system for one region and then to expand that to other areas as time and resources permit. Staff development can follow a similar course, starting off with one person to lead and develop the system and adding others when necessary. One major benefit of starting small is that the system's growth can be driven by demand, allowing one to avoid having to justify a large supply-driven system that perhaps does not live up to full expectation.

The first steps will definitely be the most difficult part of developing a GIS. The staff may not know what to do or how to start things effectively, collecting and building the first databases will be troublesome, choosing the right computers and software will be a headache, and so on. One good solution is to simply hire suitable expertise to get everything going. Such a person need not know a great deal about education, but should give the best advice on what to buy, how to get the equipment configured and operational, and how to build the first sets of data. Having such a person working hand in hand with the staff allocated to GIS development will give these people suitable training and give them confidence in taking the system's development further.

Having argued that GIS development should be demand driven as far as possible, it should be recognized that a good deal of marketing will probably be necessary to cultivate an audience and group of users for a

GIS. I say this because most education planners and managers are unlikely to be aware of the benefits and products of a GIS. These people will have to be shown and convinced that much information can be supplied through maps, and that geography adds new and complimentary dimensions to other aspects of education. In all likelihood, it will take some effort to break through the dogma that regional averages are the only useful pieces of information, and thus to encourage an interest in the details of education. This is the same kind of dogma that prevents a ready movement in approach between central planning and micro planning.

Users of GIS information are therefore not likely to come queuing at your door on day One. Rather they will emerge from various sectors as the service becomes known. Two products that will generate interest and demand will be simple thematic maps showing variation in indicators, and 'road-map' type maps showing the positions of schools in each region. In planning a GIS, think about who will be using the system and how they might respond to its development.

A final integral part of your planning process is to consider what data will be needed. Self-evidently, answers to this question will be determined by what issues are of interest, what products are expected, what data are either available immediately or will have to be collected, and what regions of the country are of interest. Assembling data into databases is often the lengthiest, most frustrating and most costly part of getting a GIS going. The following section explores these aspects in some detail.

Building sets of data

You may ask: Why 'build' sets of data? The concept of building is rather appropriate to what happens in compiling information into a database: the pieces need to be assembled; they need to fit into a design; once put together the pieces often mean much more than they do individually; it often takes a fair amount of time to collect and fit the pieces together; the process has to be done systematically; and a final structure must be kept in mind. As in a building, the GIS developer will be dealing with different classes of building material or data. For our purposes, these fall in two broad groups: the sets of spatial data (school locations, census enumeration areas, roads, and districts, for example) and the sets of attribute information (for example, demographic statistics, enrolments, teachers's qualifications).

Sources of spatial data

There are five main ways of obtaining data to build a spatial database: getting the data in digital form from someone else, digitizing maps, scanning maps, translating satellite images and aerial photographs, and by using global positioning systems (GPS) to record co-ordinates on the ground. The first method is used preferably because given the costs of collecting your own spatial data using the remaining four techniques, it is always advisable to spend a good deal of time investigating what data might already be available. There are no clear guidelines on this, but the longer you look for data the more of it you are likely to find. GISs are developing at a rapid rate in most parts of the world, and many organizations now have information that is relevant to the needs of an education GIS. District, region and other administrative boundaries are often available since they provide reference points for mapping work in many different sectors. For similar reasons, the positions of towns and villages, major roads, railway tracks, power lines, water canals and pipelines, rivers, and lakes are often available in a digital or GIS format.

Sources for these data will vary from country to country. The national survey office is the logical first call to be made. GISs have developed earlier and at a faster rate in the environmental fields, so organizations such as geological surveys, wildlife conservation departments, meteorological services, water supply agencies, and agricultural and forestry departments will often be valuable sources of spatial data. People operating GISs in these organizations will also know of others doing similar things. For population census information, the national central statistics office may have maps of enumeration areas, health departments should have digital maps of clinics, and transport departments may have spatial databases on the roads in the country. Non-governmental and commercial organizations should not be missed, for example town planners, land surveyors, geological exploration companies, and civil engineers. In larger centres, computer vendors selling GIS software often know about sources of data and may even have some data available themselves.

In many cases, spatial data provided by other people and organizations may not be exactly what you want. The degree of resolution may be too great or it may be too coarse, or information from one part of the country may be missing, for example. Nevertheless, these data are often useful in getting you going and in providing your maps with background reference information. Depending on the organization, you may have to pay for the data and even sign an agreement which restricts the use of the data to certain conditions and for particular purposes. Because spatial data are often costly to collect and compile, many people are now investigating how these data can be subject to copyright regulations and how the data can be used commercially. Swapping of data is also done routinely, such that one organization exchanges a set of data for another that it wants. If you are developing a set of spatial data on individual schools in your country, you may think of swapping it for other data. Few other people are likely to have digital information on the positions of schools.

Chapter 4 discusses some points which provide guidance on the choice of GIS programmes. One of these concerns the compatibility of data created by different programmes. The worst thing is to have a system which cannot use data produced elsewhere. While many programmes have data conversion facilities, some data could get lost or be corrupted, and a substantial amount of effort may be required to do the conversions. So, it is most important that the data be provided in a format that your system can use directly or readily convert into a useable form. Essential points to remember are that you know the data formats your system can handle, what conversion possibilities exist, and that you enquire what data formats can be produced by the system providing the data to you. Standard and widely used formats used for moving graphical and spatial information between GISs include DXF (digital exchange format), DLG, and ArcInfo 'e00' file formats.

Because spatial databases vary greatly in their degree of precision, it is wise to have a good idea of what kinds of resolution were used when the data were compiled in the first place. It is always possible to change the data from fine to coarse levels of precision, but the reverse can't be done. Having coarse data is useful because computers can process the information and print maps more quickly. A coarse set of data is fine if you only want to have rough thematic maps to show how indicators compare in different regions of a country. That database will, however, be unreliable if you want to superimpose the positions of schools and be sure that every school is correctly mapped in the region of which it is an administrative part.

As a last step in securing data from another source, request as much documentation as possible to provide information on the precision, map sources, authors, and extent of coverage of the data. This information may be available in terms of national standards established in the country in which you are working.

Digitizers are the most widely-used method of collecting raw spatial data. These tools enable one to trace features on maps and then transfer the co-ordinates of the traced information to a computer. An electrical field of conductors under the surface of the digitizing tablet provides a digital grid of co-ordinates. A map is mounted on the tablet and the position of any feature on the map can be recorded in relation to the co-ordinates of the underlying grid. This is done by pointing a stylus at the feature to be copied and then pressing or clicking the stylus. The tablet and computer record the exact co-ordinates at the point at which the stylus was pressed. Point objects are recorded by a single click to record their positions, while a line is traced by moving the stylus along the line and pressing the stylus whenever the line changes direction. To trace an area or polygon, the stylus is moved successively along the area's perimeter ensuring that the last node is at exactly the same point as the first node. The number of nodes or points digitized is always dependent on the judgement of the person doing the digitizing, since only he or she can decide how often to press the stylus along a curved line. More points will give a smoother and more accurate curve, while fewer points mean less data to process, less time spent doing the tracing and a jagged, less accurate feature.

Once a map has been attached to a tablet, reference points for the co-ordinates have to be recorded. The way in which this is done will vary from tablet to tablet, and will also depend on the software used for digitizing. At the very least, two reference points have to be stipulated, each as a pair of x,y co-ordinates. The most useful co-ordinates to record are those printed on the map itself, perhaps latitude and longitude or another co-ordinate system in metres or feet. These reference points provide a scale for the grid, so that units of distance on the tablet are translated into figures reflecting real distances or positions on the surface of the earth. The points also control the orientation of the map, with increasing values along the x-axis reflecting increasing distances in a due east direction, for example. Needless to say, the map must remain firmly attached in the same fixed position on the tablet for the reference points to continue providing the established scale and orientation.

The size of map that can be digitized depends on the size of the tablet, but even with the largest possible tablet you will, in all likelihood, be digitizing several contiguous maps or sections of maps. Reference points using the same co-ordinate systems and units become most important here, because features on adjacent maps will have to be correctly aligned and then joined. For example, roads and regional boundaries that transect one or more maps will have to be matched-up and joined correctly. This will be difficult to do if they have not been digitized using the same scale and orientation. GIS programmes usually provide various tools for joining features and other tasks requiring the editing of digital features – these are discussed below. Likewise, information traced from different maps will have to be transformed to the same map projection and co-ordinate system. Even though adjoining maps are usually drawn using the same map projections, their base or standard lines of latitude and longitude will be different. Conversion programmes or utilities that accompany many GISs are used to translate data from one co-ordinate system to another.

Digitizing maps is hard work, requiring days, weeks or months spent bending over a tablet, carefully tracing the map features. The time spent digitizing will depend on the extent of features that must be copied and the level of precision used in copying those features. It is best to set aside a large block of time for any single digitizing session so that close attention can be paid to where you are on the map, what features have been done, what elements have to be re-done, and what can be done during another session. It is also advisable to digitize all features that belong to the same class in the same session. Many programmes allow you to assign all the features traced to the same layer, so you can capture all the roads and have them classified as roads immediately. Having completed all the roads, for example, you can then move on to tracing other features on other layers.

Because digitizing is time consuming and thus expensive, and depends upon the operator's discretion in deciding how precisely features are digitized, many people now use scanners to capture information off maps. These machines are similar to photocopiers in that a light beam is used to detect and record features of a particular colour or darkness on a sheet of paper. Instead of translating this information into an image on another piece of paper, scanners record the information in a digital, raster format which can be stored by a computer. A scanned image of a whole map does not provide any intelligent information on what the digital features are. For example, a scanned line could be a road, river, contour line or regional boundary. Solving this problem requires that you either systematically go through the digital image on your computer classifying the features appropriately, or that different layers of the map be scanned and classified separately. The latter solution requires that copies of the map must be available, each depicting a different feature: one for roads, one for rivers, etc. These copies are usually available as transparent films which are used to provide separate colour layers for the original printing of maps.

Satellite imagery is a rapidly growing source of spatial data. This growth is due to the improving quality of images, because new and better programmes are becoming available to interpret the images into spatial data, the costs of satellite images are becoming more reasonable, the growing numbers of people who realize the value of the images, and because they provide recent information that is not available from other sources. Satellite images are available as printed images and as digital information, the latter usually being stored on tapes because of the large volume of data. These are raster data, each grid block or pixel having a certain colour or reflectance depending on the nature of the object represented in the pixel. The size of each pixel varies according to the resolution used by the satellite imaging system: each pixel represents an area of 10 by 10 metres in some systems, while pixels in other systems depict larger areas. An area of grass will be depicted on the pixel using a different colour or wave length from that representing a road, or houses, or water, or different kinds of tree species, etc. Image interpreters are computer programmes that scan through the image, recording the wave length in each pixel, and classifying the pixels accordingly. The programmes also aggregate those adjoining pixels with identical wave lengths into polygons or lines deemed to represent the same spatial features. This is as much as the programmes can do, and people doing the image interpretation then have to assign attributes so that certain lines are indeed roads, others are rivers, some areas are housing developments, while others are identified as forests dominated by a particular kind of tree. This latter process is called *ground truthing* and requires that the attributes of objects in the images be verified by visiting them on the ground or using some other method of positively identifying what they represent.

The use of satellite imagery as sources of spatial data is likely to increase, especially in developing countries. This is because many large areas in these countries have not been mapped recently. Reliable information on where people live is therefore not available, and rapidly growing human populations mean that new settlements are likely to arise and remain largely unknown in the absence of current spatial data. The production of maps showing changes in population densities and distribution has obvious and important implications for education planners. Aerial photographs can be processed using similar procedures to those used for satellite images. However, aerial photography is expensive, covers smaller areas, and is not done as routinely.

The final method of collecting data is to actually visit the feature on the ground and to then record its co-ordinates, usually in degrees of latitude and longitude. This is the traditional method of surveying areas, the positions of features being determined using theodolites, survey beacons and trigonometry. These older methods required specialized knowledge and equipment. Fortunately, in recent years global positioning systems (GPS) have become available as tools for obtaining rapid and accurate co-ordinates. GPSs receive radio signals transmitted by a group of satellites orbiting the earth, decode the signals and provide a calculated set of co-ordinates. The co-ordinates are displayed on a screen from where they can be recorded for later addition to a set of spatial data. Alternatively, the co-ordinates can be stored in the GPS memory, along with codes identifying each position, and then loaded directly into a spatial database. Some GPS can store thousands of such points, allowing the user to map very many features in any one session. A GPS can continually record the position of a car as it drives along to map a road. For schools in remote areas, GPSs provide accurate co-ordinates that would otherwise be hard to obtain.

Cleaning and building spatial data sets

While some of the sources of data described here may provide you with complete spatial data which can be loaded and directly used by your GIS, the other methods usually require that the data be further processed before they can be used in a spatial database. This is because digitized and scanned data, satellite image data, and pairs of latitude-longitude points collected using GPSs all comprise simple graphical data determined by sets of co-ordinates. There is no intelligent information to go with them necessarily, nothing that classifies them as points, lines and polygons in a spatial database, and nothing to give each object a unique code. Steps must therefore be taken to ensure that each object meets the requirements of one of these three types of features.

The details will vary according to whether you digitized or scanned maps, or used image processing on satellite images, or recorded co-ordinates using a GPS. However, each object will have to be processed so that the GIS recognizes it as the object you intend it to be. Raster data from images have to be translated into vectors if your programme requires that format. Points must be single pairs of co-ordinates with a label, lines must be a set of nodes and internodes joined to each other, and polygons must be closed. Different sets of data will have to be transformed into the same map projection and data from adjoining maps must be joined together. Some of the data 'errors' that have to be corrected are illustrated in *Figure 6.*

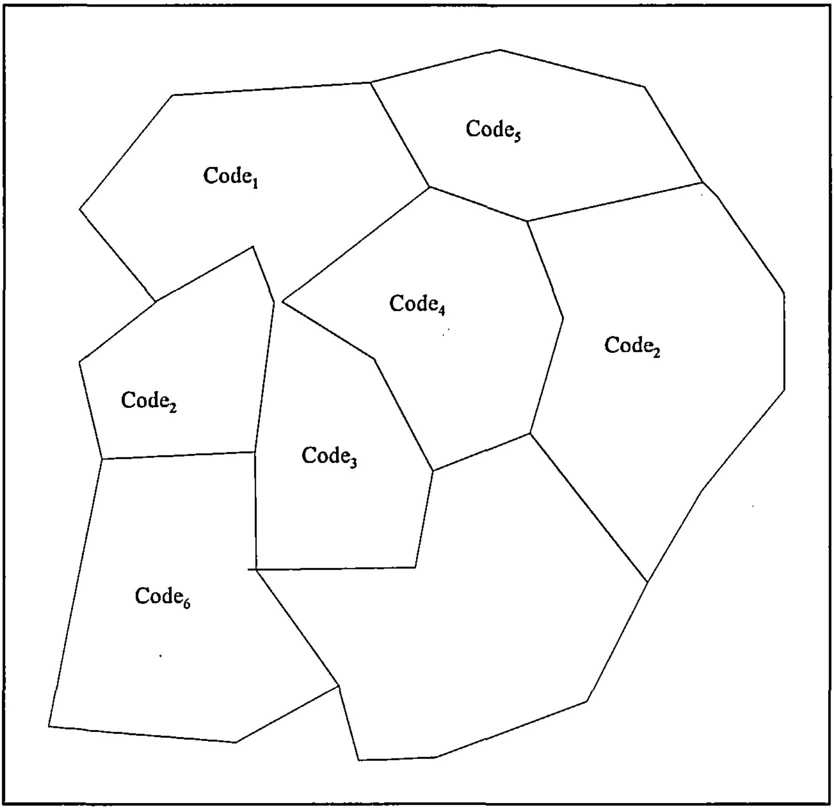
Lines that 'overshoot' the points at which they either intersect or join other lines must be eliminated, and lines that should join at a node must be corrected to ensure that they indeed join.

