NEW AGE

GIS BASICS

SHAHAB FAZAL

NEW AGE INTERNATIONAL PUBLISHERS

GIS BASICS

This page intentionally left blank

GIS BASICS

SHAHAB FAZAL

Department of Geography
Aligarh Muslim University
Aligarh

PUBLISHING FOR ONE WORLD

NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS

New Delhi • Bangalore • Chennai • Cochin • Guwahati • Hyderabad
Jalandhar • Kolkata • Lucknow • Mumbai • Ranchi
Visit us at **www.newagepublishers.com**

Published by New Age International (P) Ltd., Publishers

ISBN (13) : 978-81-224-2639-7

PUBLISHING FOR ONE WORLD

NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS
4835/24, Ansari Road, Daryaganj, New Delhi - 110002
Visit us at **www.newagepublishers.com**

Preface

In today's world – the flow of information especially digital information has become the critical ingredient for success in any activity. That is why, the period we live in is often referred to as an **information age**.

It is a simple fact that everything human beings do, do takes place at a certain location on the earth – it has a geographic component, although we tend not to think about it much. The digital information revolution of the late twentieth century has allowed this geographic information to be more easily accessed, analyzed and used than ever before. This led to the development of GIS as a discipline and emergence of GIS as a core of digital technology.

The technology of GIS is spread over the domain of several disciplines such as Mathematics, Statistics, Computer Sciences, Remote Sensing, Environmental Sciences and of course Geography. Similarly, diverse is the list of its applications – Commerce, Governance, Planning and Academic Research. These application areas are also growing and expanding every day due to its power and vast possibilities.

Traditionally, the discipline of Geography dealt with spatial description and analysis. Now in the era of multidisciplinary approach, students, researchers, professionals from different disciplines find their way into the emerging discipline of GIS making it popular.

The rapid expansion and popularization of GIS means that now GIS is not just for the specialists, but for everyone, but these GIS users have different requirements. There are numerous amounts of GIS learning material available in the form of textbooks as well as posted on various websites. These literatures in general tend to be rather advanced and designed for specialists while requirements of GIS beginners are some what ignored.

The present book is an attempt to provide basic fundamentals of GIS for beginners. The book is evolved following the basic education approach, spreading onto three stages of learning. The first stage is about basic fundamentals, here development in technology instigating the learning processes are discussed. This is spread over first three chapters, which introduces the beginners to the GIS as a discipline, its history, development and evolvement process. The second stage is about the scope of the field, here the emphasis is on issues of technological advancements and revolution in spatial learning and their basic concepts. In this section, four chapters (fourth to eighth) cover the breadth and also depth of GIS, here geographic data, nature, structure, source and real world models are elaborated. Lastly, the third stage of learning, where the approach works towards the development of critical thinking, using the knowledge base acquired from the earlier chapters. The last four chapters discuss geographic query, analysis, selection and future of GIS, project design and management.

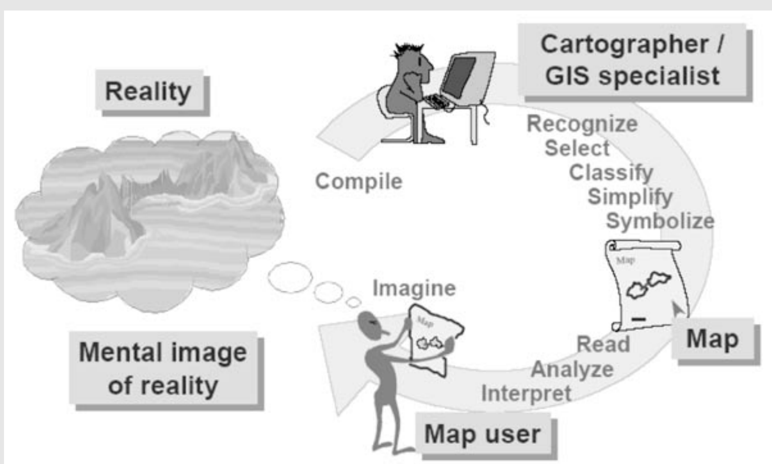
Shahab Fazal

CONTENTS

This page intentionally left blank

CHAPTER 1

GEOGRAPHICAL INFORMATION SYSTEMS — REPRESENTING GEOGRAPHY



Society is now so dependent on computers and computerized information that we scarcely notice when an action or activity makes use of them. Over the past few decades we have developed extremely complex systems for handling and processing data represented in the only form acceptable to computers: strings of zeros and ones, or bits (binary digits). Yet is has proved possible to represent not only numbers and letters, but sound, images, and even the contents of maps in this simple, universal form. Indeed, it might be impossible to tell whether the bits passing at high speed down a phone line, or stored in minute detail on a CD-ROM (compact disk-read-only memory) represent a concerto by Mozart or the latest share prices. Unlike most of its predecessors, computer technology for processing information succeeds in part because of its ability to store, transmit, and process an extremely wide range of information types in a generalized way.

The utility of computer has become so important nowadays, that almost all our activities have some bearing on computers. Its ability to quick and efficient processing of the given task has revolutionized our life. Spatial Information Technology is the outcome of developments in computer technology. Geography, as with for other subjects, stipulates the use of information technology to gain access to additional information sources and to assist in handling, presenting and analyzing spatial informations. Internet and computerization has opened a vast new potential in the way we perceive, communicate and analyze our surrounding spatial phenomena. Data representing the real world can be stored, processed and presented in relatively simplified forms to suit specific needs. This provides base for geographical information system.

Computerization has opened a vast new potential in the way we communicate, analyze our surroundings, and make decisions. Data representing the real world can be stored and processed so that they can be presented later in simplified forms to suit specific needs. Many of our decisions depend on the details of our immediate surroundings, and require information about specific places on the Earth's surface. Such information is called geographical because it helps us to distinguish one place from another and to make decisions for one place that are appropriate for that location. Geographical information allows us to apply general principles to the specific conditions of each location, allows us to track what is happening at any place, and helps us to understand how one place differs from another (Figure 1.1). Geographical information, then, is essential for effective planning and decision making.

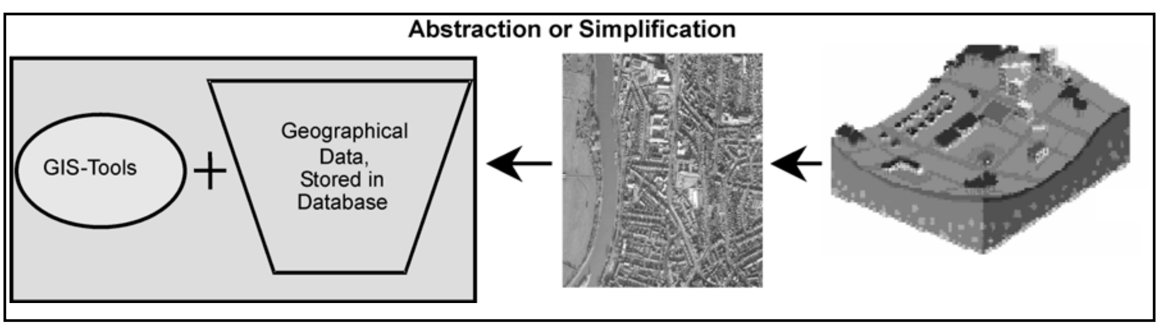


Figure 1.1: GIS builds database those results from data processing of real world informations.

Box 1: *General questions with geographic importance*

Every day people pose questions

- Where is GURGAON ?
- What are the soil characteristics there ?
- What is the land use pattern in Gurgaon District ?
- Which is the main economic activity in Gurgaon District ?
- What are the trends in rural and urban employment pattern in Gurgaon District ?
- Where would be a better location for opening a restaurant in Gurgaon District ?
- Which is the shortest route to reach Gurgaon from New Delhi railway station?

Almost everything that happens or exists occurs 'somewhere'. Knowing 'where' it happened or existed is critically important.

All human activities require knowledge about the Earth, thus geographic location is very important.

We are used to thinking about geographical information in the form of maps, photos taken from aircraft, and images collected from satellites, so it may be difficult at first to understand how such information can be represented in digital form as strings of zeros and ones. If we can express the contents of a map or image in digital form, the power of the computer opens an enormous range of possibilities for communication, analysis, modelling, and accurate decision making (Figure 1.2). At the same time, we must constantly be aware of the fact that the digital representation of geography is not equal to the geography itself-any digital representation involves some degree of approximation.

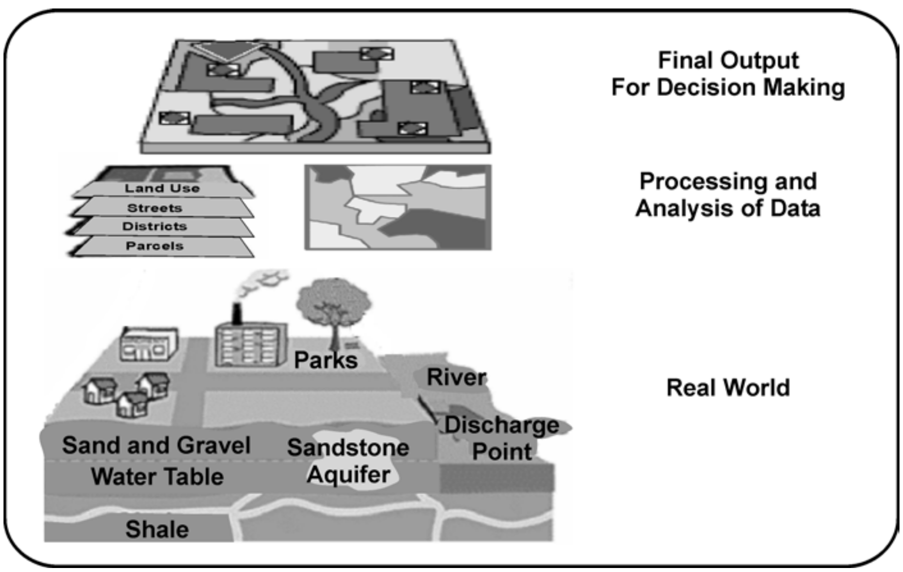


Figure 1.2: GIS simplifies the real world informations to bring it into computer. Different techniques are used to analyze data for decision making.

Information Technologies in Geography

GIS is one of many information technologies that have transformed the ways geographers conduct research and contribute to society. In the past two decades, these information technologies have had tremendous effects on research techniques specific to geography, as well as on the general ways in which scientists and scholars communicate and collaborate.