Most GIS programmes have utilities which help identify and even fix these errors for you. For example, any gap between two lines can be closed by the programme if the gap is narrower than a given distance. Imagine a digital drawing with units of measurement in metres that covers an area of several thousand square kilometers. Since an error of 5 metres is likely to be unintentional, you could have the programme correct the error without the correction having to be verified. Accordingly, the GIS utility can scan the drawing for lines or gaps of less than 5 metres and delete all the overshoots or close all the gaps, as the case may be.

Once a digital drawing or map is complete it can be loaded into the spatial database. This involves the GIS scanning through the drawing and processing objects that meet the desired criteria. For example, you might have a drawing of polygons and the programme will systematically scan through the drawing, recognizing and storing each polygon that has its lines completely enclosing an area. Many programmes allow you to use text items on the drawing to label and, therefore, code the objects uniquely. Point objects can be shown as a text item with the origin point of the text (usually the left hand, bottom corner) being the actual position of the text. Similarly, text items can be drawn inside polygons and the programme then *picks-up* each text entry as the label for each polygon.

Another major step to be accomplished while creating a spatial database is to ensure that all objects are correctly classified, so that related or similar objects will be loaded into the same database. Most GISs store objects of the same kind in the same layer, theme or database. You thus need to have all roads classified as roads, making sure that none of the lines loaded into the roads layer are actually telephone lines, for example. All schools, perhaps drawn as text items on your digitized drawing, must indeed be schools and not towns or clinics. Great care must also be taken to ensure that each object is appropriately and correctly labeled with its own unique geocode. Typical errors include giving two objects the same label, giving an object a label that does not fit correctly into the length of the field assigned for codes, or neglecting to label one or more objects. These errors are often difficult to identify and correct, so attention to these details will be well worth the effort.

Figure 6. Some typical errors in spatial data. There is a gap between two polygons ($Code_1$ and $Code_2$), there is a dangling line in one polygon ($Code_6$), one polygon is missing a code label, and two polygons have identical labels ($Code_2$)



Each GIS programme has its own requirements for cleaning data and for loading spatial objects into its database. It will pay you well to spend substantial time going through the manual and whatever other instructions there might be to help get these important processes done. Failure to do so will result in an incomplete or irrelevant database, causing you to spend more time going over the steps again to get it right.

Sources and management of attribute data

From a GIS point of view, much more attention is usually paid to collecting and building sets of spatial data. While this bias is to be expected, much of the spatial data will be rather useless in the absence of good information about the spatial objects. Having good attribute information will increase the value of your maps and spatial analyses substantially, and, by good, I mean that the information should be accurate, current and relevant to the results that the GIS is intended to produce.

For education planning and management, the richest sources of data will be from schools and other educational institutions, and from information collected during population censuses and surveys. The former are usually controlled by departments of education, while demographic. surveys are usually run by central statistics offices. Many countries conduct annual education censuses, a detailed questionnaire being sent to each school on which the school staff is required to report information on enrolments, numbers of teachers and their qualifications, the number and condition of classrooms, the ages and gender of students, the curriculum taught at the school, pass rates, numbers of repeaters and graduates, the provision of services such as electricity, water and telephones, etc. In processing these data, each school is usually given a unique code number so that different sets of data can be stored and then joined in relational databases. Much of this data can be usefully linked with spatial data to examine geographic variation between schools, and *Chapter 5* provides examples of the kinds of results and insights to be gained from this approach. The easiest way to ensure that these linkages can be made is to use the code numbers given to the schools in the census database as the geocodes for schools in the spatial database. Any data collected during a census can then be immediately linked to your spatial data.

School census data generally provides information on what is happening in schools, while population census data provides information on what happens in areas surrounding the schools, for example, on the socio-economic characteristics of parent communities, demographic trends affecting enrolment growth, and numbers of children not at school. The form in which population census data may be available will vary from country to country, depending on the formats used for processing the data, confidentiality requirements that might prevent people outside the central statistics office from obtaining the data, and the geographical units of analysis. From an education perspective and for the kinds of analyses that should be done in the education sector, it is ideal to have spatial data on the distribution of enumeration areas or census localities, and then to have the attribute data available in as disaggregated a form as possible. This will allow fine scale analyses and mapping to be done around schools and other areas of interest. The principle here is that you should try to retain as much detail in both the spatial and attribute data. It is always possible to aggregate the data into large units of analysis as and when these bigger areas have to be considered.

A variety of other kinds of attribute data will be of interest to education planners and managers. Information on administrative structures in different areas should be available, for example, the school inspectors responsible for different areas, regional governing bodies, and local leadership structures. Relationships between education and other social services, for example, clinics, *crèches*, youth centres, vocational training establishments, water supplies, telephone services and electricity supplies, can be examined by obtaining relevant attribute data and linking these to your spatial database.

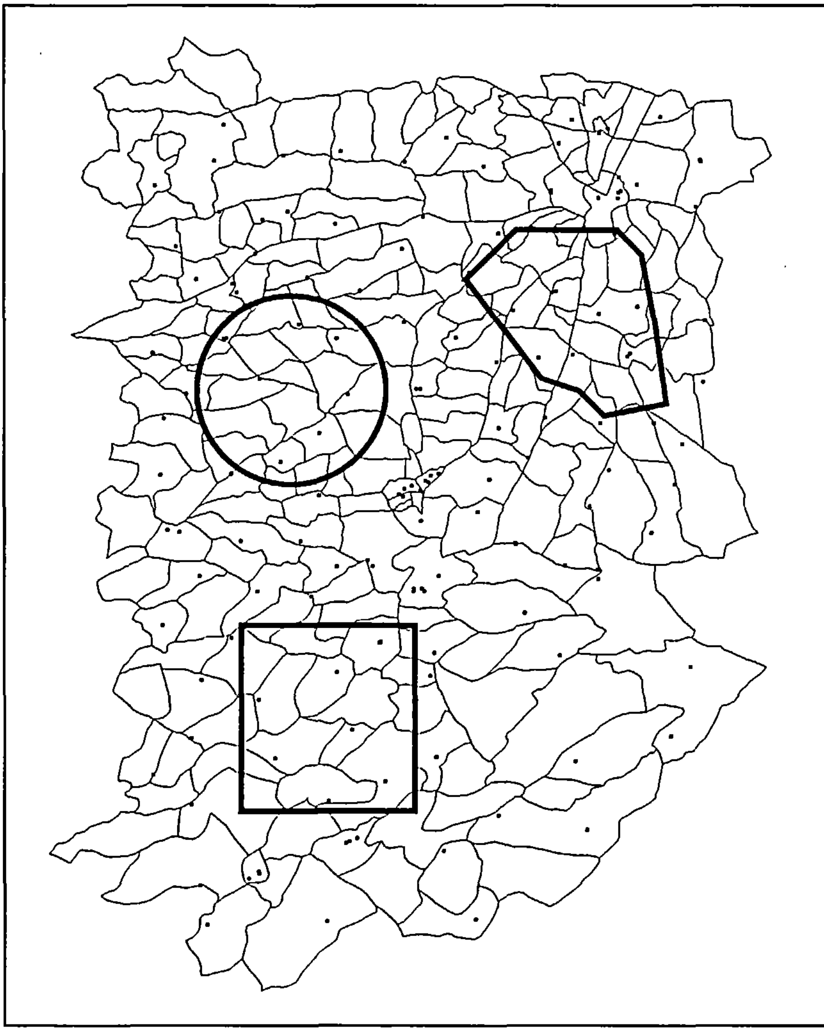
From a management point of view there is likely to be one critical difference between spatial and attribute data. This is that different groups of people will probably be responsible for collecting and building the data sets. In a well-organized and disciplined system this should not be a major problem, but difficulties will arise if those who capture and process attribute data see no value in making it available for GIS purposes. Likewise, if spatial data sets are built without paying sufficient attention to the availability and formats of attribute data, much of a GIS's power may be compromised. Close collaboration with those responsible for collecting attribute information is therefore recommended to ensure that appropriate data are available and that the GIS staff understand the formats and analytical value of the attribute data. Perhaps the ideal office to place a GIS and its staff is in the section of an education department that routinely conducts the education censuses.

Extracting and analysing GIS data

A GIS serves no purpose unless information from it can be extracted and analyzed. GIS programmes vary in how they allow you to extract data, but several concepts are key. The first is that information can be extracted either by searching for the data geographically or by selecting attribute information that meets certain criteria. For example, you could select all schools in a part of a region by drawing a selection area on your GIS computer screen and then having the GIS retrieve all schools within that window. Alternatively, you could apply selection criteria to an attribute database to find all schools that offer English, for example, in a particular grade. The GIS would then draw the positions and names of those schools on a map. The second is that combinations of approaches are also used frequently, such that a set of schools can be selected as spatial objects in one geographical area and selection criteria based on their attributes can secondly be applied to retrieve those that meet your requirements. Thirdly, selection criteria can be based on the relationships between different sets of spatial objects. Schools, for example, can be selected if they fall within or outside a specified range around other geographical objects.

In simple selections of spatial objects, the GIS user draws a shape on the computer screen and requests that all specified features within or outside that shape are selected. Options allow for the selection of objects that fall wholly within or outside the area, or objects that are both within the area and that have boundaries crossing that of the selection area (*Figure 7*).

Figure 7. Selection areas drawn on a map of schools (blocks) and census areas (polygons) in the shape of a window, circle or polygon. Spatial objects inside or crossing the boundaries of these shapes are selected



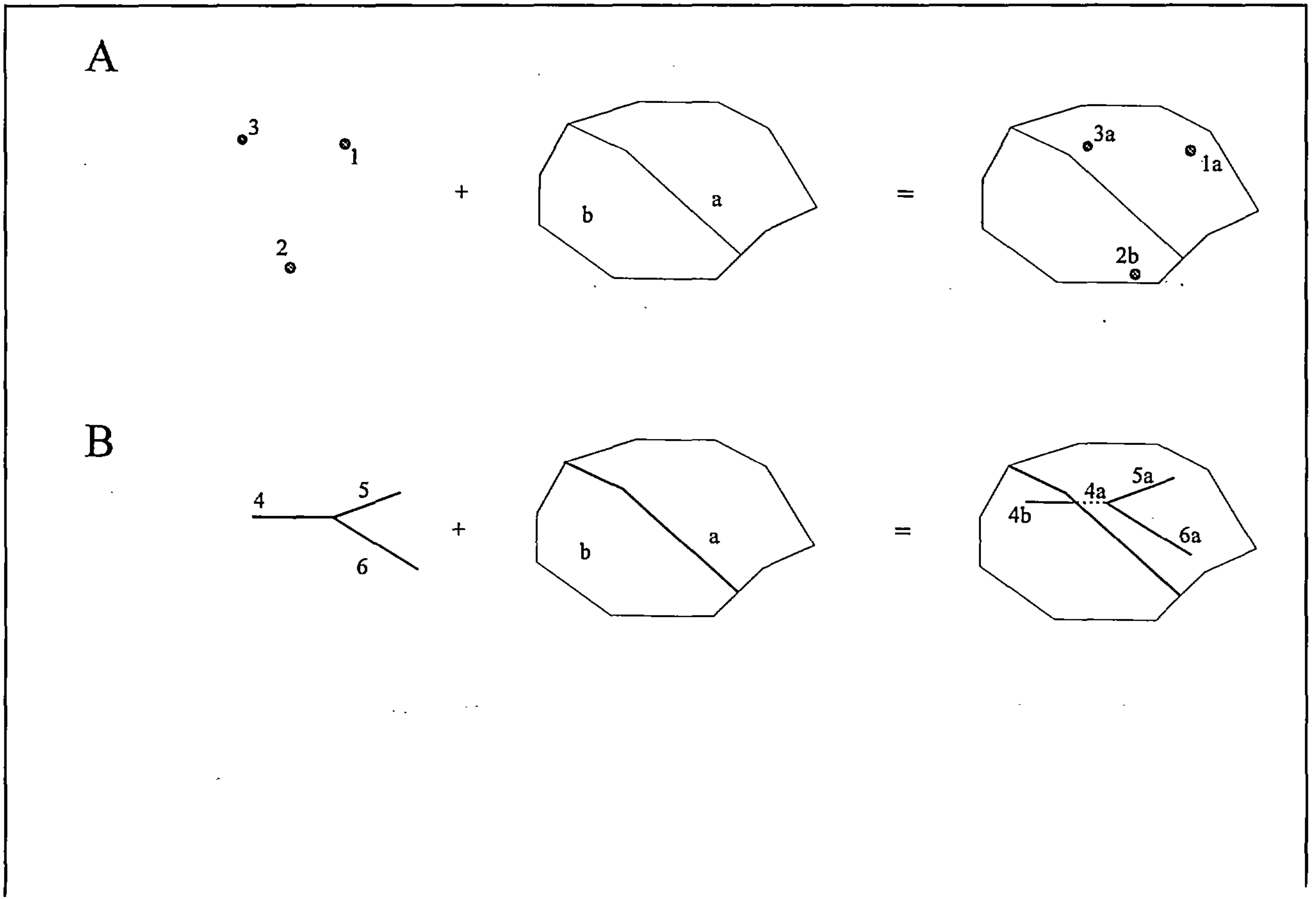
Note: When using 'inside' selection criteria, only those objects that are completely within the selection area are retrieved. 'Crossing', however, also selects those objects that have their boundaries crossing the edges of the selection areas.