Discipline-Specific Tools

1. *Cartography and Computer:* Assisted Drafting: Computers offer the same advantages to cartographers that word-processing software offers writers. Automated techniques are now the rule rather than the exception in cartographic production.
2. *Photogrammetry and Remote Sensing:* Aerial photogrammetry, a well – established technique for cartographic production and geographic analysis, is now complemented by the use of 'remotely sensed' information gathered by satellites in outer space. Information technologies have made both sorts of information far more readily available and far easier to use.
3. *Spatial Statistics:* Statistical analysis and modelling of spatial patterns and processes have long relied on computer technology. Advances in information technology have made these techniques more widely accessible and have allowed models to expand in complexity and scale to provide more accurate depictions of real-world processes.
4. *Geographic Information Systems (GIS):* These systems allow geographers to collate and analyze information far more readily than is possible with traditional research techniques. As will be noted below, GIS can be viewed as an integrating technology insofar as it draws upon and extends techniques that geographers have long used to analyze natural and social systems.

General Communication, Research, and Publication Technologies

1. *Communication and Collaboration:* Electronic mail, discussion lists, and computer bulletin boards make it far easier for colleagues to communicate ideas and share ideas, locally, nationally, and internationally. Distance – learning techniques make it possible to hold interactive classes and workshops simultaneously at distant locations.
2. *Access to Library and Research Materials and Sources:* Network access to both primary and secondary research resources is expanding rapidly. From their offices, scholars can now get information held by libraries, government agencies, and research institutions all over the world.
3. *Publication and Dissemination:* Information technologies are reducing substantially the cost of publishing and distributing information as well as reducing the time required to circulate the latest news and research results.

The Course of Technological Innovation

These advances in the application of information technologies in geography began several decades ago and will continue to expand their effects into the foreseeable future. Scholars who have studied the spread of technological innovations in society sometimes divide the process into four phases:

1. *Initiation:* An innovation first becomes available.
2. *Contagion:* Far-ranging experimentation follows to see how the innovation can be adapted to meet a wide variety of research and commercial needs. Some, but not necessarily all of these experiments will work.
3. *Coordination:* The most promising applications of the innovation gradually gain acceptance and are developed collaboratively. The coordination of experimentation helps to distribute the potentially high costs of further development and implementation.
4. *Integration:* An innovation is accepted and integrated into routine research tasks.

In geography, many innovations in the application of information technologies began in the late 1950s, 1960s and early 1970s. Methods of sophisticated mathematical and statistical modelling were developed and the first remote sensing data became available. Researchers began also to envision the development of geographic information systems. The mid-1970s to early 1990s was a period of contagion. The first commercially available software for GIS became available in the late 1970s and spurred many experiments, as did the development of the first microcomputers in the early 1980s. This was an exciting time in which the development of powerful software coupled with the availability of inexpensive computers permitted many researchers to test new ideas and applications for the first time. In the early 1990s, or perhaps just a bit earlier, many innovations entered the coordination phase even as other experimentation continued at a fast pace. The strengths and weaknesses of many information technologies were by then apparent, and researchers began to work together to cultivate the most promising applications on a large scale. Arguably, the complete integration of information technologies in geography has yet to be achieved except perhaps in a few relatively specialized research areas. Complete integration across the discipline may, in fact, be many years away.

GIS as an Integrating Technology

In the context of these innovations, geographic information systems have served an important role as an integrating technology. Rather than being completely new, GIS have evolved by linking a number of discrete technologies into a whole that is greater than the sum of its parts. GIS have emerged as very powerful technologies because they allow geographers to integrate their data and methods in ways that support traditional forms of geographical analysis, such as map overlay analysis as well as new types of analysis and modelling that are beyond the capability of manual methods. With GIS it is possible to map, model, query, and analyze large quantities of data all held together within a single database.

The importance of GIS as an integrating technology is also evident in its pedigree. The development of GIS has relied on innovations made in many different disciplines: Geography, Cartography, Photogrammetry, Remote Sensing, Surveying, Geodesy, Civil Engineering, Statistics, Computer Science, Operations Research, Artificial Intelligence, Demography, and many other branches of the social sciences, natural sciences, and engineering have all contributed. Indeed, some of the most interesting applications of GIS technology discussed below draw upon this interdisciplinary character and heritage.

Geographic Information Systems: A Generic Definition

GIS is a special-purpose digital database in which a common spatial coordinate system is the primary means of reference. Comprehensive GIS require a means of:

1. Data input, from maps, aerial photos, satellites, surveys, and other sources.
2. Data storage, retrieval, and query.
3. Data transformation, analysis, and modelling, including spatial statistics.
4. Data reporting, such as maps, reports, and plans.

Three observations should be made about this definition

First, GIS are related to other database applications, but with an important difference. All information in a GIS is linked to a spatial reference. Other databases may contain locational information (such as street addresses, or zip codes), but a GIS database uses geo-references as the primary means of storing and accessing information.

Second, GIS integrates technology. Whereas other technologies might be used only to analyze aerial photographs and satellite images, to create statistical models, or to draft maps, these capabilities are all offered together within a comprehensive GIS.

Third, GIS, with its array of functions, should be viewed as a process rather than as merely software or hardware. GIS are for making decisions. The way in which data is entered, stored, and analyzed within a GIS must mirror the way information will be used for a specific research or decision – making task. To see GIS as merely a software or hardware system is to miss the crucial role it can play in a comprehensive decision-making process.

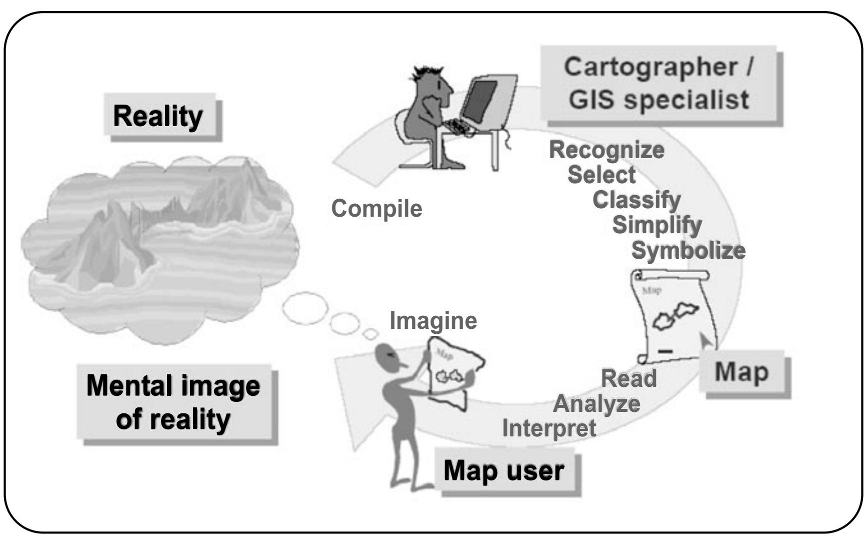


Figure 1.3: Different stages of information transfer in GIS.

What Actually GIS is?

GIS is expressed in individual letters G – I – S and not at pronunciation GIS. It stands for geographic or geographical information systems. Geographic Information Science is a new interdisciplinary field. It is built upon knowledge from geography, cartography, computer science, mathematics etc.

GIS can be defined as ***'A system for Capturing, storing, checking, integrating, manipulating, analysing and displaying data which are spatially referenced to the Earth. This is normally considered to involve a spatially referenced computer database and appropriate applications software'.***

GIS needs spatial data, this makes it unique. Here spatial means – related to the space – the real world location. That is why GIS is based on basic geographic concepts.

A Geographic Information System is an integration of computer hardware and software which can create manipulate, and analyze a geographically referenced data base to produce new maps and tabular data GIS includes the capabilities of Computer Aided Design (CAD) and Data Base Management Systems (DBMS), but is more than just a combination of those systems. In a GIS, a relationship between the graphic map data and the tabular data base is maintained so that changes to the map are reflected in the data base GIS allows automatic determination of the relationships between maps, and can create new maps of those relationships.

Geographic Information System (GIS) can also be defined as:

The organized activity by which people

- Measure aspects of geographic phenomena and processes;
- Represent these measurements, usually in the form of a computer database, to emphasize spatial themes, entities, and relationships;
- Operate upon these representations to produce more measurements and to discover new relationships by integrating disparate sources; and
- Transform these representations to conform to other frameworks of entities and relationships.

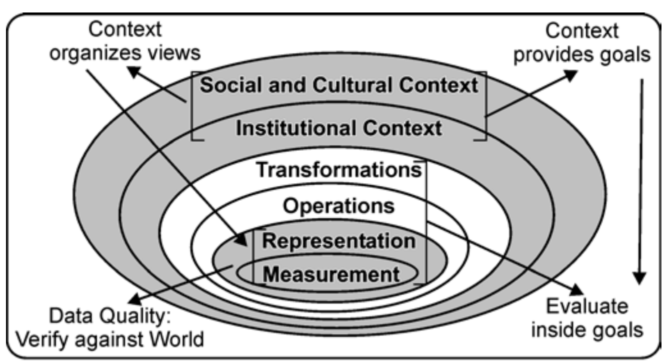


Figure 1.4: GIS framework.

These activities reflect the larger context (institutions and cultures) in which these people carry out their work. In turn, the GIS may influence these structures.

Other definitions

Many people offer definitions of GIS. In the range of definitions presented below, different emphases are placed on various aspects of GIS. Some miss the true power of GIS, its ability to integrate information and to help in making decisions, but all include the essential features of spatial references and data analysis.

A definition quoted in William Huxhold's Introduction to Urban Geographic Information Systems:

> *'. . . The purpose of a traditional GIS is first and foremost spatial analysis. Therefore, capabilities may have limited data capture and cartographic output. Capabilities of analyses typically support decision making for specific projects and / or limited geographic areas. The map data-base characteristics (accuracy, continuity, completeness, etc.) are typically appropriate for small-scale map output. Vector and raster data interfaces may be available. However, topology is usually the sole underlying data structure for spatial analyses.'*

C. Dana Tomlin's definition, from Geographic Information Systems and Cartographic Modelling:

> *'A geographic information system is a facility for preparing, presenting, and interpreting facts that pertain to the surface of the earth. This is a broad definition . . . a considerably narrower definition, however, is more often employed. In common parlance, a geographic information system or GIS is a configuration of computer hardware and software specifically designed for the acquisition, maintenance, and use of cartographic data.'*

From Jeffrey Star and John Estes, in Geographic Information Systems: An Introduction:

> *'A geographic information system (GIS) is an information system that is designed to work with data referenced by spatial or geographic coordinates. In other words, a GIS is both a database system with specific capabilities for spatially-reference data, as well [as] a set of operations for working with data . . . In a sense, a GIS may be thought of as a higher-order map.'*

The GIS View of the World

GIS provide powerful tools for addressing geographical and environmental issues. Consider the schematic diagram below. Imagine that the GIS allows us to arrange information about a given region or city as a set of maps with each map displaying information about one characteristic of the region. In the case below, a set of maps that will be helpful for urban transportation planning have been gathered. Each of these separate thematic maps is referred to as a **layer, coverage, or level.** And each layer has been carefully overlaid on the others so that every location is precisely matched to its corresponding locations on all the other maps. The bottom layer of this diagram is the most important, for it represents the grid of a locational reference system (such as latitude and longitude) to which all the maps have been precisely registered.

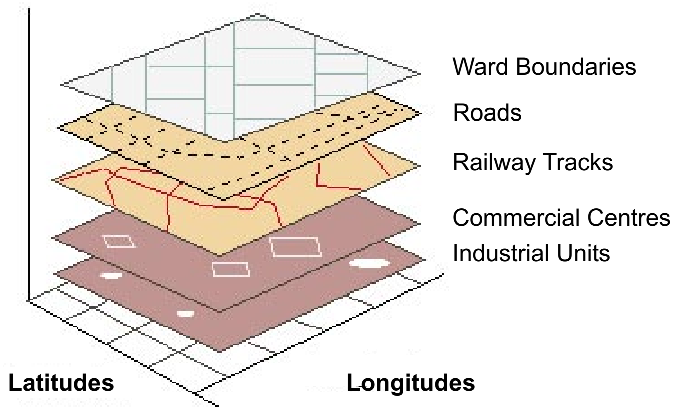


Figure 1.5: GIS: an integrating technology.

WHY IS GIS IMPORTANT?

- 'GIS technology is to geographical analysis what the microscope, the telescope, and computers have been to other sciences.... (It) could therefore be the catalyst needed to dissolve the regional-systematic and human-physical dichotomies that have long plagued geography' and other disciplines which use spatial information.
- GIS integrates spatial and other kinds of information within a single system – it offers a consistent framework for analyzing geographical data.
- By putting maps and other kinds of spatial information into digital form, GIS allows us to manipulate and display geographical knowledge in new and exciting ways.
- GIS makes connections between activities based on geographic proximity
  - looking at data geographically can often suggest new insights, explanations.
  - these connections are often unrecognized without GIS, but can be vital to understanding and managing activities and resources.
  - *e.g.* we can link toxic waste records with school locations through geographic proximity.

Box 2: Definitions of GIS and the groups who find them useful.

A container of maps in digital form	the general public
A computerized tool for solving geographic problems	decision makers, planners
A spatial decision support system	managers, operations researchers
A mechanized inventory of geographically distributed features	utility managers, resource managers
A tool for revealing what is otherwise invisible in geographic information	scientists, investigators
A tool for performing operations on geographic data that are too tedious if performed by manual methods	resource managers, planners, GIS experts

- GIS allows access to administrative records – property ownership, tax files, utility cables and pipes – via their geographical positions.
- Maps are fascinating and so are maps in computers and there is increasing interest in geography and geographic education in recent times. GIS gives a 'high tech' feel to geographic information.

Contributing Disciplines

GIS is a convergence of technological fields and traditional disciplines. GIS has been called an 'enabling technology' because of the potential it offers for the wide variety of disciplines which must deal with spatial data. Each related field provides some of the techniques which make up GIS. Many of these related fields emphasize data collection – GIS brings them together by emphasizing integration, modelling and analysis, as the integrating field, GIS often claims to be the science of spatial information.

GEOGRAPHY: Geography is broadly concerned with understanding the world and man's place in it. Geography has long tradition in spatial analysis. The discipline of geography provides techniques for conducting spatial analysis and a spatial perspective on research.

CARTOGRAPHY: Cartography is concerned with the display of spatial information. Currently it is the main source of input data for GIS is maps. Cartography provides long tradition in the design of maps which is an important form of output from GIS. Computer cartography (also called 'digital cartography', 'automated cartography') provides methods for digital representation and manipulation of cartographic features and methods of visualization.

REMOTE SENSING: This emerging technique which records images from space and the air are major source of geographical data. Remote sensing includes techniques for data acquisition and processing anywhere on the globe at low cost, consistent update potential. The main advantage of it is that interpreted data from a remote sensing system can be merged with other data layers in a GIS.

PHOTOGRAMMETRY: Using aerial photographs and techniques for making accurate measurements from them, photogrammetry is the source of most data on topography (ground surface elevations) used for input to GIS.

SURVEYING: Surveying is concerned with the measurement of locations of objects on the Earth's surface, particularly property boundaries. Surveying provides high quality data on positions of land boundaries, buildings, etc.

STATISTICS: Many models built using GIS are statistical in nature, many statistical techniques used for analysis in GIS. Statistics is important in understanding issues of error and uncertainty in GIS data.

COMPUTER SCIENCE: Computer science is one of the main engines for GIS development. Artificial intelligence (AI) uses the computer to make choices based on available data in a way that is seen to emulate human intelligence and decision-making – computer can act as an 'expert' in such functions as designing maps, generalizing map features. Computer-aided design (CAD) provides software, techniques for data input, display and visualization, representation, particularly in 3 dimensions. Advances in computer graphics provide hardware, software for handling and displaying graphic objects, techniques of visualization. Similarly, database management systems (DBMS) contribute methods for representing data in digital form, procedures for system design and handling large volumes of data, particularly access and update.

MATHEMATICS: Several branches of mathematics, especially geometry and graph theory, are used in GIS system design and analysis of spatial data.

Major Areas of Application

GIS technology, data structures and analytical techniques are gradually being incorporated into a wide range of management and decision-making operations. Numerous examples of applications of GIS are available in many different journals and are frequent topics of presentations at conferences in the natural and social sciences.

In order to understand the range of applicability of GIS it is necessary to characterize the multitude of applications in some logical way so that similarities and differences between approaches and needs can be examined. An understanding of this range of needs is critical for those who will be dealing with the procurement and management of a GIS.

FUNCTIONAL CLASSIFICATION: One way to classify GIS applications is by functional characteristics of the systems; this would include a consideration of characteristics of the data such as themes, precision required and data model. Secondly, GIS a function as which of the range of possible GIS functions does the application rely on? *e.g.* address matching, overlay? Thirdly, a product *e.g.,* does the application support queries, one-time video maps and/or hardcopy maps? A classification based on these characteristics quickly becomes fuzzy since GIS is a flexible tool whose great strength is the ability to integrate data themes, functionality and output.

GIS AS A DECISION SUPPORT TOOL: Another way to classify GIS is by the kinds of decisions that are supported by the GIS. Decision support is an excellent goal for GIS, however: decisions range from major (which areas in India are best suited for establishing SEZ with foreign aids?) to minor (which way to turn at next intersection?). Decision support is a good basis for definition of GIS, but not for differentiating between applications since individual GIS systems are generally used to make several different kinds of decisions.

GIS USERS: GIS field is a loose coalescence of groups of users, managers, academics and professionals all working with spatial information. Each group has a distinct educational and 'cultural' background with varied interests and priorities. As a result; each identifies itself with particular ways of approaching particular sets of problems. The core groups of GIS activity can be seen to be comprised of:

a. mature technologies which interact with GIS, sharing its technology and creating data for it such as surveyors and engineers, cartographers, scientists using remote sensing techniques.
b. management and decision-making groups such as resource inventors, and resource managers, urban planners, municipal officials managing land records for taxation and ownership control, facilities managers, managers involved in marketing and retail planning or vehicle routing and scheduling.
c. science and research activities at universities and government labs – these groups of GIS activity seeking to find distinctions and similarities between them.

SOME IMPORTANT AREAS WHERE GIS IS BEING USED ARE:

- *Different Streams of Planning:* Urban planning, housing, transportation planning architectural conservation, urban design, landscape planning etc.
- *Street Network Based Application:* It is an addressed matched application, vehicle routing and scheduling: location, development and site selection and disaster planning.
- *Natural Resource Based Application:* Management and environmental impact analysis of wild and scenic recreational resources, flood plain, wetlands, acquifers, forests, and wildlife.
- *View Shed Analysis:* Hazardous or toxic factories siting and ground water modelling. Wildlife habitat study and migrational route planning.
- *Land Parcel Based:* Zoning, sub-division plans review, land acquisition, environment impact analysis, nature quality management and maintenance etc.
- *Facilities Management:* Can locate underground pipes and cables for maintenance, planning, tracking energy use.

THE APPEAL AND POTENTIAL OF GIS

The great appeal of GIS stems from their ability to integrate great quantities of information about the environment and to provide a powerful repertoire of analytical tools to explore this data. Imagine the potential of a system in which dozens or hundreds of maps layers are arrayed to display information about transportation networks, hydrography, population characteristics, economic activity, political jurisdictions, and other characteristics of the natural and social environment. Such a system would be valuable in a wide range of situations – for urban planning, environmental resource management, hazards management, emergency planning, or transportation forecation, and so on. The ability to separate information in layers, and then combine it with other layers of information is the reason why GIS hold such great potential as research and decision-making tools.