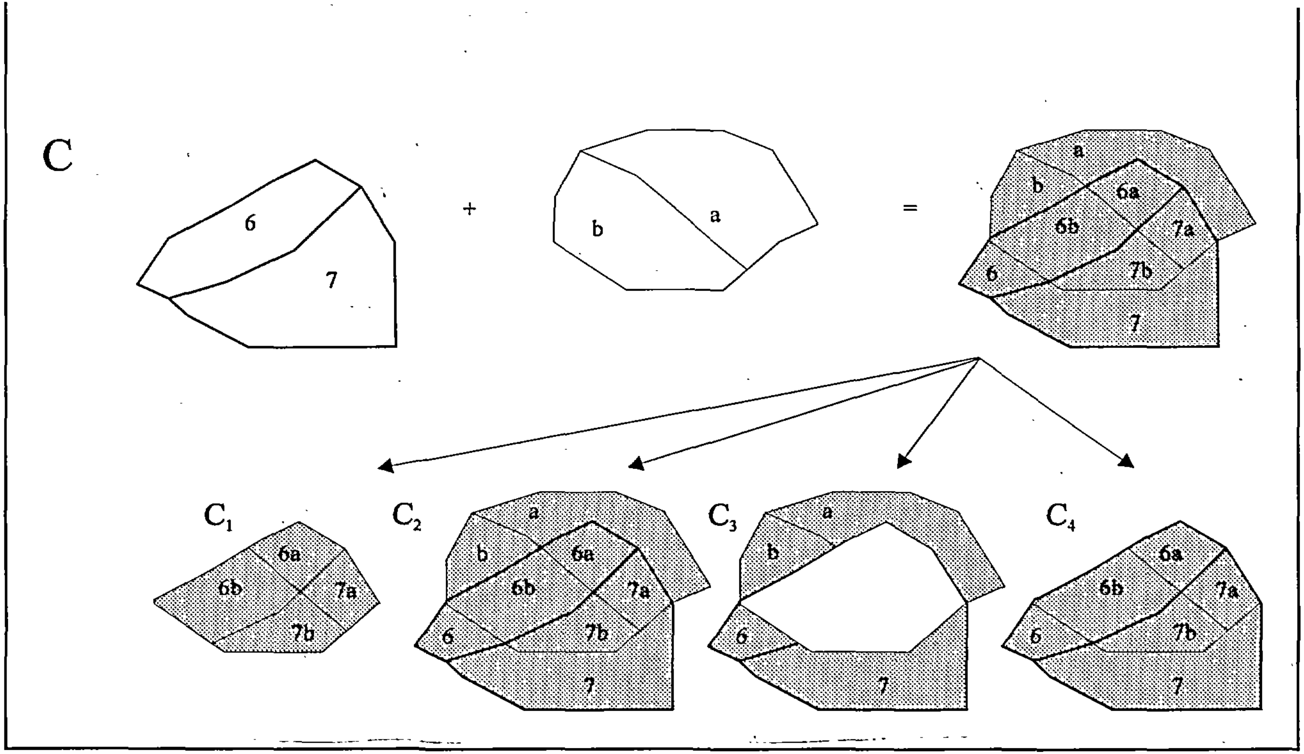
The area used to select objects can be any size on the map and usually conforms to one of three shapes: a rectangle, circle or polygon. Commands available within the GIS allow you to draw any one of these three shapes to specify the area you are selecting. Having established your selection area, the programme then compares the co-ordinates of that area with those of objects in the spatial database and then returns what is called a *selection set*. This is the set of objects selected according to the criteria stipulated. Selections generally work on only one kind of spatial object at a time, so the user has to specify whether schools, roads, windmills, or clinics, for example, are to be searched for in the spatial database.

Once a selection set has been established, it is available for a number of uses. It can be displayed on the screen, and printed or plotted. Statistical and spatial analyses can be performed on it, graphs can be generated, and it can be copied to another file or spatial database. Further selection procedures can also be applied to retrieve what would then be a subset of this selection set. These additional selections can either be done spatially by refining the geographical area of interest or by using attribute criteria.

Selections using relationships between geographical objects take a number of forms. Distance criteria are established between objects in the same way as you draw a rectangle, circle or polygon such that objects to be selected are either within, outside or crossing the boundaries of other spatial objects (*Figure 8*). For example, you can select all those schools that are within district mapped as a polygon, or find those schools that are outside the transmission area of radio signals that could be used for educational broadcasting, or find those villages that are within the catchment areas of more than one school. Terms commonly used and employed to describe operations that require analyses of relationships between spatial data include: *intersections*, *joins*, *unions*, *overlays*, *point-in-polygon*, and *line-in-polygon*. These terms vary from programme to programme, and some of their analytical applications are illustrated in *Chapter 5*.

Figure 8. Analyses of the geographical relationships between spatial objects





Note: These involve overlays of points (A) or lines (B) with polygons, or overlays between different polygons (C). The points and lines are then relabelled with attributes to show which polygons they fall in. Note, that a single line can be split and labelled such that segments of the line in different areas are given appropriate labels (4a and 4b). Overlays between polygons can produce new sets of polygons consisting of only those areas that are common to the original polygons (c_1), areas consisting of all the original polygons (c_2), areas excluding those that are common (c_3), or areas belonging to one of the original sets of polygons but split into those units with which they intersect (c_4).

Attribute information can be queried and selected in many ways. As in the case of spatial data, selected attribute records are returned to the user as a selection set. The rules for doing so are the same as those used for standard database programmes. Boolean functions which include, exclude and join sets of information can be established to search for data which meet criteria using combinations of *and*, *or* and *not*. Other criteria select information based on values that are *equal to*, *greater than* or *smaller than* certain limits. Additional fields can be added to an attribute database into which the results of calculations can be placed. For example, you may want to generate an indicator reflecting needs at schools, the value for each school representing a calculation based on a combination of indicators of need. Likewise, statistical programmes may generate variables which reflect membership of a statistical cluster.

Analyses of attribute data depend on the usefulness of the data and the degree to which relevant questions can be asked of the data. What must be made clear is that the data will not produce results themselves, no matter how voluminous the data files and no matter how powerful a computer is used for processing.

Careful thought needs to be given to the issues at hand, what questions are likely to yield the most interesting and topical results. Statisticians often apply a 'shot-gun' approach to their data, tabulating the information in every conceivable fashion. All that results from this approach are very thick books containing hundreds of tables. Very few of the tables can be digested, and even fewer can be understood or interpreted. In generating maps from spatial and attribute data, pause to consider what data is best seen on a map, what users are likely to make of the maps, whether the maps are useful in raising issues, and whether the spatial relationships shown on the maps make any sense.

All of this requires that you understand the spatial, and especially the attribute data. The data have to be probed, graphed, checked, updated. What do these data really indicate, how much do the data vary from school to school, how current are the data, are the data reliable, are better proxies available for the same indicator?

Pay special attention to variation in the data, since variation is often the most interesting aspect to question. Maps, after all, show how features on the earth vary from place to place, and it is the attributes of those features that tell us most!

Chapter 4

Equipment, software and staff

Until a few years ago, GISs really had to run on large, costly computers. The programmes were both expensive and difficult to use. Since the technology and philosophy behind GISs was relatively new, few people were in a position to use or even to understand the advantages of these programmes. Very little map information was available in digital form, so much time and money had to be devoted to the capture and development of spatial databases by those who ventured into this area of work.

Much of this has changed. Relatively cheap personal computers now perform so efficiently that they can run most GISs, there are very many GIS programmes available on the market at reasonable prices, and the programmes are much easier to use. The use of the programmes has become common in many applications and countries, and it is now easier to find and train people to do GIS work. These developments, and the availability of a wide variety of spatial data assembled over recent years, all make it very much easier for an education manager or planner to consider developing a GIS. Nevertheless, decisions on what to buy and who to involve are never easy, especially when you are bombarded by the persuasive talk of competing sales people. Hopefully, information in this will help you emerge from that assault having made the best decisions.

While GISs run on a variety of computer operating systems, the majority are produced to work on personal computers (PCs) running in a DOS or Windows environment. These kinds of computers will certainly meet the needs of most GIS applications in planning and managing education, and the discussion here is aimed at these kinds of systems. Larger computer systems (mini-computers or mainframes) are more often used in natural resource applications where the volume of spatial data is usually many times greater than in the social sectors.

In making choices between the different options, it is important that some kind of development strategy be in place, such that the pieces that are acquired and the steps that are taken fit into a sequence of development that is logical and likely to be successful. Several principles are worth keeping in mind:

Compatibility Equipment and software obtained should be compatible with other equipment and software that might interact with the GIS. This is especially true if spatial and attribute data are to be exchanged between different computer users in an education department.

Upgrading Open pathways should exist for further development of a GIS. Both computers and software are improving at a great rate, so try to ensure that whatever you buy initially is likely to be compatible with future developments. Try to select the likely 'stayers' amongst the many competing computer operating systems and GIS developers. Keep away from anything that remotely looks like becoming redundant.

Local support In the initial stages of developing a GIS you may need support from a variety of people outside your organization; people to help configure the hardware, install and provide training on the use of software, and other, more advanced GIS users who can help with database development. Having access to as much of this kind of help will be important.

The development of a GIS may cost a fair amount of money. Spending a little more money by buying good equipment and software may cost more initially, but it will save time and money in the future. Likewise, being able to hire decent people by paying them a good salary will cost less in the long run because your products will be delivered far more quickly. The examples in *Chapter 5* should persuade you that considerable savings can be achieved by using a GIS to help plan and manage education. Hardware and software costs for a modest system, consisting of a reasonable computer, small colour printer, and an elementary GIS programme would probably amount to less than US$5,000 in most countries.

Hardware

Three main pieces of hardware have to be considered: computers, printers and digitizing tablets or scanners. The former two are absolute necessities, while a digitizer or scanner will only be needed if maps on paper have to be copied. Although the number of computers and printers will depend on the size of the operation, having more than one of these items is always useful in case one breaks down.

In computer jargon, GISs are hungry for memory. The programmes themselves are substantial in size, and spatial databases are often large. Many GISs also create large temporary files on the hard disk while processing spatial data. It is thus advisable that the hard disk storage capacity of a computer be large. GISs also require good quantities of working memory, usually called random access memory (RAM) where most processing of GIS data occurs. The more RAM available, the faster the processing of data and the quicker the results will be available. The speed of the computer's micro-processor(s) is another important determinant of a GIS's performance, and the faster the better. If you think of using a personal computer bought some time ago, ensure that a maths co-processor is available. In essence, it is worthwhile having a powerful computer with plenty of storage capacity.

Another kind of memory is that used for processing of information to and from the video screen attached to the computer. Since spatial databases are often built from maps drawn by a graphics programme (for example, a drawing of lines, areas, points and text items), the computer and screen need to display these drawings clearly and rapidly. More importantly, the majority of products produced by queries and analyses of GIS data are produced as graphics to be assembled into maps for later printing. All of this means that an adequate amount of video memory should be allocated to the video operations of the computer system. A wide variety of video screens and controller cards for video displays are available, varying in terms of their control of colour, resolution of display, size, and the speed with which graphical information is displayed. A colour screen is an absolute necessity because so much graphical information is best viewed in different colours. The resolution and speed of display by a screen should be as great as possible. Larger screens are helpful in allowing more information to be viewed simultaneously. People doing GIS work spend large amounts of time peering at their screens, so high quality screens will improve productivity and reduce eye strain.

The most important consideration in selecting printers is the size of maps you wish to print. Large format printers are much more expensive than smaller ones, but large maps are obviously more impressive and useful. Having a colour printer is really a necessity. Portraying several variables (for example, different kinds of schools, different administrative boundaries) on the same map or showing how indicators vary from school to school or region to region is almost impossible without colour. Dot matrix printers are no longer a useful way of printing and should not be considered for any GIS use. The following kinds of printing devices are now used most frequently for printing maps produced from GISs:

Pen plotters	Use pens holding different ink colours, now becoming less fashionable, but relatively inexpensive for printing on large sheets of paper.
Thermal printers	Use special paper which is heated to produce a range of colours in the printed image.

Ink jet printers — Contain ink cartridges of several colours, the printer head shooting tiny drops of ink onto the paper, mixing the inks to produce a range of colours.

Electrostatic printers — This technology is used by laser printers, an electric current being used to apply an ink powder (toner) to the paper.

The printer you buy will require various consumables, for example, ink or toner cartridges, pens and, perhaps, special paper. Ensure that these consumables are available locally and try to ensure that they will continue being available.

High quality scanners suitable for scanning maps to obtain spatial data are specialized items of equipment. They are comparatively expensive and should only be considered when there is a proven need to copy substantial volumes of map information into a spatial database. Large format digitizers, which are best suited to copying big map sheets, are also rather expensive. And even if there is a need to develop your own databases, you may only need the equipment for a relatively short while. One possibility is to get another organization to do the digitizing and scanning for you. This approach is becoming increasingly popular, but it is important to specify several things clearly, for example, exactly what digital data you want, what co-ordinate systems are to be used, and what levels of resolution are required. In addition to saving on the costs of a digitizer or scanner, you will also save on staff costs that go into the training and then the long periods spent digitizing and scanning maps.

Small digitizers, with a tablet size of about an A3 sheet of paper, are relatively inexpensive and can be used to good effect on small maps. Larger maps can also be copied on them, but this is tedious, requiring that you successively digitize pieces of a map, calibrating and setting control points for each new piece, and then joining all the pieces together into one drawing.

GIS and other software – making the right choices

A large selection of GIS software is now available. The choice of what to buy will be based on a number of factors, but the most important of these must be the products that you want the system to deliver. Bear in mind, also, that your future needs may be more demanding than you currently think. For example, a small system capable of drawing simple thematic maps might seem adequate for present purposes. However, as the system develops and demands increase, you may find that its data formats can not use information from other sources, and it can not perform a variety of spatial analyses. Some GIS programmes are sold as modules, such that a basic system can be used to do many of the simpler tasks while additional units for more advanced processes and utilities can be added as demands require.

This raises a second important issue: try to select a programme that is successful and popular. Software companies doing good business can invest in future improvements and upgrades to their products. New versions of software are compatible with older versions, so those already using the programmes can immediately import their data and benefit from the new developments. Several other advantages go with using programmes that are in wide circulation: you are more likely to find people using the same programme who can help with initial developments, the programmes will provide for standard data formats so you can be more sure of being able to use data obtained from other sources, and you are more likely to be able to find support from local computer vendors. These are all important factors to consider in buying a GIS programme, irrespective of whether a popular and well-known system is chosen.

There are two other reasons for buying a programme that has been widely tested and proven. One is that the programme's use of memory is likely to be more easily configured. As mentioned earlier, GISs make substantial use of a computer's memory. However, both software and hardware require that memory use and allocation be properly set or configured, a requirement often easier said than done. The other reason is that the programme is more likely to have standard drivers to communicate with printers, digitizers, screens and other equipment. Drivers are small software utilities that allow a programme to correctly format the information it sends to or receives from these peripheral pieces of equipment.