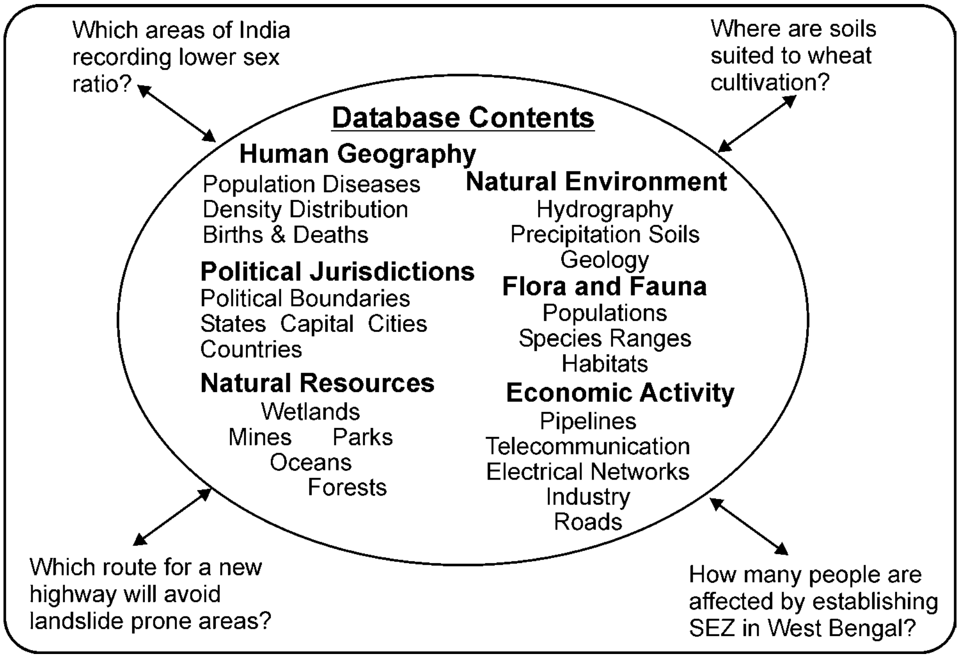


Figure 1.6: Application potential of GIS for geographical studies.

Development of GIS

Since the mid-1970s, specialized computer systems have been developed to process geographical information in various ways. These include:

- Techniques to input geographical information, converting the information to digital form.
- Techniques for storing such information in compact format on computer disks, compact disks (CDs), and other digital storage media.
- Methods for automated analysis of geographical data, to search for patterns, combine different kinds of data, make measurements, find optimum sites or routes, and a host of other tasks.
- Methods to predict the outcome of various scenarios, such as the effects of climate change on vegetation.
- Techniques for display of data in the form of maps, images, and other kinds of displays.
- Capabilities for output of results in the form of numbers and tables.

Components of GIS

HARDWARE: It consists of the computer system on which the GIS software will run. The choice of hardware system ranges from Personal Computers to multi user Super Computers. These a computers should have essentially an efficient processor to run the software and sufficient memory to store enough information (data).

SOFTWARE: GIS software provides the functions and tools needed to store, analyze, and display geographic information. The software available can be said to be application specific. All GIS software generally fit all these requirements, but their on screen appearance (user interface) may be different.

DATA: Geographic data and related tabular data are the backbone of GIS. It can be collected in-house or purchased from a commercial data provider. The digital map forms the basic data input for GIS. Tabular data related to the map objects can also be attached to the digital data. A GIS will integrate spatial data with other data resources and can even use a DBMS.

METHOD: A successful GIS operates according to a well-designed plan, which are the models and operating practices unique to each task. There are various techniques used for map creation and further usage for any project. The map creation can either be automated raster to vector creator or it can be manually vectorized using the scanned images. The source of these digital maps can be either map prepared by any survey agency or satellite imagery.

PEOPLE: GIS users range from technical specialists who design and maintain the system to those who use it to help them perform their everyday work. GIS operators solve real time spatial problems. They plan, implement and operate to draw conclusions for decision making.

NETWORK: With rapid development of IT, today the most fundamental of these is probably the network, without which no rapid communication or sharing of digital information could occur. GIS today relies heavily on the Internet, acquiring and sharing large geographic data sets.

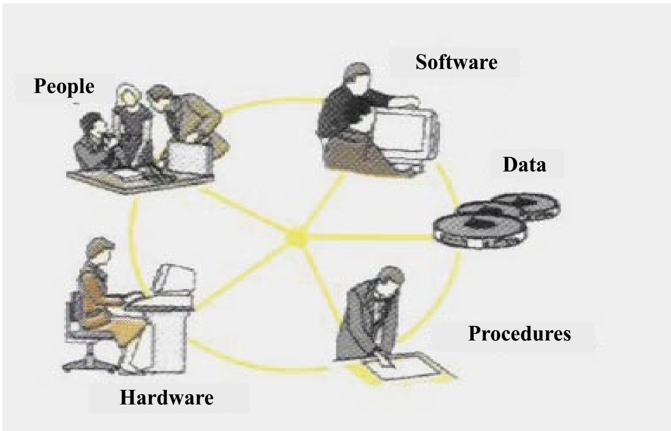


Figure 1.7: Six basic components of GIS.

Although it is very easy to purchase the constituent parts of a GIS (the computer hardware and basic software), the system functions only when the requisite expertise is available, the data are compiled, the necessary routines are organized, and the programs are modified to suit the application. A computer system can function at what may appear to be lightning speed, yet the entire time span of a GIS project can stretch to months and even years. These facets of an overall GIS are interlinked. In general, procurement of the computer hardware and software is vital but straightforward. The expertise required is often underestimated, the compilation of data is expensive and time consuming, and the organizational problems can be most vexing. These facets of an overall GIS are discussed in detail later.

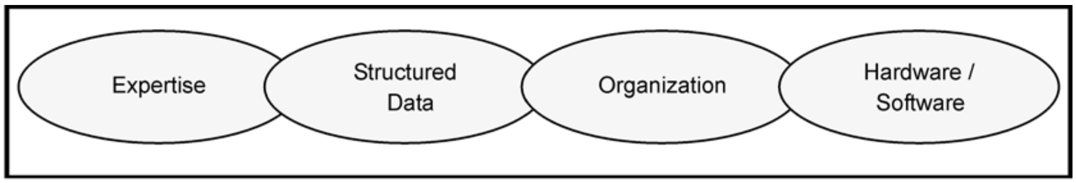


Figure 1.8: A GIS chain – Equal role of the above links in GIS organization.

Traditionally, geographical data are presented on maps using symbols, lines, and colours. Most maps have a legend in which these elements are listed and explained – a thick black line for main roads, a thin black line for other roads, and so on. Dissimilar data can be superimposed on a common coordinate system. Consequently, a map is both an effective medium for presentation and a bank for storing geographical data. But herein lies a limitation. The stored information is processed and presented in a particular way, usually for a particular purpose. Altering the presentation is seldom easy. A map provides a static picture of geography that is almost always a compromise between many differing user needs. Nevertheless, maps are a substantial public asset. Surveys conducted in Norway indicate that the benefit accrued from the use of maps is three times the total cost of their production.

Compared to maps, GIS has the inherent advantage that data storage and data presentations are separate. As a result, data may be presented and viewed in various ways. Once they are stored in a computer, we can zoom into or out of a map, display selected areas, make calculations of the distance between places, present tables showing details of features shown on the map, superimpose the map on other information, and even search for the best locations for retail stores. In effect, we can produce many useful products from a single data source.

The term geographical information system (GIS) is now used generically for any computer-based capability for the manipulation of geographical data. GIS is computer-based capability for the manipulation of geographical data. A GIS includes not only hardware and software, but also the special devices used to input maps and to create map products, together with the communication systems needed to link various elements. The hardware and software functions of a GIS include:

- Compilation
- Storage
- Updating and changing
- Management and exchange
- Manipulation
- Retrieval and presentation

- Acquisition and verification
- Analysis and combination

All of these actions and operations are applied by a GIS to the geographical data that form its database. All of the data in a GIS are georeferenced, that is, linked to a specific location on the surface of the Earth through a system of coordinates. One of the commonest coordinate systems is that of latitude and longitude; in this system location is specified relative to the equator and the line of zero longitude through Greenwich, England. But many other systems exist, and any GIS must be capable of transforming its georeferences from one system to another.

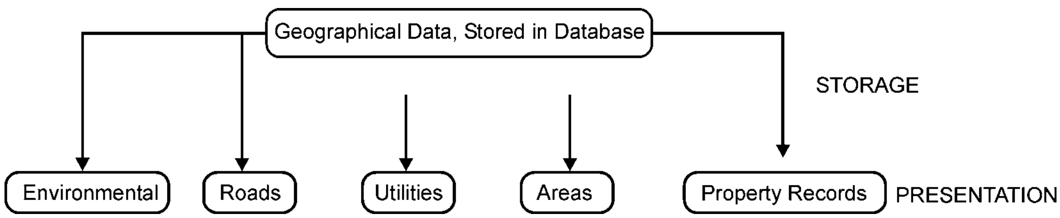


Figure 1.9: A map can be a presentation medium and a storage medium. GIS manipulates data to produce results.

Geographical information attaches a variety of qualities and characteristics to geographical locations (Figure 1.10). These qualities may be physical parameters such as ground elevation, soil moisture level, or classifications according to the type of vegetation, ownership of land, zoning, and so on. Such occurrences as accidents, floods, or landslides may also be included. We use the general term attributes to refer to the qualities or characteristics of places, and think of them as one of the two basic elements of geographical information, along with locations.

In some cases, qualities are attached to points, but in other cases they refer to more complex features, either lines or areas, located on the Earth's surface; in such cases the GIS must store the entire mapped shape of the feature rather than a simple coordinate location. Examples of commonly mapped features are lakes, cities, counties, rivers, and streets, each with its set of useful attributes. When a feature is used as a reporting zone for statistical purposes, a vast amount of information may be available to be used as attributes for the zone in GIS. In market research, for example, it is common for postal codes to be used as the basis for reports on demographics, purchasing habits, and housing markets.

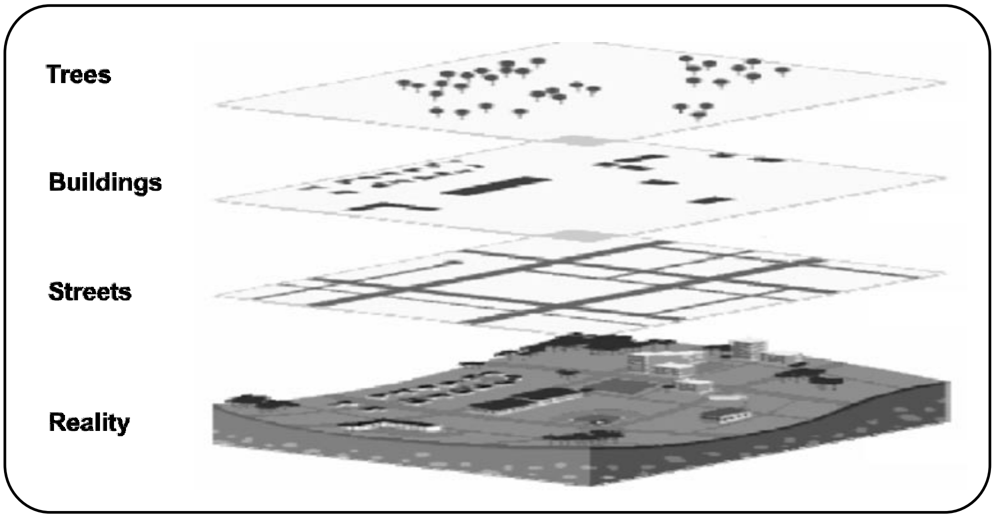


Figure 1.10: GIS stores data in different theme layers in the computer, each layer is linked to a common referencing system.

The relationships between geographical features often provide vital information. For example, the connections of a water supply pipe network may be critical for technicians, who need to know which valves to close in order to increase water pressure in certain sectors. The details of properties bordering a road are necessary if all property owners affected by roadwork are to be properly notified. Connections between streets are important in using a GIS to assist drivers in navigating around an unfamiliar city. The ability of a GIS to store relationships between features in addition to feature locations and attributes is one of the most important sources of the power and flexibility of this technology. Some GISs can even store flows and other measures of interaction between features, to support applications in transportation, demography, communication, and hydrology, among other areas.

Stored data may be processed in a GIS for presentation in the form of maps, tables, or special formats. One major GIS strength is that geographical location can be used to link information from widely scattered sources. Because the geographical location of every item of information in a GIS database is known, GIS technology makes it possible to relate the quality of groundwater at a site with the health of its inhabitants, to predict how the vegetation in an area will change as the irrigation facilities increases, or to compare development proposals with restrictions on land use. This ability to overlay gives GIS unique power in helping us to make decisions about places and to predict the outcomes of those decisions. The only requirement is that the geographical information from each source be expressed in compatible georeferencing systems.

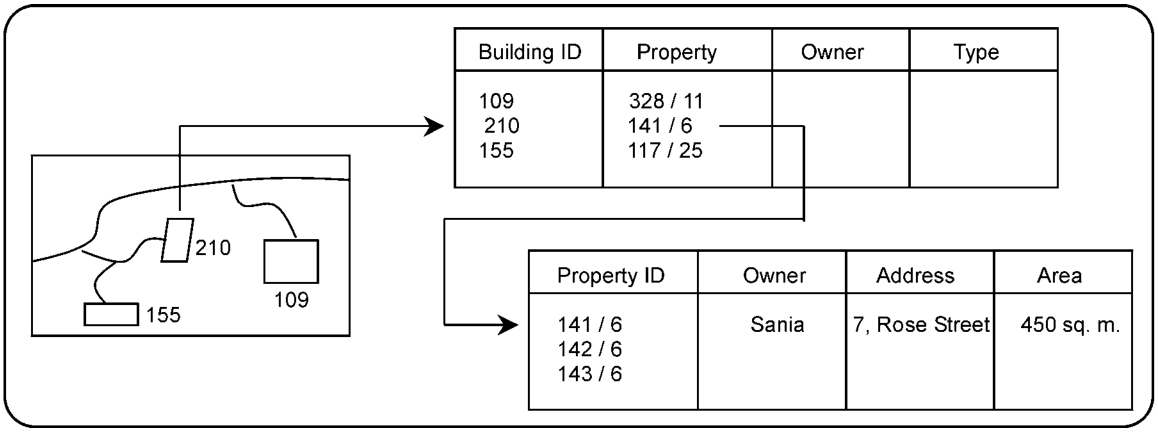


Figure 1.11: GIS functions on the interaction between digital map data and its attribute informations.

A GIS can process georeferenced data and provide answers to questions involving, *e.g.*, the particulars of a given location, the distribution of selected phenomena, the changes that have occurred since a previous analysis, the impact of a specific event, or the relationships and systematic patterns of a region. It can perform analyses of georeferenced data to determine the quickest driving route between two points and help resolve conflicts in planning by calculating the suitability of land for particular uses.

A GIS can process georeferenced data and provide answers to questions involving, *e.g.*, the particulars of a given location, the distribution of selected phenomena, the changes that have occurred since a previous analysis, the impact of a specific event, or the relationships and systematic patterns of a region. It can perform analyses of georeferenced data to determine the quickest driving route between two points and help resolve conflicts in planning by calculating the suitability of land for particular uses.

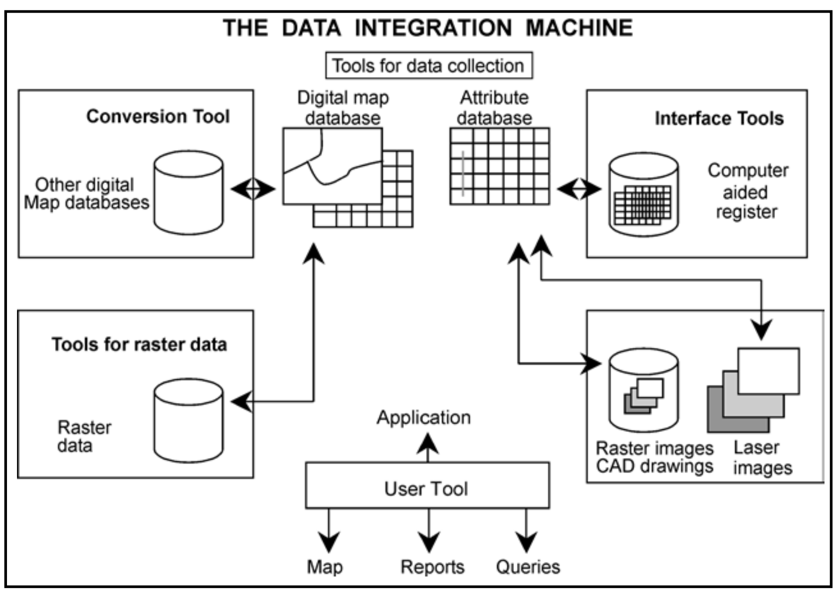


Figure 1.12: GIS is a typical data integration machine. It receives, process and transmits data.

GIS can process data from a wide range of sources, including data obtained from maps, images of the Earth obtained from space satellites, video film of the Earth taken from low-flying aircraft, statistical data from published tables, photographs, data from computer-assisted design (CAD) systems, and data obtained from archives by electronic transmission over the Internet and other networks. Data integration is one of the most valuable functions of a GIS, and the data that are integrated are more and more likely to be obtained from several distinct media-multimedia is an active area for research and development in GIS (Figure 1.12).

Technically, a GIS organizes and exploits digital geographical data stored in databases. The data include information on locations, attributes, and relationships between features. But a database can only approximate the real world, since the storage capacity of a database is minuscule in comparison with the complexity of the real world, and the cost of building a database is directly related to its complexity. The contents of a book of 100,000 words can be stored in digital form in roughly 1 million bytes (the common unit of computer storage is a byte, defined as 8 bits; 1 megabyte is slightly more than 1 million bytes). The information on a topographic map is comparatively dense, and it commonly takes 100 megabytes to capture it in digital form. A single scene from an Earth observing satellite might contain 300 megabytes, the information content of 300 books. Thus even crude approximations to the complexity of real-world geography can rapidly overtake the capacity of our digital storage devices.

Although the contents of a GIS database are equivalent to a map, there are important differences. On a map, a geographical feature such as a road or a power line is shown as a

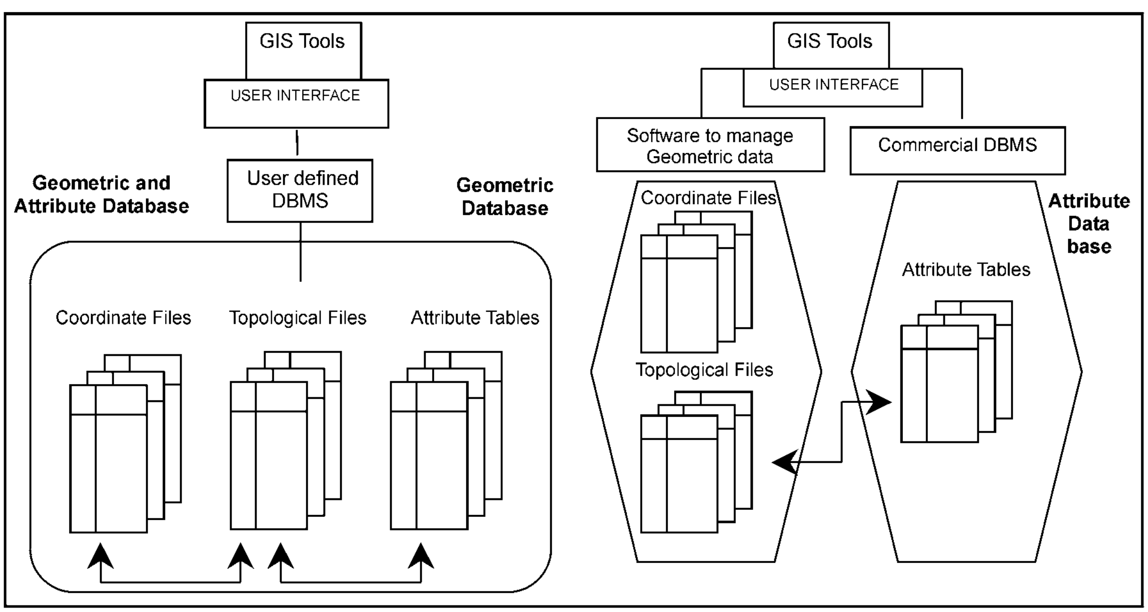


Figure 1.13: DBMS solution for GIS.

symbol using a graphic that will readily be understood by the map reader. In a geographical database a road or power line will be represented by a single sequence of points connected by straight lines, and its symbolization will be reattached when it is displayed. A tube well will be represented by a single point, with the attribute 'tube well', and will be replaced by a symbol when displayed. This approach is economical since the geometric form of the tube well symbol will be stored only once rather than repeated at each tube well location, and it also allows analysis to be more effective.