Try to check how easy a programme is to use, and where its strengths and weaknesses lie. Many GISs can achieve the same results, but do so using quite different paths. GIS programmes also differ in what they do best: some are very good at producing maps, but are unwieldy if they have to be used for digitizing; some allow easy access to attribute data in a great variety of database formats while others have rather restricted access to other databases; some provide excellent utilities for analysing spatial information but suffer in other respects, and so on. If the people developing a GIS are to be working in isolation from other GIS users and local support, ensure that the programme is accompanied by a good set of training manuals and other reference material.

A GIS programme will obviously be at the core of a development of a set of spatial and associated attribute databases. However, several other programmes provide useful support and help extend the analytical power of a GIS. Having a database programme which uses the same format as that used for attribute data in the GIS is particularly useful. This allows the user to manipulate and analyze attribute information in a database programme, an easier and more robust way than to use the utilities provided by many GIS programmes for the same purposes. Statistical and spreadsheet programmes are also useful for manipulating attribute data.

Used in conjunction with a GIS programme, they add considerable power in making many calculations easy to do.

Many GIS programmes come with a packaged programming language. This is a piece of software that can be used to write your own, generally small, programmes to customize some of the work you do, such as various routine tasks. Advanced users often develop specialized procedures for their own applications. Programming languages can also be used to create customized menu structures, where the items listed on the menus are your sets of spatial data, the names of regions in the country, queries that you often run, and results that must be generated frequently.

A final point concerning software is that you should look out for third-party software developers of utilities and accessories to GIS programmes. These are usually small enterprises that produce programmes that link up with and add to existing, larger applications. Many of the utilities produced by these groups make routine or mundane tasks much easier to perform or, indeed, extend the range of functions performed by the parent programme.

Staffing – who will do the work?

While computers, printers and programmes may involve a fair amount of money, the most expensive component in having a GIS will be the salary costs of those employed to operate the system. Over and above concerns about costs will be the need to ensure that those asked to develop and run a GIS actually do what is required of them.

Processes are often as important as products in education management and planning. GISs, however, are intended to deliver products for education planning and management, and it is therefore useful to involve people who are product-driven by nature. These are the people to ensure that a GIS will deliver what it can. This is not to decry the value of processes, since people will also have to be involved in the process of developing a GIS and others will become interested in the process of looking at education through geographical perspectives.

In addition to selecting staff who are interested in the final products, look for several other qualities. Attention to detail is important, especially in the database development stage when maps are being digitized, spatial objects are being coded, and sources of data are being sought. A sustained interest in what a GIS can do for education will also be important to see the whole development process through and to keep the system running. While a strong technical ability has obvious value, the system should not be led by technical interests alone. This is because technical challenges will not be encountered frequently enough to maintain a desired level of interest.

Even though GISs are now much easier to use than previously, good training and technical support for staff are needed. The concepts and approaches to GISs differ from many other programmes, so it is important to lead people into this way of thinking. Once their thinking and commitment are there, sustain it with encouragement and whatever other rewards can be delivered. Since many education departments are likely to offer meagre salaries, there is a great danger that staff, once trained, will be off to greener pastures. Having two or more people knowledgeable about the system will provide backups in the event that one person leaves. Computers, programmes and databases are often easier to replace than the people who make all these components work.

Chapter 5

GIS applications in managing and planning education

Products that a GIS can deliver for an education manager or planner will now be discussed. The selection presented here is not intended to be exhaustive, but rather to illustrate the kinds of things that can be done with a GIS. Three broad groups of applications will be discussed: those which use maps to display information about education, those that use GISs for analytical purposes to derive information which would not be readily available in other ways, and those that use GISs to provide access to information systems. Some of the examples really apply to more than one category. Many of the illustrations will be based on situations in Namibia, the country where most of the author's experience in developing and using a GIS for education was obtained.

Mapping information

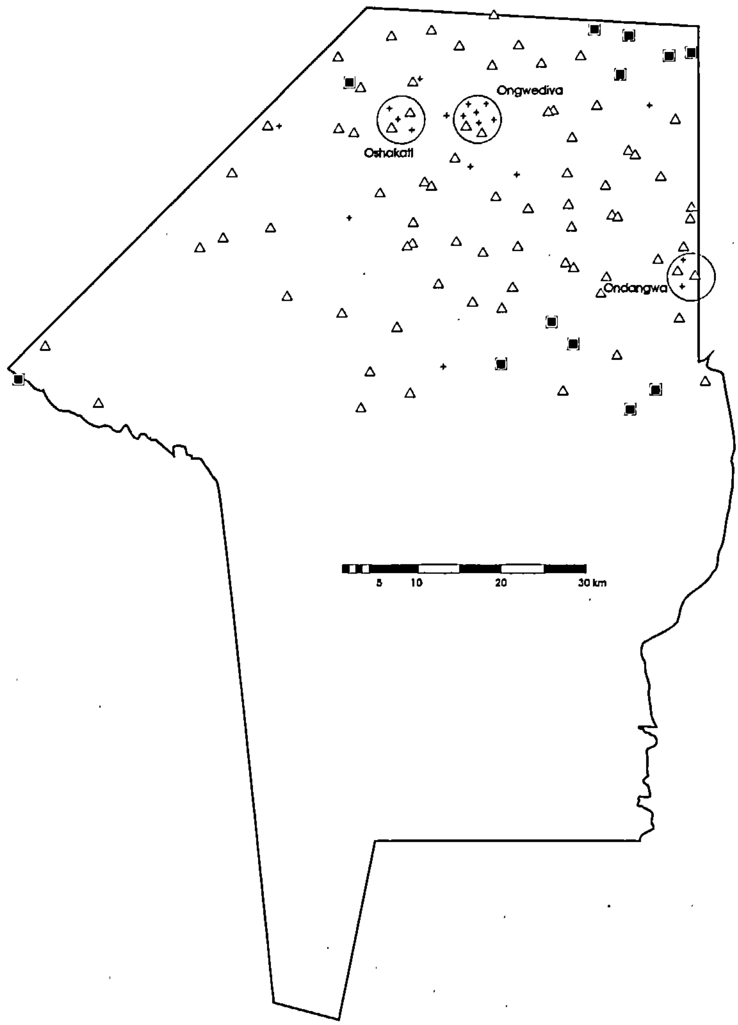
As illustrated, maps are tools for providing information about features on the earth's surface, and, in this particular case, specifically about education. There are, of course, other ways of providing information about education: in articles and books, in graphs, lectures, and tables of statistics, etc. What is useful about maps in contrast to some of these other methods is that people often find them easier to read and the issues easier to grasp. A table can list all the regions in a country, giving precise statistics on how each region compares with the others. However, not many people have the patience and interest to scan all the numbers and then interpret their relevance. The same information on a map, using different colours to show how the indicators vary from region to region, is much more likely to attract peoples' attention. Maps also provide new ways of seeing, understanding and interpreting information.

Amongst the least spectacular but very useful products from a GIS are simple maps showing the locations of schools, their names, and landmarks to guide people seeking those schools. Such maps are seldom available from survey offices or other mapping agencies. The ability to produce these maps is one benefit, but of greater advantage is the ability to produce them to any scale, for any region or area, and for any year, according to the needs and interests of the user. Up-to-date maps can be produced every year as new schools are added to an education system and others close down. Detailed maps can be drawn of schools within tiny areas, such as in towns, showing names of the major streets. Maps of areas used for the administration of education (such as inspection circuits and school districts) or areas used for political jurisdiction can be printed as the need arises.

In the GIS world, the ability to produce customized maps is called Automated Mapping (AM). This usually refers to the production of maps on a regular basis as conditions change, for example, maps produced each week showing the positions of traffic accidents during the previous week. An equivalent process in the education world is to produce customized maps showing the location and names of schools which meet certain criteria. Managers and planners often need to know which are primary and which are secondary schools, which have produced good and which have poor examination results, what schools offer selected subjects in certain grades so that new textbooks can be delivered to them, what schools have yet to be connected to an electricity supply – the examples could go on and on. Given spatial data on school locations and appropriate attribute information on these issues, the GIS user can simply select schools meeting set criteria in a given geographical area of interest and draw the necessary maps.

In many developing countries, specific projects seek to improve or reform conditions in schools, for example, to improve facilities in schools, to provide in-service training for teachers, to increase participation by girls, or to improve school management practices, for example. While the problems that these projects seek to redress may be clear, what is not clear is where the project's activities should be targeted. In many cases, considerable resources are wasted, providing activities and resources to schools in areas where they are hardly needed, while more deserving schools elsewhere remain in need. In these situations, GISs can be important planning tools by mapping indicators of need to show just which micro-areas require assistance the most. One example of this approach is shown in *Figure 9*, where levels of qualifications of teachers at schools in one area of Namibia are shown. The map allows one to see where teachers are poorly qualified, making it reasonably clear where effective in-service training programmes could be delivered. The map also shows some of the elements that might be important in deciding where teachers' resource centres could be placed so that a reasonable number of unqualified teachers could have access to these centres.

Figure 9. A map of schools as points in northern Namibia showing the average levels of teachers' qualifications at each school



+ Teachers have more than 2 years of professional training on average

△ Teachers have 1 - 2 years of professional training on average

▣ Teachers have less than 1 year of professional training on average
(Note: These are averages for all teachers at each school)

Another example is shown in *Figure 10.* The project required that US Peace Corps Volunteers be allocated to clusters of schools, such that each Volunteer could provide in-service training to 20-30 teachers teaching Grades 1-4. These teachers had to be within a reasonable distance (about 5 kilometres) of where the Volunteer would be placed, and should be teaching in schools that were deemed to be in need of support, as gauged by a composite indicator which reflected the teachers' qualifications, class sizes, physical conditions of the schools, repetition rates, and text book availability. A statistical programme was first used to derive this indicator for each school from school census information, and a map was then produced to show the numbers of teachers and indicator value for each school. An accompanying scale allowed one to immediately see clusters of schools which were 'needy' in terms of the indicator and had the desired number of teachers to be trained. Proposed clusters could then be drawn as circles on the map. Finally, local school inspectors, familiar with the schools and terrain, were given maps showing the proposed clusters with the request that they confirm or reject the suitability of the proposals.

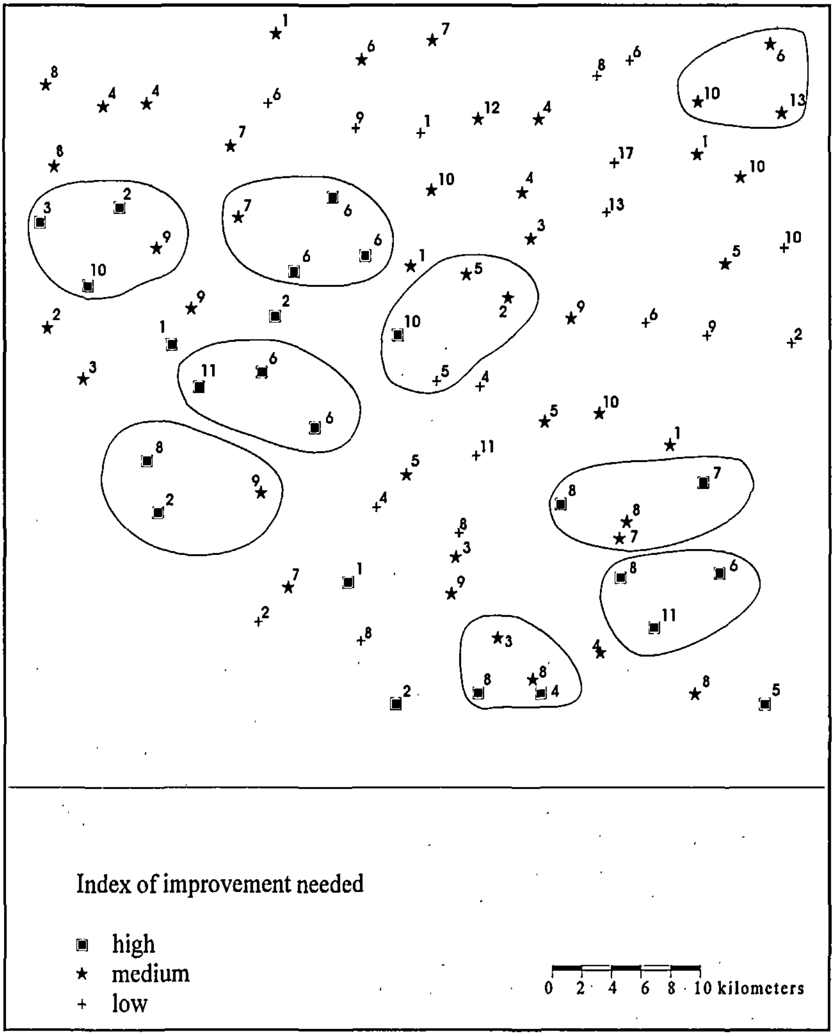
Since the schools had been accurately mapped using global positioning systems and the attribute data were quite current, most proposals derived by this desk study were found to be viable. The region in which the Peace Corps Volunteers were to be placed was about 100 by 300 kilometers in size. Some 600 schools were in this area and all were potential candidates for having Volunteers placed nearby to assist in upgrading their teachers. Twelve Volunteers were allocated to the region, so 12 clusters of schools had to be found. In the absence of a GIS, placing the Volunteers would have been much more difficult.

One approach would have been to simply ask local education administrators to identify schools meeting the desired criteria. It would then have to be assumed that all schools were known equally well by the administrators, such that remote schools were not excluded by virtue of their locations or conditions in them being poorly known. The chances of this assumption being met were clearly small, given the size of the area, the rural conditions, and number of schools.

Another approach would have been to have had a team of people visit all schools in the area, systematically collecting information and assessing viable clusters. This would have been expensive, costing many times more than the use of a GIS to achieve a comparable result.

These two examples have used individual school information, mapped as points, to show how indicators vary from one school to another. From here it is easy to see how a GIS can be used to produce maps for micro-planning or school mapping purposes.

Figure 10. Map of schools using different block symbols to reflect different measures of need for improvement



Note: Numbers next to each school are the numbers of teachers teaching lower primary grades at the school. Each polygon encloses a number of nearby schools in relative need of improvement, with between 20 and 30 teachers, and close enough to be served by one Peace Corp Volunteer.