Databases are vital in all geographical information systems, since they allow us to store geographical data in a structured manner that can serve many purposes. Many GISs impose further structure by using a database management system (DBMS) to store and manage part or all of the data in a largely independent subsystem under the GIS itself. A DBMS is a general-purpose software product, and GISs that use this approach are often able to function in conjunction with a wide range of DBMS products. The database underlying a GIS achieves many objectives. It ensures that data are:

- Stored and maintained in one place
- Stored in a uniform, structured, and controlled manner than can be documented
- Accessible to many users at once, each of whom has the same understanding of the database's contents
- Easily updated with new data

This contrasts with the traditional way of organizing and storing data on paper in filing cabinets, in which data are often:

- Stored in ways that are understandable to one person only
- Easily corrupted by use, or edited in ways that are meaningful only to the editor
- Inaccessible to anyone other than the creator of the system
- Stored in formats and at scales that are so diverse that they cannot be compared or collated
- Difficult to update

GIS Diversity

Although the general definition of GIS given here is quite valid, in practice the diversity of GIS has spawned various definitions. First, users have contrived working definitions suited to their own specific uses. Thus they may vary according to whether operators are planners, water-supply and sewage engineers, support service personnel, or perhaps professional and public administrators or Earth scientists. Second, those with a more theoretical approach, such as research workers, software developers or sales and training staff may use definitions that are different from those used in practical applications. Systems can be tailor made by assembling them from available software tool kits of semi-independent modules, assorted computer hardware components, and other interoperable devices. Many applications can be addressed by acquiring a single, generic GIS product and a standard configuration of hardware. There are many views of GISs, including:

- A data processing system designed for map production or visualization
- A data analysis system for examining conflicts over plans or optimizing the design of transport systems
- An information system for responding to queries about land ownership or soil type
- A management system to support the operations of a utility company, helping it to maintain its distribution network of pipes or cables
- A planning system to aid the design of road systems, excavations, or forest harvest operations
- An electronic navigation system for use in land or sea transport.

GISs are often designated according to application. When used to manage land records they are often called land information systems (LISs); in municipal and natural resource applications they are important components of urban information systems (UISs) and natural resource information systems (NRISs) respectively. The terms spatial and geospatial are often used almost interchangeably with geographical, although spatial is also used to refer more generally to any two-or three-dimensional data whether or not it relates directly to the surface of the Earth. The term automatic mapping/facility management (AM/FM) is frequently used by utility companies, transportation agencies, and local governments for systems dedicated to the operation and maintenance of networks. Nonetheless, GIS is now accepted internationally as an umbrella term for all digital systems designed to process geographical data.

The software capabilities required for a GIS often overlap those needed by other computer applications, particularly image processing and computer-assisted design (CAD). Image processing systems are designed to perform a wide range of operations on the images captured by video cameras, still cameras, and remote-sensing satellites. Today, the distinction between image processing and GIS is becoming increasingly blurred as images become more and more important sources of GIS data. Broadly, though, it is convenient to think about image processing systems as concerned primarily with the extraction of information from images, and GIS as concerned with the analysis of that information.

CAD systems have been developed to support design applications in engineering, architecture, and related fields. Broadly, CAD systems emphasize design over analysis and often lack the capabilities needed to process the complex attributes and information of georeferenced data or to integrate georeferenced data from many sources. Nevertheless, the distinction between CAD and GIS has become increasingly blurred in recent years; by adding appropriate features, many former vendors of CAD systems are now able to compete effectively in the GIS market.

The major challenges to system developers and users alike are now very different, and related to the comparative ease of use of the technology, the problems of finding and accessing suitable data, and the lack of trained personnel able to exploit the technology's potential to its full.

Our complex society Modern societies are now so complex, and their activities so interwoven, that no problem can be considered in isolation or without regard for the full range of its interconnections. For example, a new housing development will affect the local school system. Altered age distribution in a village will affect health and social expenditure. The volume of city traffic will put constraints on the maintenance of buried pipe networks, affecting health. Street excavations may drastically reduce the turnover of local retail shops. Traffic noise from a new road or motorway may well drive people from their homes. The actions needed to solve such problems are best taken on the basis of standardized information that can be combined in many ways to serve many users. GISs have this capability.

Populations are now extremely mobile; changing jobs and moving to another location have become commonplace. When key personnel leave a company, they take their expertise with them; if that expertise involves specific knowledge of, say, the water supply and sewage network of a community, the loss can be serious if the information is otherwise inadequately documented. Here, too, GIS has an advantage in that it can act as an effective filing system for dissimilar sectors of a complex society.

Benefits of Computerizing Information

Almost all aspects of modem society use digital information, and the total amount that flows through our communication networks daily is truly staggering. GIS offers its users the ability to process quantities of data far beyond the capacities of manual systems. Data in GIS are stored in a uniform, structured manner, as opposed to manual systems in which data are stored in archives and files, in agencies, on file cards, on maps, or in long reports. Data may be retrieved from GIS databases and manipulated far more rapidly and reliably than data in manual systems. In addition, data are quickly compiled into documents using techniques that include automatic mapmaking and direct report printouts. The potential gains from switching from manually prepared maps and ordinary files to computerized GIS are considerable, in both the public and private sectors. Various studies showed that considerable benefits may be achieved, provided that the strategy used to implement GIS is suitably chosen. The study also showed that benefits are often related to objectives and that the following benefit/cost ratios may be attained by introducing GIS (Figure 1.14):

1. If computerized GIS is used for automated production and maintenance of maps, the benefit/cost ratio is 1:1.
2. If the system is also used for other internal tasks such as work manipulation and planning, the benefit/cost ratio may be 2:1.
3. The full benefit of the system is first realized when information is shared among various users. The benefit/cost ratio may then be 4:1.

Nonetheless, it is obvious that investment in GIS is at least as productive as investment in other sectors. These benefits are not automatic. They depend largely on proper choice of an acquisition and implementation strategy, following careful study of the objectives and requirements of GIS investment, and careful selection of the appropriate system.

Objective	GIS Operation	Production of Data	Use of Data
Task	• Storage • Update • Manipulation • Maintenance • Retrieval	• Analysis of Data • Map Production • Planning • Project Management	• Map Production • Coordination of Tasks • Information Updating • Information Sharing • Management & Planning • Execution of Task
Benefit / Cost Ratio	**1:1**	**2:1**	**4:1**

Figure 1.14: The benefit/cost ratio of GIS data is significantly high.

Without these safeguards, many GIS projects eventually fail to deliver the promised benefits and may eventually fail entirely, at considerable cost to the institution. Even with a carefully selected strategy it is difficult to estimate benefits precisely. The ratios discussed above are average over many projects varying widely in scale and scope. Some figures, however, are impressive, with benefit/cost ratios of up to 8 to 10:1 or more.

But benefits are a function of many factors, including the goals and objectives of the project, the strategy adopted in its implementation, and the structure of the system built to serve the objectives.

Systematic planning and implementation often set profitable GIS projects apart from those that are unprofitable. Projects based on carefully estimated cost and benefit calculations are often more profitable than projects driven by pure technology. Profitable projects are user oriented rather than production oriented. Profitable projects start by being defined so clearly and convincingly that they are funded outside the ordinary operating budget. The measurable benefits of GIS are usually expressed as gains in efficiency in terms of time saved, but there are also many cases of direct increases in income and reductions in costs. Measurable benefits may include:

- Improved efficiency due to more work being performed by the same staff, or the same work performed by a smaller staff
- Reduction in direct operating costs through better bases for financial management, less costly maintenance of facilities, and joint uses of available data
- Increases in income due to increased sales, or sales of new products and services

Experience indicates that when GIS makes some traditional jobs superfluous, staff are not made redundant but instead put to tasks in the GIS environment that create more value. Intangible benefits may also accrue. They cannot be expressed directly in monetary terms, but attempts should always be made to include them when benefits are evaluated. Intangible benefits may include:

- Improved public and private decision making in administration, planning, and operations
- Improved information and service to the public
- Increased safety, and reduction in the impact of disasters through better planned evacuation and more efficient management of emergency services
- An improved environment for future generations
- Better presentation of plans and their associated effects
- Improved decisions regarding new development, and better analysis of market and site conditions

The greatest long-range global benefits of GISs are probably in the sectors where decisions have an environmental impact. The environment and the natural relationships within it are complex and not yet fully understood. It is, however, widely known that environmental degradation is implicated in the causes of many modern problems.

Users of GIS

Today, the widespread acquisition of digital computers by businesses, universities, researchers and households has allowed technologies such as GIS to penetrate many aspects of our lives. Nevertheless, computer processing of geographical data remains problematic, and GIS are widely regarded as difficult to learn about and to use. The author hope that subsequent chapters of this book will provide a conceptual and technical understanding of GIS that will allow readers to make effective use of its capabilities in one or more of the many areas of its application.

Users of GIS naturally fall into two groups. Some are professional operators of GIS, who spend much of their lives working with the technology in their jobs. They are well trained in the particular software they use and are well aware of its capabilities. In many cases they do not use the results of their work themselves, but pass them to end users. The results may be maps, designed and produced by the GIS operator, results of analysis to be used in planning harvesting of trees, or work orders for maintenance staff in a major utility company.

The second groups of users spend a relatively small proportion of their lives using GIS. They may maintain a GIS capability on their personal workstation in order to produce an occasional map, to find a park in an unfamiliar city, to plan a driving route for a vacation, or to carry out analysis of map data in connection with a research project. In these cases the opportunities for lengthy training are much less, so the GIS must be simple and easy to use. This second group also comprises end users and primary users who make professional decisions based on GIS products. The group includes:

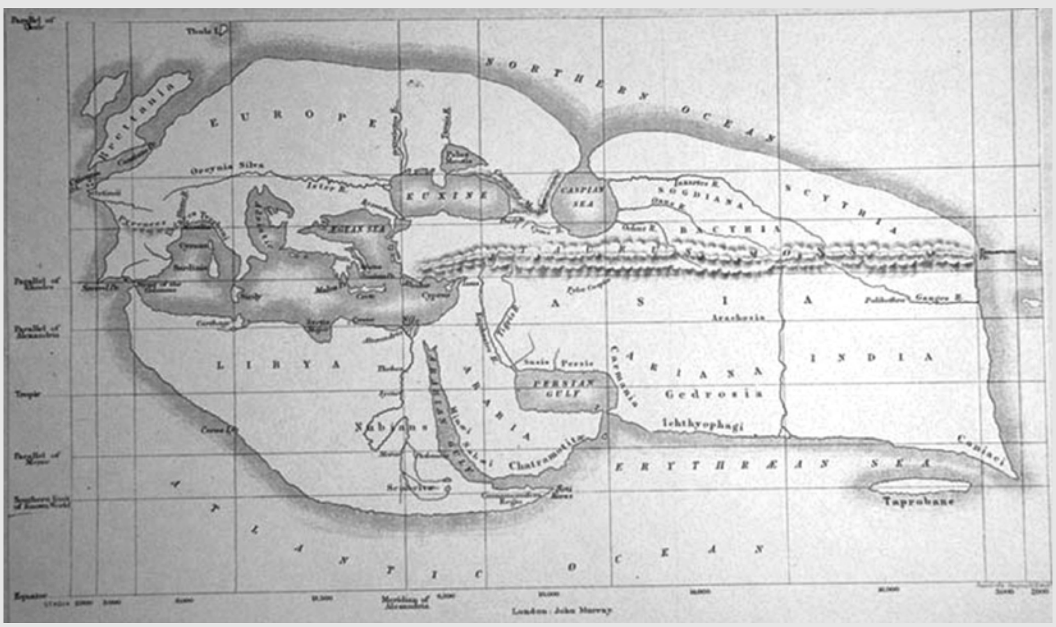
- Operation and maintenance engineers; a typical decision may be whether to replace or repair a damaged water main.

- Regional planners; characteristic tasks involve presentations of plans to municipal authorities in a realistic, varied, visual manner.
- Building authority functionaries; representative jobs include processing building permit applications involving access roads, water supply, or sewage.
- Revenue officials, typically dealing with tax assessment and taxpayer addresses.
- Road engineers, whose responsibilities include locating new roads to minimize cut-and-fill operations.
- Information officers; information produced may include complete packages to newly established firms with details on industrial areas, schools, and transportation.
- Local officials, who may require updated overviews on the effects of effluents on water quality at municipal hand pumps.
- Fire brigades, for whom rapid, reliable information on the locations of fires and the presence of hazards such as explosives would be invaluable.
- Forest managers planning harvest operations, computing volumes of annual growths, estimating road costs and identifying sensitive wildlife areas.

This page intentionally left blank

CHAPTER 2

HISTORY AND DEVELOPMENT OF GIS



Geographical information systems evolved from centuries of mapmaking and the compilation of spatial data. The earliest known maps were drawn on parchment to show the gold mines at Coptes during the reign (1292 – 1225 B.C.) of Rameses II of Egypt. Perhaps earlier still are Babylonian cuneiform tablets that describe the world as it was then known. At a later date, the Greeks acquired cartographic skills and compiled the first realistic maps. They began using a rectangular coordinate system for making maps around 300 B.C. About 100 years later, the Greek mathematician, astronomer, and geographer Eratosthenes (ca. 276 – 194 B.C.) laid the foundations of scientific cartography. One of the earliest known maps of the world was constructed by Claudius Ptolemaeus of Alexandria (ca. A.D. 90 – 168).

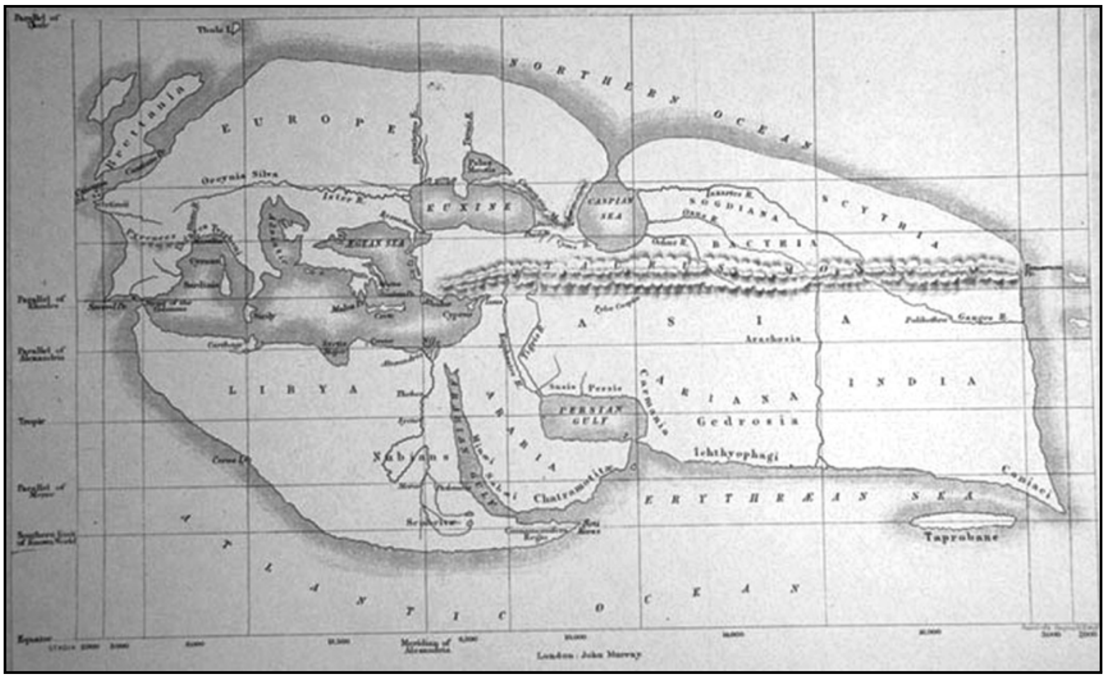


Figure 2.1: The map prepared by Eratosthenes.

The Romans were more concerned with tabulations and registers. The terms cadastre (an official property register) and cadastral (of a map or survey that shows property or other boundaries) originate from the late Greek *kattá-stíkon*, which means 'by line'. But it was the Romans who first employed the concept to record properties, in the *capitum registra*, literally, 'land register'. In many countries, the term cadastre designates map and property registers.

Throughout history, as societies organized, it became necessary to meet the expense of this. Some of the better known earlier examples include taxation levied by emperors and kings to meet military expenses. These direct levies are the foundations of today's complicated revenue systems involving the taxation of income, property, and goods. Since

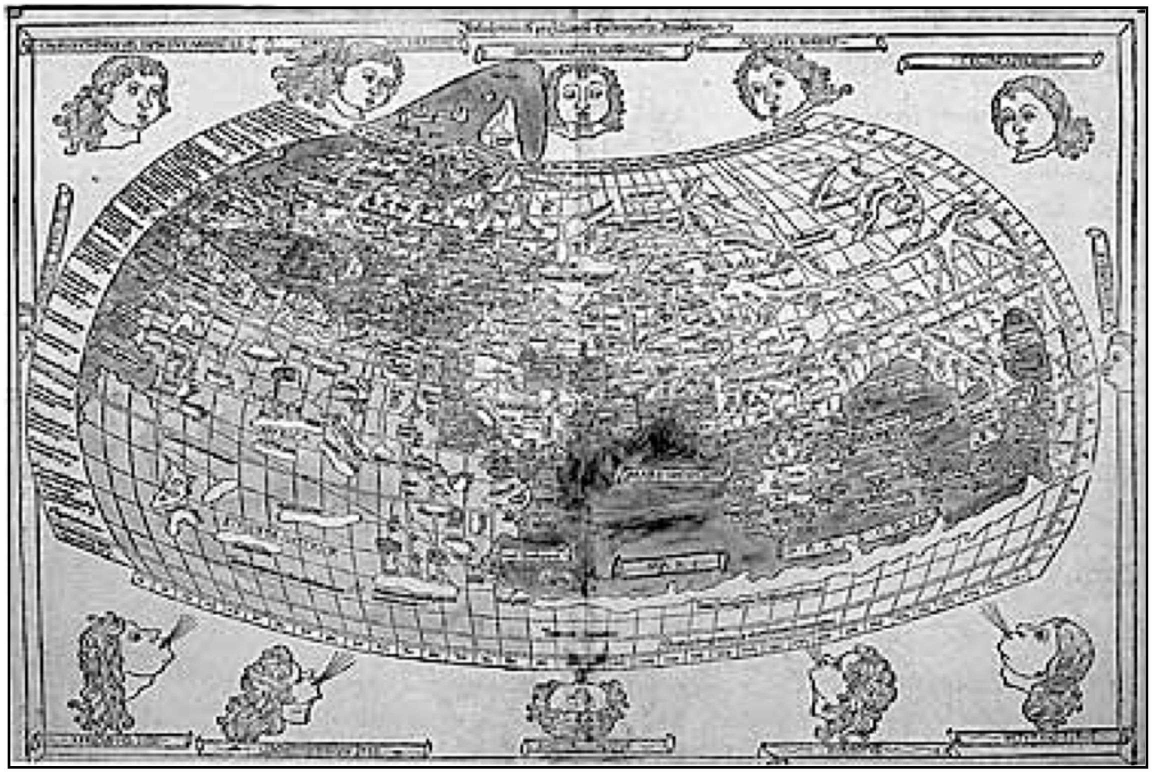


Figure 2.2: Ptolemy's map of the world, about A.D. 150, republished in 1482. Notice the use of latitude and longitude lines and the distinctive projection of this map.

both the ancient Egyptians and the Romans taxed property, property registration was early systematized to assure tax revenues.

The earliest maps were drawn almost exclusively to facilitate commercial sea voyages. On them, coasts were meticulously detailed and harbours were plumbed, while interiors remained unknown, apart from details of important trade and caravan routes.

The Arabs were the leading cartographers of the Middle Ages. European cartography degenerated as the Roman Empire fell. But in the fifteenth century, old skills were revived and Claudius Ptolemaeus's Geographia was translated into Latin to become the then existent view of the world. Although cartography was neglected, in many countries property registry thrived. The best known example is the Domesday Book, the record of the lands of England compiled in 1086 for the first Norman king, William the Conqueror (1027–87). The data included specifications of properties and their value, and a count of inhabitants and livestock, as well as incomes earned and taxes paid.

The travels and explorations of Marco Polo, Christopher Columbus, Vasco da Gama, and others resulted in increased trade. In turn, maps were needed of previously unmapped seas and coasts. As the European countries and the newly discovered regions evolved to

Figure 2.3: Al-Idrisi's map of the world, 1456. He completed a map of the known world in the 12th century. Drawn with south at the top, this later example has been inverted for easier viewing.

more organized societies, the need for geographical information increased. Ordnance developments, such as the introduction of artillery, made maps important in military operations, and military agencies became the leading mapmakers. In many countries, the military mapmakers became responsible for both topographic land maps and navigational charts. Vestiges of this trend remain: map agencies, particularly nautical chart agencies, seem characteristically military. For example, the official mapmaking agency of Great Britain is the Ordnance Survey. The introduction of mass printing techniques enabled maps to be produced as consumer articles rather than as works of art, as was often the case earlier when maps were drawn by hand.