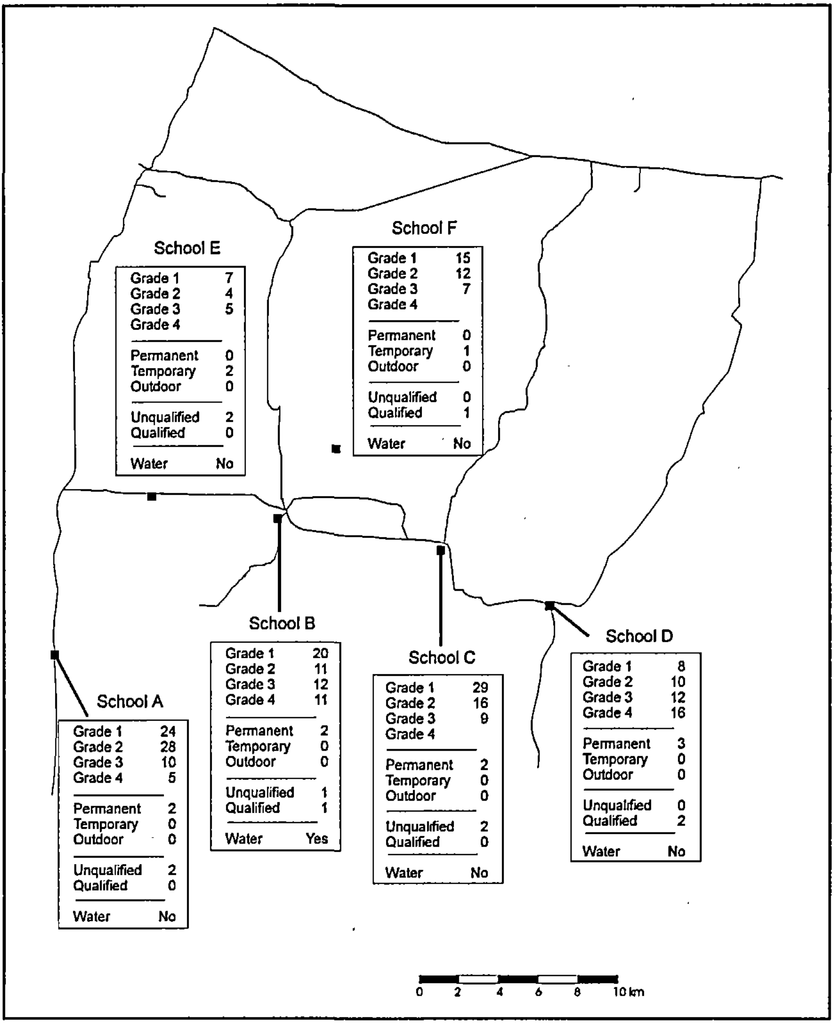
One particular difficulty in Namibia is that many schools do not offer a full range of grades so that pupils are often unable to complete one phase of their schooling. Schools very close to one another often offer the same range of incomplete grades, and there is often no school in the area to which a pupil can proceed to complete higher grades (*Figure 11*). By drawing on statistical information about schools, maps can rapidly be drawn to show what grades are offered, enrolments in each grade, classroom availability, teachers' qualifications and other kinds of information that can be useful in suggesting changes to rationalize the provision of education in a small area. In *Figure 11*, for example, two schools have very small enrolments (E - 16 pupils, F - 34 pupils) and both could be closed if the pupils moved to school B. The teachers at these schools could also be moved to alleviate shortages of teachers elsewhere. The range of grades offered at School B could be extended if one or two teachers were moved there since it already has two permanent classrooms, has a water supply, and offers Grade 4. Moving pupils to that school and adding additional classrooms would create more demand and make the provision of Grade 5, and Grade 6 and 7 later, viable.

This approach to planning requires that the education planner diagnose conditions in schools and surrounding areas on a local basis, deriving from that analysis proposals for a better organization of the school system. Using a GIS, the work in preparing school maps can be speeded up considerably and often automated. Information showing estimates of projected enrolments in areas surrounding schools can be added by overlaying polygons, as can information on transport, water supply and other logistical facilities.

In *Figure 12*, information for schools in each quarter degree square of latitude and longitude has been aggregated by overlaying the positions of schools on a grid of polygons, as shown in *Figure 8*. The objective in this case was to identify areas in which enrolments in Grade 1, the first year of schooling, have increased or decreased during the past five years. Such an indicator provides an estimate of changing levels of demand for education in different areas. Schools in those blocks showing a recent decline in Grade 1 enrolments are likely to see declines in the number of pupils in higher grades in the years ahead. This method of aggregation was chosen as an 'unbiased' way of averaging information for nearby schools. Each quarter degree square is 15 x 15 minutes of latitude and longitude, or about 730 square kilometres in area.

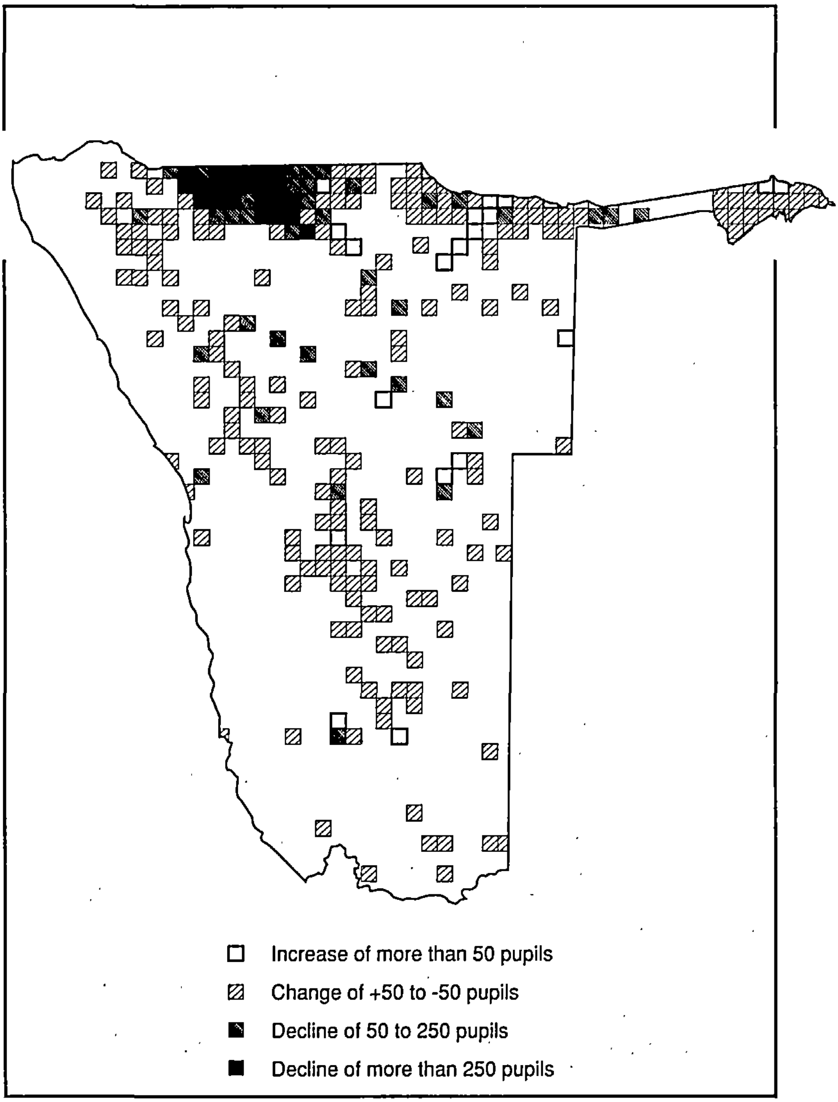
Since these lines do not follow any political or administrative boundaries, the aggregated values are less likely to be interpreted in terms of political or administrative effects. These effects are one of the major fallacies associated with regional averages where the average value for each region is usually implied as being the consequence of a policy specific to that region. Since there is usually considerable variation between schools in a region, a regional average can often be more artificial than informative.

Figure 11. Information that could be used to suggest curriculum extensions and rationalization



Note: Each block of text provides enrolments in each grade, numbers of classrooms (permanent, temporary or outdoor), numbers of unqualified and qualified teachers, and information of whether the school has water. None of these schools offers a complete curriculum. The thin lines are roads or tracks and the square blocks are schools.

Figure 12. Changes in Grade 1 enrolments from 1991 to 1995, per 15 x 15 minute latitude and longitude grid block



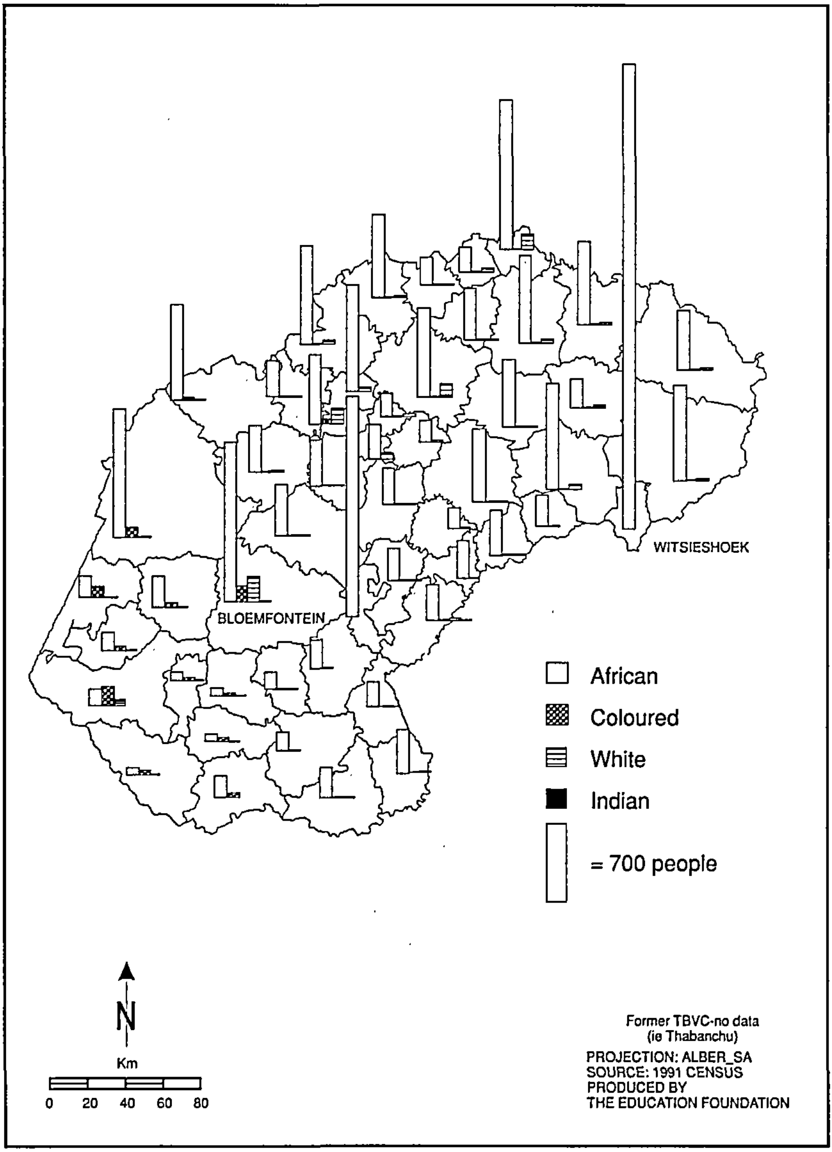
The analysis used to produce *Figure 12* used point-in-polygon intersections of school locations (as points) with polygons marking the boundaries of the grid blocks. The process simply involved 'overlaying' the schools and grid blocks and then classifying each school according to the grid block on which it lay. The same could have been done using two sheets of transparent paper, one showing the schools and the other the grid lines. By aligning the sheets so that they covered the same area, you could count and classify the schools in each block. The utilities provided in a GIS make these kinds of processes much easier to do, and with much less chance of error.

Although regional averages often carry with them great dangers of misinterpretation, it is hard to avoid using them when trying to summarize a volume of statistics. Conventionally, this is done by publishing tables of statistics, each row or column giving the average (or some other descriptive statistic) for each region. Graphs are also used to good effect to compare conditions between regions or districts. A more recent approach is to use GIS maps to make these comparisons. Compilations of such thematic maps, showing how a wide range of indicators varies from region to region, have been published for France and South Africa, and perhaps for other countries. Because many people relate more easily to a map, these Atlases provide a more compelling and informative picture of how conditions vary from area to area. Equity is the major issue that these maps confront by highlighting disparities between regions using different degrees of shading or even small pie or other graphs for each region (*Figure 13*).

The assumption of uniformity in regional, thematic maps is more problematical the bigger the region, and maps showing how these assumptions are violated for large regions often provide quite new ways of seeing education. They draw attention to variation and processes that would not otherwise be obvious.

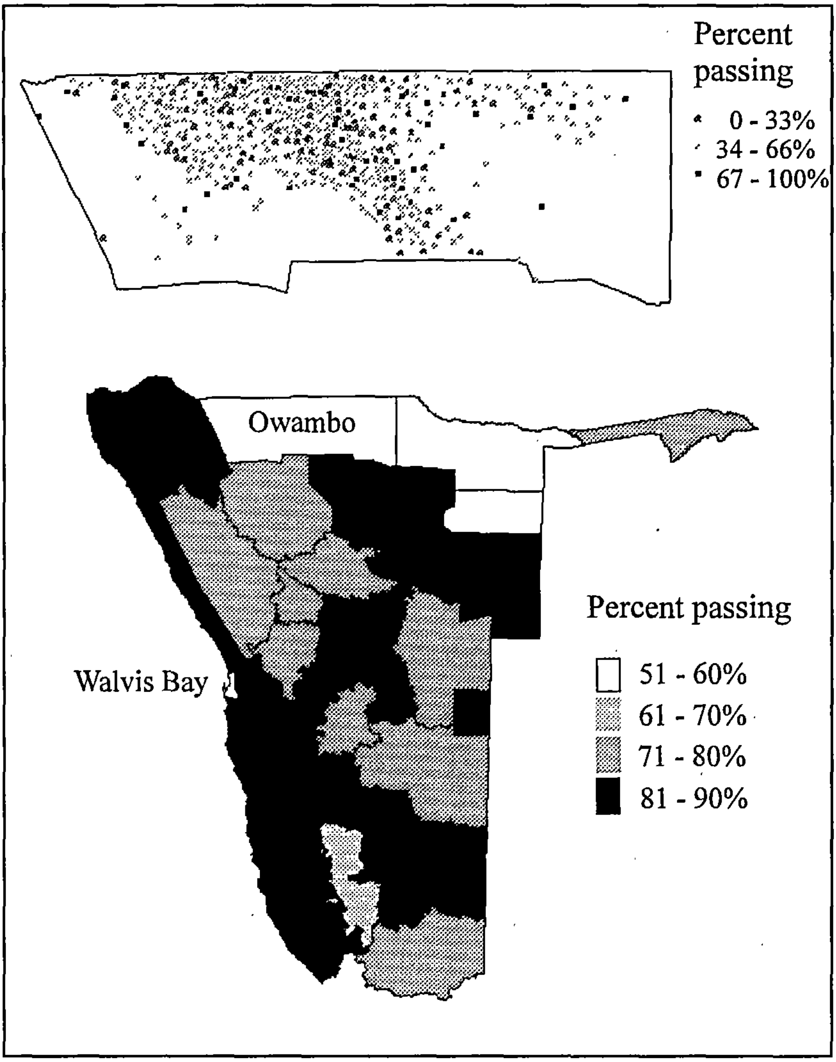
One example of this is shown in *Figure 14*, which shows two sets of information on the failure rates of pupils in Grade 1. The lower map shows substantial disparities between different administrative regions. The regions mapped here are those that divided Namibia along ethnic lines, and many of the disparities are commonly attributed to resources being unequally allocated to different ethnic groups and in different areas. Regions with high failure rates are those that were poorly supplied with resources before Independence in 1990. The map brings out these regional differences very clearly. Not much is new in the map, however. Everyone knows of these massive disparities, and all that is achieved is another fairly dramatic portrayal of the inequities.

Figure 13. Map of the Orange Free State province in South Africa using histograms to show numbers of children in each ethnic group not at school



Note: From the *Education Atlas of South Africa*, reproduced with the permission of the Education Foundation, South Africa.

Figure 14. Percentages of Grade 1s who passed at the end of 1992 at each school in the Owambo region (top) and in each magisterial and ethnic administration areas into which Namibia was divided before independence (bottom)



The upper map, however, reveals that even within the same region, conditions are by no means uniform, exploding the myth that failure rates are *only* dependant on the resources allocated to each region. This second map is more useful because it starts to suggest that other processes and dynamics may be at play, for example, that failure rates in individual schools may be far more dependant upon policies and decisions specific to each school. This is suggested by the fact that neighbouring schools, within just a few kilometres of each other, show quite different results.

These schools serve the same communities, with pupils coming from similar home backgrounds. From an educational and political point of view, it is clear that resources should be allocated equitably and, perhaps, even affirmatively. However, information revealed in the second map indicates that improved allocations, alone, are unlikely to make much of a difference to conditions in schools. The map suggests that procedures for passing and failing pupils need to be improved. Given this realisation and appropriate measures to ensure that similar norms are applied in all schools, greater equity can now be hoped for between regions and greater equity between pupils in nearby schools.

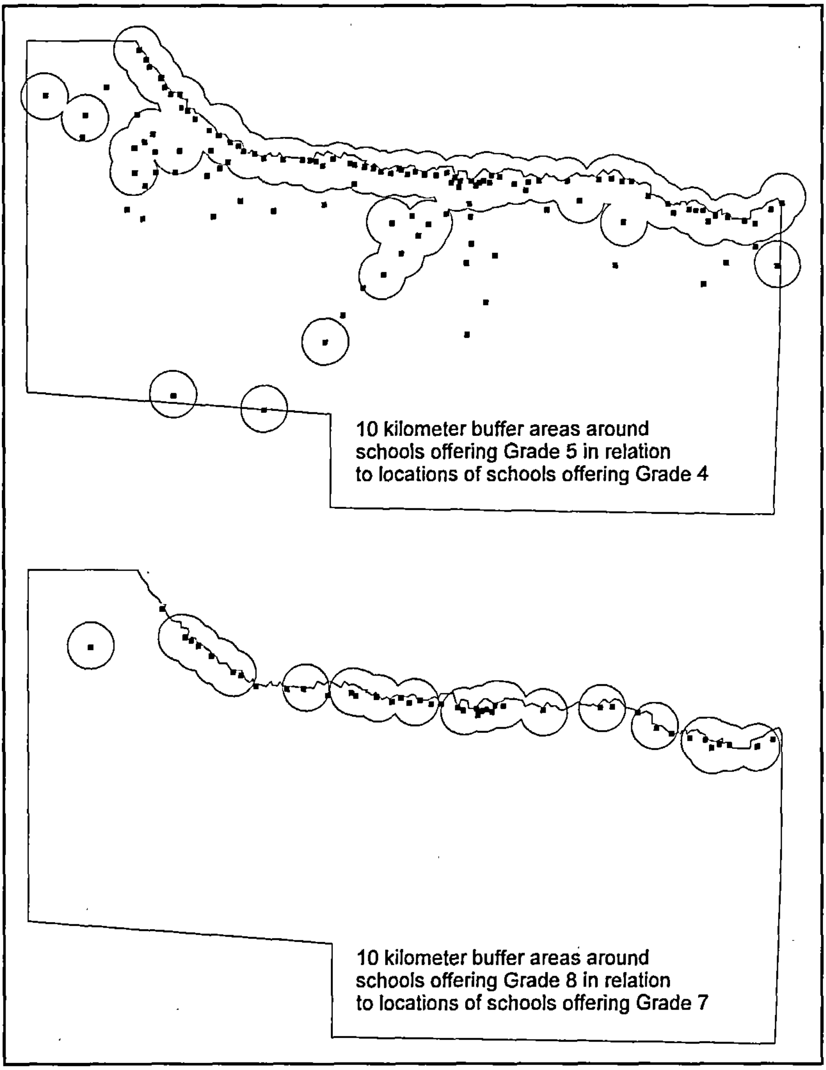
The quality of maps produced in an education environment will obviously depend on the quality of data available to the GIS user. If the positions of certain schools have yet to be determined accurately, it would be wise to 'flag' these as tentative positions. Likewise, attribute data that is not quite current but used in a map should be indicated as having been collected in an earlier year. Despite the tools available in a GIS, drawing a map takes time, and the more included in the map the longer it takes. Inevitably, you have to seek a compromise between drawing a quick map which might show all the desired features, and a much more detailed one which shows all that is necessary in addition to being a fine work of art. For example, a black and white map of schools could show the locations and names of schools, the positions of major roads, a scale, and nothing else. Another would be in colour, showing different classes of roads, contours, vegetation types, farm boundaries, electricity and telephone lines, all villages and other settlements, and railway lines. All of this would be accompanied by a detailed legend, a border showing principal lines of longitude and latitude, the sources of data, information on the projection and co-ordinate system, and other ancillary information about the map. The balance between these extremes will have to be sought for each map drawn.

Spatial analyses

The greatest power of a GIS lies in its use for the analysis of relationships between spatial objects. A good example of this is to consider how pupils have to move if they are to progress from one grade to the next. Many schools in Namibia offer a limited range of grades, so pupils have to move to another school if they are to continue their education. Two pairs of grades are shown in *Figure 15*: Grade 4 and 5 (two grades in the middle of the primary phase), and Grade 7 (the last year of primary school) and Grade 8 (the first secondary year). For each set, the question asked is how many schools and pupils are more than 10 kilometres away from the nearest school offering the next grade up, Grade 5 and Grade 8. The first step is to create a *buffer* zone around each school which offers Grade 5 and Grade 8. In this example, the buffer has a radius of 10 kilometres as the distance of interest to see how many pupils are unlikely, perhaps, to be able to walk 10 kilometres to a school once they finish Grade 4 and Grade 7, respectively. Buffer zones are created by the GIS as polygons around schools with Grade 5 and 8, those zones around nearby schools (closer than 10 kilometres, in this case) being merged into the same polygon. The next analytical step is to classify all schools in terms of whether they are inside or outside a buffer polygon, using a point-in-polygon intersection analysis (*Figure 8*). Finally, all that is needed is to count the number of schools outside the buffer zone and to sum the number of pupils that are unlikely to be able to walk to a school offering the next grade (*Figure 15*). This analysis is based on distances between schools and, of course, ignores the possibility that some pupils actually live nearer to a school with Grade 5 or Grade 8. A better analysis would be of spatial information on the locations of homes of pupils, but such information would be hard to assemble. The example in *Figure 15* provides a surprise. Far more of a gap had been expected between the last primary and first secondary grade than between two grades in the middle of the primary phase. Yet in this region only three schools and some 196 Grade 7 pupils are more than 10 kilometres away from schools offering Grade 8, compared with 21 schools and 398 Grade 4 pupils being more than 10 kilometres from schools with Grade 5 classes.

Another use of intersections is given in *Figure 16.* Namibia's population census in 1991 was conducted using a geographical framework of some 2200 enumeration areas or census blocks. Each enumerator was given one of these blocks to census. Paper maps showing the boundaries of all these enumeration areas were borrowed from the Central Statistics Office and then digitized. The resulting map of the country resembled a jigsaw puzzle, with a myriad of lines showing the boundaries of the census blocks. Aggregated data for each block could then be linked to its corresponding area to draw thematic maps showing, for example, the number of children not at school *(Figure 16)*. While these kinds of maps presented information from the Census accurately, there were two difficulties with the maps. First, the census block boundaries did not follow commonly recognized boundaries, so many people found it hard to interpret the maps. Second, the blocks varied greatly in size, being large in sparsely populated areas and small in areas where the population was dense. This meant that a tiny area having 100 children not at school would be hardly visible on a map of the whole country.

Figure 15. Buffer areas, 10 kilometres in radius, around schools providing Grade 5 and Grade 8



Note: The dots are schools providing Grade 4 and 7. Pupils in schools outside the buffer areas are likely to have difficulty in gaining access to the next grades.

By contrast, a large area having the same number of children not at school would be large and would therefore capture a disproportionate amount of attention on a map.

To solve these problems, all the Census information was dissected and then aggregated into a grid of some 1850 latitude and longitude squares – in sparsely populated areas, the squares comprised an area of 15 x 15 minutes, while in densely populated areas they were 7.5 x 7.5 minutes in size. The transformation involved overlaying and intersecting the grid of 1850 squares with the 2200 census polygons (*Figure 17*). Each enumeration area was dissected into pieces, the pieces being labeled or classified according to which latitude/longitude square they fell into while retaining the code of the enumeration area of which they were originally part. The area covered by each piece was calculated by the GIS and then expressed as a percentage of the area of the original enumeration area. It was then possible to divide up the data for the population in the enumeration area on a proportional basis into each grid square. For example, if a section of an enumeration area in a grid square made up 15 per cent of the surface area of the original enumeration area, 15 per cent of the population was allocated to that square. The proportions of the population allocated to each section in a square were finally summed to derive an estimated total population for each square.

The GIS also calculated the area of each latitude/longitude grid square, allowing the population data to be expressed in densities. For our purposes, the number of children not attending school could be shown as the number of children per 100 square kilometres *(Figure 16*). An obvious assumption in this analysis is that the population is evenly or homogeneously distributed in each enumeration area. While it is known that human populations are not evenly spread, the only way of avoiding this difficulty is to have spatial data with a very fine resolution. If an additional set of information on the locations of villages, farms and towns was available, it could have been used to adjust the proportions of the population to each grid block.

Examples given here have involved overlaying points (schools) with polygons (grid squares) and one set of polygons (enumeration areas) with another set (grid squares). Exactly the same procedures can be used to analyze relationships between lines and polygons. An overlay of roads with buffer polygons generated around towns where road maintenance crews are based would produce results showing what roads need to be maintained by crews in each town, what kinds of roads are involved, and the distances covered by each type of road. That kind of information would also be useful for budgetary purposes, budgets for materials being allocated on the basis of unit costs for maintaining different lengths of roads.

Figure 16. Different presentations of maps of areas where children aged 7 to 3 years are not at school

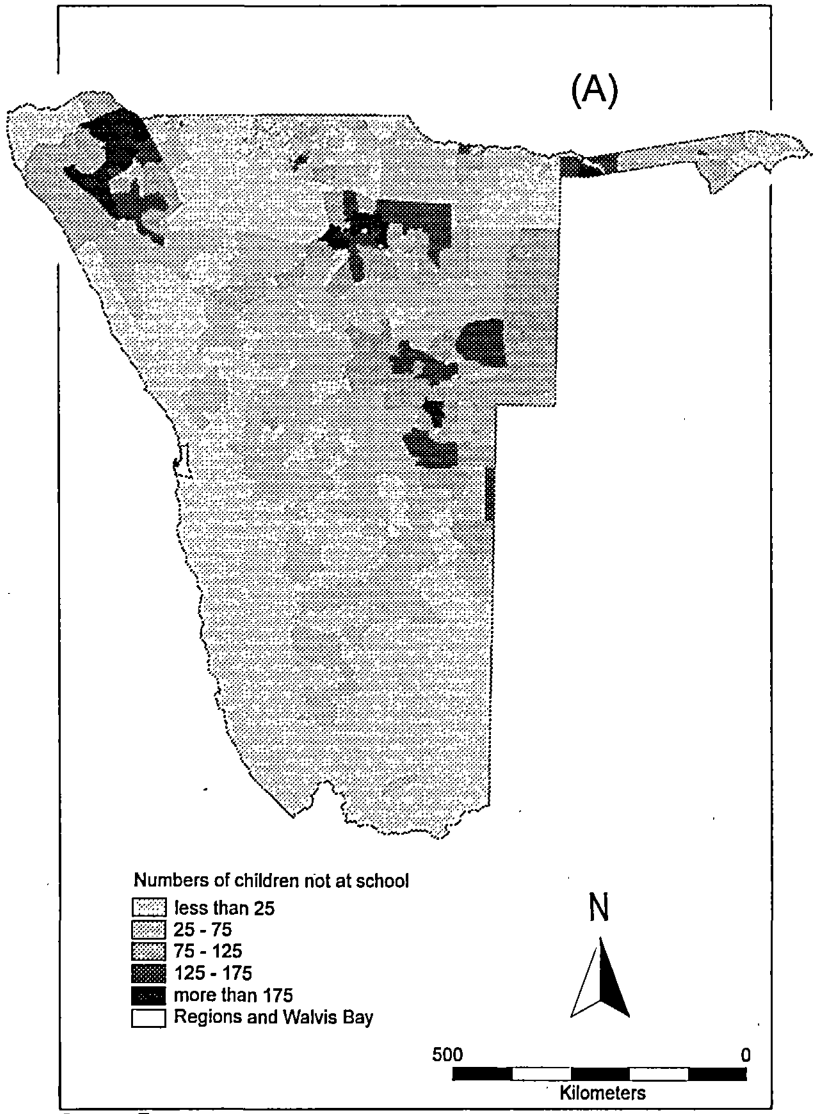
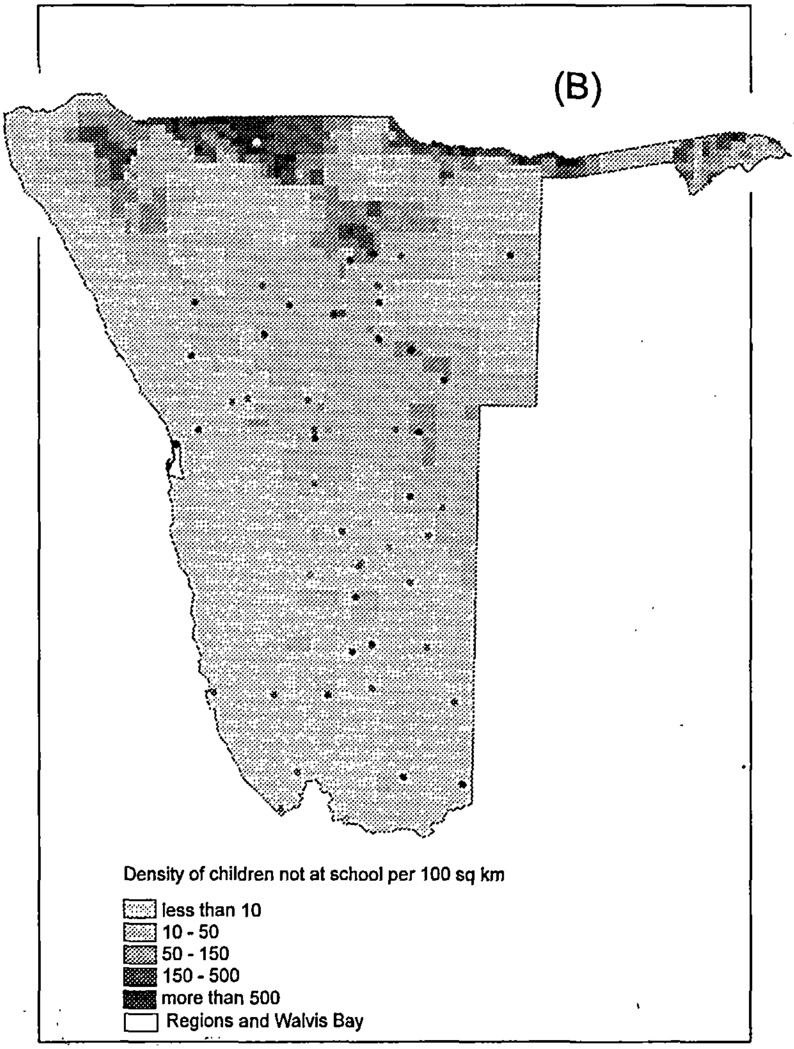


Figure 16 (cont'd)



Note: 'A' shows the number of children not at school in each of some 2200 census enumeration areas, while 'B' presents the same information in terms of the density of children in some 1860 latitude and longitude grid blocks. Census information was not available for the Walvis Bay enclave on the west coast.

Human settlements are often fairly evenly distributed in rural areas where villages are spaced at rather regular intervals across the countryside. Schools in these areas draw their pupils from surrounding settlements, the children walking each day to and from school. While the positions of schools may be known accurately, it is very difficult to obtain information on their catchment areas. One way of getting approximate catchment areas is to generate polygons around each school so that the boundaries of each polygon fall exactly half way between schools. This would be an easy task for two schools – simply plot the boundary at the half way mark between the two. Where many schools are involved, however, the calculations required to achieve this result would be very complex.

Thiessen polygons are created by GISs to solve this problem *(Figure 18)*. Maps showing the positions of schools and Thiessen polygons may be useful in showing areas where children have to walk unnecessarily large distances. Such maps would also be helpful in setting zoning criteria in urban areas, and in zoning primary schools as feeders to secondary schools.

These polygons provide only estimates of catchment areas, based on the assumption of equal distance to each school. Additional information and a better degree of resolution would be achieved if spatial data on the locations of settlements were overlaid with the polygons. That analysis would provide information on the comparative number of settlements within the catchment area of each school.

School-mapping programmes sometimes involve identifying homogeneous catchment areas around schools. These are zones sharing similar demographic features, where enrolments can be expected to change at similar rates, creating comparable demands at schools within those areas as well as similar educational characteristics. Part of the reason for identifying these homogeneous areas is to reduce the volume of information that has to be considered in drawing up a school map. Using a GIS, it may not be necessary to join similar areas into homogeneous zones since large amounts of data can be processed easily. However, even if there is a need to create homogenous areas, most GISs have tools to generate new polygons comprising of adjoining areas that share similar features (*Figure 19*). This is done by removing borders between two or more polygons classified as sharing the same features. Attributes of the original polygons can then be assigned to the new polygon in an aggregated or averaged form. For example, demographic indicators in census enumeration areas could be used to classify each area as belonging to any one group. The polygons could then be processed so that all those areas which belong to the same group and share borders are merged into a new set of regions. A common term for this procedure is *dissolving*.

Figure 17. Overlay of census enumeration areas (thick lines and large code labels) and a grid (thin lines and small code labels) of quarter degrees of latitude and longitude to show how census information can be transformed into a grid system

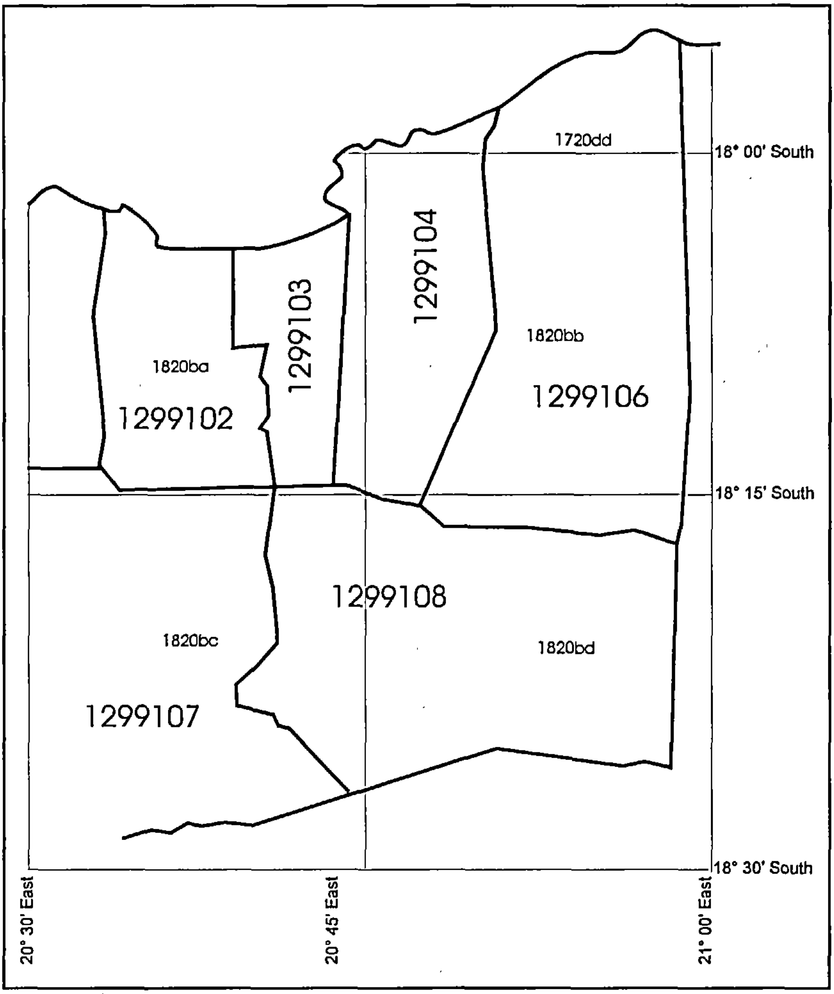
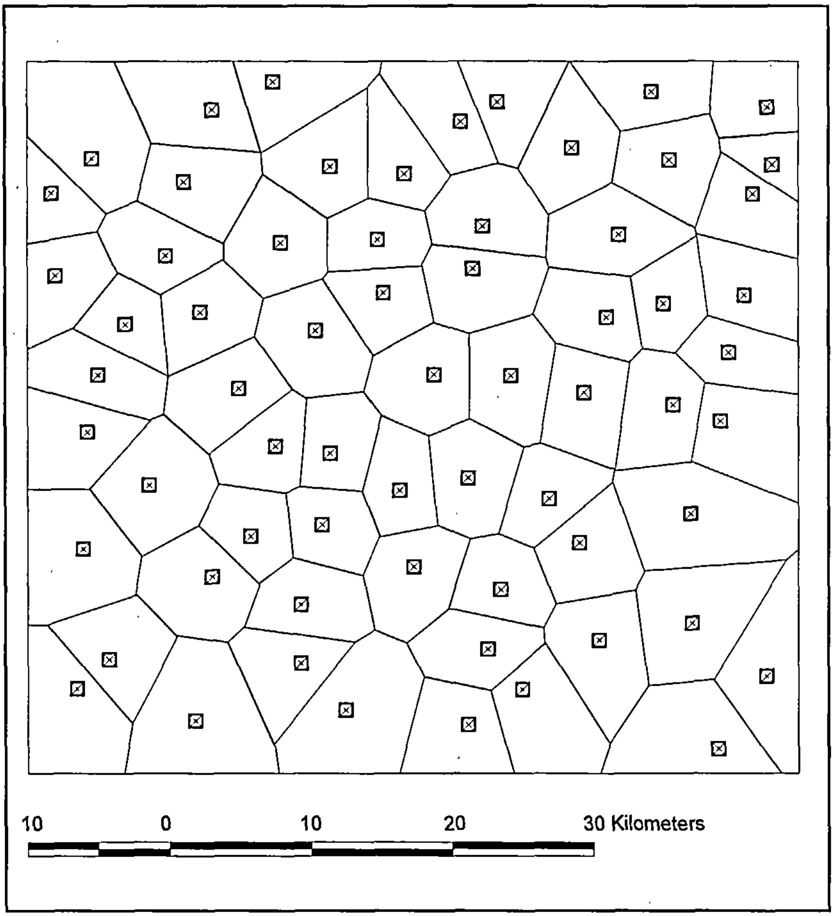


Figure 18. Thiessen polygons created around schools (small square blocks) offering Grade 1



Note: The boundaries around each school are an equal distance from all neighbouring schools.

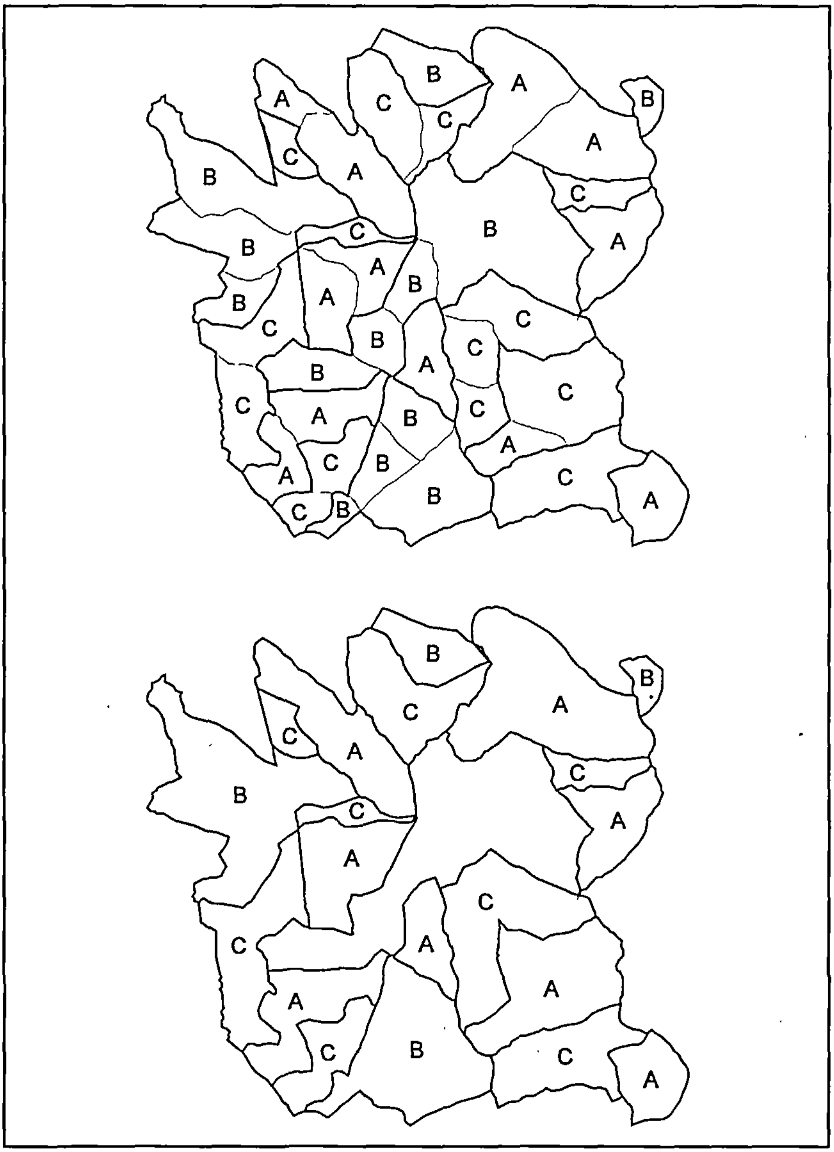
Properties along streets and roads in towns usually have numbers assigned to them, each having a distinct number that is consecutive with adjacent property numbers. In a process known as *address matching*, GISs can be used to plot the estimated positions of addresses along streets. Each street is assigned a range of values in a database to determine the lowest number at the beginning of the street and the highest property number at its end. Even-numbered properties can be assigned to one side of the street and odd numbers to the other, as is often done for real properties. Attribute data on those properties, for example, the names of people living in the houses, can then be linked by matching their addresses with the assigned range of street addresses. This is a much more economical method of linking addresses than using point objects for each property along a road or having the boundaries of each property stored as a polygon. One application for schools in towns is that maps can be created showing the homes of all the pupils and then related to maps showing actual or potential catchment areas of schools in the same town.

If spatial information on roads and railway lines in a town are available, maps showing the home addresses of pupils could be compared with the routes taken by buses and other transportation facilities to and from school. By determining the number of pupils living along each street, analyses can be done to determine the best routes to be taken en route to and from school. The analysis would take into account not only the number of pupils to be transported, but the distances and unit costs of travel involved. In a rural setting, routes could be analyzed to determine the best roads to be taken in delivering materials to schools and the best routes to be used by school inspectors during their routine visits to schools.

Information systems

The past few years have seen a substantial growth in the development of information systems, especially in the education sectors of developing countries. Much of this development has been stimulated by assumptions that education managers lack information to do their jobs, and that the provision of good information will stimulate good management. Education planners have, accordingly, set about collecting, processing and storing large volumes of information in elaborate databases. Rather less thought has been given to how that data can be retrieved by planners and managers. There are several difficulties in retrieving data: people need to be led into the information so they are encouraged to use the data, complicated menu structures or commands are needed to gain access to the information required, different people have different requirements at different times so each query differs from the next one, and a good understanding of how the data are stored and structured is often needed. GISs cannot overcome all of these obstacles, but they can help in several respects.

Figure 19. Polygon boundaries dissolved between polygons sharing the same attribute information, as indicated by the letters 'A', 'B' or 'C'



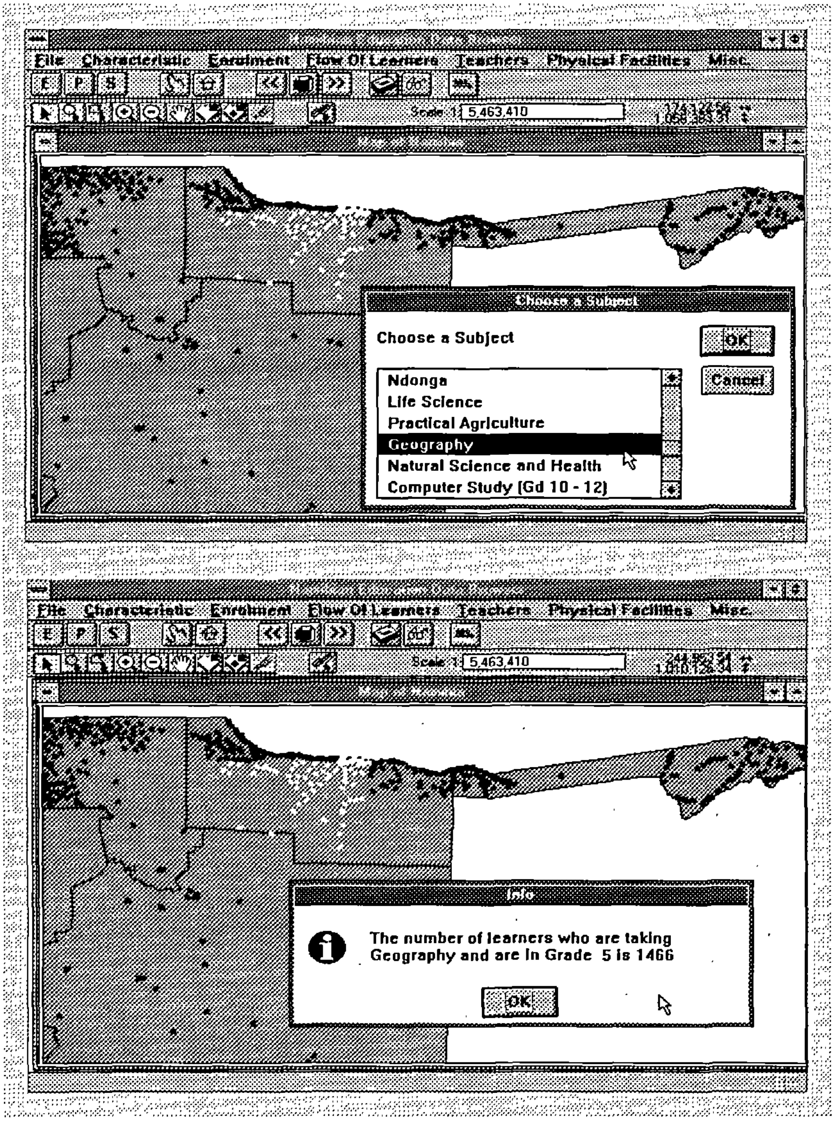
A map is a useful medium through which to gain access to information. Most people know their countries well enough to be able to select a recognized region, area or school that interests them. This means that a spatial object or set of objects can be selected by the user on a map on a computer screen, allowing attribute information linked to these features to be most easily interrogated or selected through a GIS. For example, you are asked to provide the names and addresses of all principals in a town. In the likely absence of a list of schools in that town, this would be a troublesome request, perhaps requiring you to scan through lists of schools in a telephone book to establish just which schools are located in the town. With a GIS, you would probably have current maps to show which schools are in the town. Better, still, you could select an area enclosing the town on your computer and have the GIS give you a list of the schools, the principals' names and addresses.

With spatial data being linked to a database of staff and addresses, the same selection procedure could immediately provide the desired list of principals' names and addresses. Information about selected schools can also be aggregated, as shown in the example in *Figure 20.* The user selects a group of schools and then follows a series of menu options to choose the grade and subject of interest to find out how many pupils take that subject in the grade. The information is provided in a matter of seconds, compared with the hours or days that it might take to obtain the same result using other methods.

One very neat feature common to all GISs is the ability to point to a spatial feature on the screen, click a button, and then have all attribute information linked to that object displayed on the screen. Such a feature could be region, town, road, or school – indeed any object stored in a spatial database. For a school, this might give you the name of the school, names of staff, enrolments in each grade, and, if you have graphics linked through the GIS to the school, an architectural plan of the school's buildings and a photograph of the principal. This feature is extremely useful for rapidly extracting information about single features.

The examples given in this illustrate some of the methods provided by GISs to display information about education on maps; to analyze spatial relationships between schools, their catchment areas and facilities that affect them; and to retrieve information in databases. The present treatment of these tools under three subheadings is simply for convenience. In reality, the GIS user will be mixing mapping tools, procedures to retrieve information and analyses of spatial information on a regular basis. For example, buffers might be created around schools, then intersected with census areas, and the final polygons then dissolved into larger and more homogeneous regions. Increasing levels of power and 'intelligence' are being added to GISs as they are developed and improved. So, in addition to the tools now available, more scope for solving problems concerned with the geography of education can be envisaged.

Figure 20. The use of a map interface to retrieve information



Note: Each dark dot in the map is a school, but those appearing white are those schools selected. The user then selects an aspect of information from the menu. In this example, the user selects enrolment information about Grade 5 and then chooses Geography (see top map). In the bottom map, a message appears to report that 1466 pupils take Geography in Grade 5 in those selected schools.

Chapter 6

Concluding comments

This book has attempted to introduce the basic components of geographical information systems. It is an introduction aimed at those who are thinking of developing or who have actually embarked on the development of a GIS for education planning and management. Of necessity, this is a short book and many other books available on the market provide more detail. While there is much to recommend in these books, they will not provide all the answers needed and they will not give you a working GIS. Put simply, there is no single recipe for developing a GIS. You need to pick up the basic principles and dive right in, learning as you go along. That approach will be the quickest way of getting a working system.

There are several aspects to developing a GIS which I believe to be critical ones. These are elements that you should remind yourself of continually, and they are offered here in the hope that they will make development of a GIS that much easier and more successful.

In starting off, spend a fair amount of time planning what you want from the system and, thus, what must go into the system. Do not be forced into hasty decisions by salesmen or consultants, especially if the latter come armed with funds from donor agencies. Consider the paradigm in which the GIS is intended to work and what kind of people will make best use of the products. Look for sources of spatial data from other organizations before starting to digitize or scan your own map information. Maintain close contact with other GIS users in your area, not only as sources of data but also for advice and support. Make sure that you clearly recognize what data will be held as spatial data and what will be added as attribute information. Make sure, also, that the coding system for spatial data will allow you to readily link them to attribute data. In building and acquiring different sets of spatial data, ensure that the data conform to the same map projections and co-ordinate systems. This will guarantee that all the data fit together accurately into the same spatial framework.

In selecting equipment and programmes (and staff), do not buy cheaply. Short-term savings are more likely to result in unnecessary additional costs in the long run. Ensure wherever possible that local support is available for the equipment purchased and that the system can be upgraded as hardware and software improve over the years to come.

In many developing countries, where digital spatial data may not be readily available, it would be wise to consider the use of satellite imagery as a major source of data. Uses for this kind of data are certain to grow in the years ahead, as the information becomes both cheaper and easier to process into forms used by GISs. The ability to map human settlements from satellite images, especially villages in rural areas, will add considerable power to the use of school mapping as a technique for planning the provision of education. Computers and software are often slightly intimidating, especially if the programmes are fairly complex and different from those in more common use. There is a tendency to be reluctant to try procedures without being quite sure of the commands and consequences. However, do not be scared of making mistakes. Exploring and experimenting with the programme is often a useful way of getting to know the power of your GIS. Different kinds of spatial analyses are worth trying, not only for what might be learned but also for the interesting results that might be obtained unexpectedly.

It is well-known that the number of GISs started is much greater than the number that produce results. A good deal of blame for this can be attributed to the systems being driven more by processes than products. The systems get hijacked by the processes of encouraging people to have a GIS because it is a 'good thing', by the processes of acquiring huge volumes of data, and by the processes of buying more and more equipment. Data, equipment and stimulation are certainly needed, but a clear balance must be maintained between what goes into the GIS and what can be produced at the end of the day. It is thus advisable that all stages of developing a GIS be driven by the need to produce tangible results.

One good way of ensuring this is to maintain a close link between the technical aspects of a GIS and the educational issues that it is intended to serve and support. Figuratively and conceptually, the system must be in a school administrator's office, not in a computer laboratory; the system must serve the needs of education planning and management, not the needs of computer technicians; and the system must be justified by its products.

Further reading

Antenucci, J. ; Brown, K. ; Croswell, P. ; Kevany, M.J. 1991. *Geographic information systems: a guide to the technology.* Routledge, Chapman and Hall.

Aronoff, S. 1989. *Geographic information systems: a management perspective.* WDL Publications.

Bernhardsen, T. 1992. *Geographic information systems.* Asplan Viak IT a.s./Norwegian Mapping Authority.

Carron, G. ; Ta Ngoc C. 1980. *Reduction of regional disparities: the role of educational planning.* Paris: UNESCO/International Institute for Educational Planning.

——. 1980. *Regional disparities in educational development: diagnosis and policies for reduction.* Paris: UNESCO/International Institute for Educational Planning.

——. 1980. *Regional disparities in educational development: a controversial issue.* Paris: UNESCO/International Institute for Educational Planning.

Curran, P.J. 1985. *Principles of remote sensing.* Longman Group Limited.

Eastman, J.R. ; Fuly, M. ; Toledano, J. 1993. *The GIS handbook.* USAID, Technical Paper No.11.

Fotheringham, S. ; Rogerson, P. (eds.). 1994. *Spatial analysis and geographic information systems.* Taylor and Francis.

France: Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche.1991; 1994; 1995; 1996. *Géographie de l'école.* Vanves (France): Direction de l'évaluation et de la prospective (DEP), Mined.

Hallak, J. 1977. *Planning the location of schools.* Paris: UNESCO/ International Institute for Educational Planning.

Huxhold, W. ; Levinsohn, A. 1995. *Managing geographic information systems projects.* Oxford University Press.

Johnson, A. ; Pettersson, C. ; Fulton, J. (eds.), 1992. *Geographic information systems (GIS) and mapping – practices and standards.* ASTM.

Krige, D. ; Cairns, S. ; Makalima, B. ; Scott, D. *The education atlas of South Africa.* The Education Foundation, 1994.

Langran, G. 1992. *Time in geographic information systems*. Taylor and Francis.

Laurini, R. ; Thompson, D. 1992. *Fundamentals of spatial information systems.* Academic Press.

Maguire, D. ; Goodchild, M. ; Rhind, D., (eds.). 1991. *Geographical information systems: principles and applications.* John Wiley and Sons.

Martin, D. 1991. *Geographic information systems and their socio-economic applications.* Routledge.

Medyckyj-Scott, D. ; Hearnshaw, H., (eds.). 1993. *Human factors in geographic information systems.* John Wiley and Sons.

Monmonier, M. 1991. *How to lie with maps.* University of Chicago Press.

Montgomery, G. ; Schuch, H. 1993. *GIS data conversion handbook.* GIS World.

Obermeyer, N. ; Pinto, J. 1994. *Managing geographic information systems.* Guilford Publications.

Star, J. ; Estes, J. 1990. *Geographic information systems: an introduction.* Prentice Hall.

Tomlin, D. 1990. *Geographic information systems and cartographic modelling.* Prentice Hall.

Tufte, E. 1991. *Envisioning information.* Graphics Press.

Wood, D. 1992. *The power of maps.* The Guildford Press.

Worrall, L. (ed.). 1990. *Geographic information systems: developments and applications.* John Wiley and Sons.

——. (ed.). 1991, *Spatial analysis and spatial policy using geographic information systems.* John Wiley and Sons.

IIEP publications and documents

More than 750 titles on all aspects of educational planning have been, published by the International Institute for Educational Planning. A comprehensive catalogue, giving details of their availability, includes research reports, case studies, seminar documents, training materials, occasional papers and reference books in the following subject categories:

Economics of education, costs and financing.

Manpower and employment.

Demographic studies.

The location of schools (school map) and sub-national planning.

Administration and management.

Curriculum development and evaluation.

Educational technology.

Primary, secondary and higher education.

Vocational and technical education.

Non-formal, out-of-school, adult and rural education.

Copies of the catalogue may be obtained from the IIEP Publications Unit on request.

The International Institute for Educational Planning

The International Institute for Educational Planning (IIEP) is an international centre for advanced training and research in the field of educational planning. It was established by UNESCO in 1963 and is financed by UNESCO and by voluntary contributions from Member States. In recent years the following Member States have provided voluntary contributions to the Institute: Denmark, Germany, Iceland, India, Ireland, Norway, Sweden, Switzerland and Venezuela.

The Institute's aim is to contribute to the development of education throughout the world, by expanding both knowledge and the supply of competent professionals in the field of educational planning. In this endeavour the Institute co-operates with interested training and research organizations in Member States. The Governing Board of the IIEP, which approves the Institute's programme and budget, consists of a maximum of eight elected members and four members designated by the United Nations Organization and certain of its specialized agencies and institutes.

Inquiries about the Institute should be addressed to:
The Office of the Director, International Institute for Educational Planning,
7-9 rue Eugène-Delacroix, 75116 Paris, France.

The book

Educational managers and planners are confronted almost daily with such questions as: Where are new schools needed? Where can pupils move after completing their primary schooling? Where are school facilities duplicated? Where is there a need for new teachers, or where is in-service training most urgently required? Where should inspection circuits be created? etc.

A Geographical Information System (GIS) comprises a set of tools useful for processing spatial information, and therefore can help in answering the questions above. Educational planners and managers – with a strong interest in the details of education, in local conditions and local solutions, and in having their efforts translated into tangible products – will make good use of a GIS. Those with broader interests in policies, processes and central control will also use GIS to highlight, among other things, regional disparities, problem areas, and schools in need of immediate action. The book has been designed to find something of a middle road between those with a strong interest in education and those with a greater interest in the technology of GISs.

The author

John Mendelsohn, a Namibian citizen, born in Kenya and trained in South Africa, was until recently Senior Education Officer in charge of the Management Information System in the Namibian Ministry of Education and Culture. With a PhD in Zoology, he has worked as a research biologist and educational planner. His current interests focus on research and information systems in the fields of education and environmental biology.

UNESCO International Institute for Educational Planning IIEP
ISBN 92-803-1158-1