Until the nineteenth century, geographical information was used mostly for trade and exploration by land and sea and for tax collection and military operations. New needs arose

Figure 2.4: World Map of the 15th Century - This map of the known world was produced, probably in late-15th century Genoa, by Paolo Toscanelli, and represents the extent of European knowledge before their exploratory voyages of the 1490s began. It shows that virtually no progress had been made in European geography since the 2nd century, when the Greek geographer Ptolemy collected the information on which this map is based. The Mediterranean coastline is easily recognizable, but the Indian Ocean coast is very inaccurate and the interiors of Asia and Africa are guesswork. It seems that Toscanelli was the first to put forward the idea of reaching Asia by sailing westwards—an idea taken up enthusiastically by Christopher Columbus.

in step with evolving infrastructures, such as roads, railways, telegraph and telephone lines, and gas and water supplies. Planning these facilities required information about the terrain beyond that commonly available. The accurate location of towns and cities, lakes and rivers, mountains and valleys became increasingly important. Detailed topographic information was needed to layout railway and road gradients and curve radii. Then, as now, foundations were a major challenge, so maps showing the type of soil and the quality, location, and properties of bedrock were required. As planning advanced, specialized maps became more common. The first geological map of Paris was compiled in 1811. In 1838, the Irish government compiled a series of maps for the use of railway engineers, which may be regarded as the first manual geographical information system.

Development became increasingly dependent on socio-economic factors. The rights of property owners entered the picture because the construction of airports, large dams, canals, roads and railways often necessitated the expropriation of private lands. New applications arose for property registers and maps as builders needed to compile overviews of affected properties in order that their owners might be justly compensated.

As cities grew larger and more complex, accurate urban planning became a necessity. Many countries began compiling statistical information relating to urban planning in the early nineteenth century. By 1837 the British Registrar General's Office had amassed extensive population statistics. Traditional village property ownership became a hindrance to effective farming. Many properties had become fragmented over the years, owing to inheritance settlements. In some cases a single property might comprise several hundred dispersed parcels of land. Sometimes the ownership of, or rights to a parcel were divided: one owner could have timber rights, another grazing right, and so on. Therefore, property mapping in the late nineteenth century aimed to wrest order from chaos. With reference to available land registers, the various parcels were assembled into properties that were easier to work. Borders were consolidated, clarifying ownership and facilitating the taxation of property.

Aerial photography accelerated the progress of mapmaking. The first aerial photograph was used for mapmaking, and the first mapmaking instrument devised, in 1909. Photogrammetry, the technique of making measurements from photographs, developed rapidly in the 1920s and 1930s, and the two world wars also hastened developments. After World War II, photogrammetry became widely used in mapmaking, mostly for maps in scales from 1:500 to 1:50,000. Aerial photographs themselves became important sources of quantitative information in evaluating such features as vegetation and geological formation.

Advantages of GIS Over Manual Methods

The traditional method of preparing and analyzing maps has been to overlay thematic maps manually to choose areas of coinciding constraints and opportunities. The difficulty with the manual overlay method was that they may be published at different scales or projections. The more layers of maps included in the analysis and the more complex they become, the more the likelihood of human error entering the analysis and the longer the process takes. The GIS can take maps from different sources and register them easily and is consistent in its analysis of multiple layers of map data. It is also faster than manual methods of analysis, allowing the flexibility to try alternate variables in analysis.

First Automatic Processing of Geographical Information

Although Blaise Pascal is credited with devising the first true calculating machine in 1647, large amounts of data were first processed automatically in 1890, when a tabulating device conceived by Hermann Hollerith was used in compiling the U.S. census. In Hollerith's first apparatus, census data were punched on cards which were then read electromechanically to compile data in separate registers. In the first half of the twentieth century, Hollerith's various mechanisms were developed further. Data processing using punch cards became an industry.

During World War II, data processing again advanced, primarily to meet the military need for predicting ballistic trajectories. One of the most famous computers developed for that purpose was ENIAC, an acronym for electronic numerical integrator and calculator. After the war, computer development continued. In 1953, IBM launched the model 650, which became the 'Model T of the Computer Age' by virtue of being the first electronic computer not to be hand-made. More than 100 were produced-in those days an amazing quantity. In today's computer terms, ENIAC, Whirlwind, the IBM 650, and other early electronic computers are referred to as first generation. All first-generation computers suffered from a common drawback: they used vacuum tubes, which, like light bulbs, gave off heat and had limited lifetimes. That alone limited their application. One 25,000 – tube computer of the period was continuously manned by a staff of 10, of which two were technicians assigned to continuous replacement of burned-out tubes. Nonetheless, computerization was the established technology for processing large amounts of data. By 1952, all U.S. governmental statistical data were processed by electronic computers.

By the late 1950s and early 1960s, second-generation computers using transistors became available, outperforming their vacuum-tube predecessors. Suddenly, computers became affordable in disciplines other than those of major governmental agencies. Meteorologists, geologists, and other geophysicists began using electronic mapmaking devices. Initially, the quality was poor, not least because automatic drawing machines had yet to be developed.

As the uses of second-generation computers spread, theoretical models were evolved to use statistical data. Then, as now, public and private decision making was often based on analyses of various classes of geographical data. These included demographic trends, cost-of-living variations, the distribution of natural resources, wealth and social benefits, and the demography of employment. The first geographical information system was constructed by the government of Canada in the late 1960s, and by modern standards was both unbelievably crude and expensive. It required a large mainframe computer, and its output was entirely in the form of tables. This was, in part, because no computer-controlled devices were available at that time to draw maps and in part because of the system's emphasis on analysis. Later, in the United States, a similar system, MIDAS, began processing data on natural resources.

The need for reliable geographical data multiplies with the expansion of road, rail, telecommunications and sewage networks, airports, electricity and water supplies, and other essential services vital to the infrastructure of urban areas. Terrain information on maps is now a vital planning tool, from the first conceptual stage to the final, legally binding plan. Burgeoning road networks have mandated extensive analyses of transport patterns. Indeed, since the mid-1950s, computers have been used in the United States to simulate traffic flows in relation to population distribution.

Important Milestones in the Development of GIS

Development of GIS was influenced by key groups, companies and individuals along with timely development of key concepts. The idea of portraying different layers of data on a series of base maps, and relating things geographically, has been around much longer than computers. Like maps of the Battle of Yorktown (American Revolution) drawn by the French Cartographer Louis–Alexandre Berthier contained hinged overlays to show

Box 3: The stages of GIS development

Stage	*Period*	*Description*	*Characteristics*
The Era of ***Beginning***	*1960 – 1975*	*Pioneering*	• ***individual personalities important*** • ***mainframe-based systems dominant***
The Era of ***Innovation***	*1975 – 1980*	*Experiment and practice*	• ***local experimentation and action*** • ***GIS fostered by national agencies*** • ***much duplication of efforts***
The Era of ***Commercialization***	*1980 – 2000*	*Commercial dominance*	• ***increasing range of vendors*** • ***workstation and PC systems becoming common*** • ***emergence of GIS consultancies***
The Era of ***Exploitation***	*2000 onwards*	*User dominance* *Vendor competition*	• ***embryonic standardization*** • ***increasing use of PC and networked systems*** • ***systems available for all hardware platforms*** • ***internet mapping launched***

Source: *Adopted from Heywood, Cornelius and Carver, 2004.*

troop movements or the mid-19th Century 'Atlas to Accompany the Second report of the Irish Railway Commissioners' showed population, traffic flow, geology and topography superimposed on the same base map similarly, Dr. John Snow used a map showing the locations of death by cholera in central London in September, 1854 to track the source of the outbreak to a contaminated well–an early example of geographical analysis.

But gradually changes started to occur in mapping techniques and following factors caused this change in cartographic analysis:

- Computer technology- improvements in hardware, especially graphics
- Development of theories of spatial processes in economic and social geography, anthropology, regional science
- Increasing social awareness, education levels and mobility, awareness of environmental problems.

The result of these developments were evident from integrated transportation plans of 1950s and 60s in Detroit, Chicago which required integration of transportation information -routes, destinations, origins, time. They ultimately produced maps of traffic flow and volume. Similarly, University of Washington, Department of Geography, research on advanced statistical methods, rudimentary computer programming, computer cartography resulted in developing:

- Nystuen-fundamental spatial concepts-distance, orientation, connectivity
- Tobler-computer algorithms for map projections, computer cartography
- Bunge-theoretical geography-geometric basis for geography-points, lines and areas

- Berry's Geographical Matrix of places by characteristics (attributes)- regional studies by overlaying maps of different themes-systematic studies by detailed evaluation of a single layer

The boost to GIS development began in mid 1960s, when Canada Geographic Information System (CGIS) made a massive effort. The Canada Land Inventory an effort by the federal and provincial governments to identify the nation's land resources and their existing and potential uses. The most useful results of such an inventory are measures of area, yet area is notoriously difficult to measure accurately from a map. CGIS was planned and developed as a measuring tool, a producer of tabular information, rather than as a mapping tool.

The second burst of activity occurred in the late 1960s in the US Bureau of the Census, in planning the tools needed to conduct the 1970 Census of Population. The DIME program (Dual Independent Map Encoding) created digital records of all US streets, to support automatic referencing and aggregation of census records. The similarity of this technology to that of CGIS was recognized immediately and led to a major program at Harvard University's Laboratory for computer graphics and spatial analysis to develop a general purpose GIS that could handle the needs of both applications. The project led eventually to the ODYSSEY GIS of the late 1970s.

In a separate development, cartographers and mapping agencies had begun in the 1960s to ask whether computers might be adapted to their needs and possibly to reducing the costs and shortening the time of map creation. National mapping agencies, such as the UK's Ordnance Survey, France's Institut Géographique National, the US Geological Survey and US Defence Mapping Agency began to investigate using computers to support the editing of maps, to avoid the expensive and slow process of hand correction and redrafting. The first automated cartography developments occurred in the 1960s, and by the late 1970s most major cartographic agencies were already partly computerized. Remote sensing also played a significant part in the development of GIS, as a source of technology as well as a source of data. GIS really began to take off in the early 1980s, when the price of computing hardware had fallen to a level that could sustain a significant software industry and cost effective applications.

The Microprocessor

In the 1960s and early 1970s, integrated circuits were developed and computer programs refined. The result: third-generation computers which brought computerization to virtually all professional disciplines, especially those processing large amounts of data.

The next major breakthrough came in 1971-1972 with the development of the microprocessor. In 1974, a microprocessor was used to build the first fourth-generation desktop computer. Seven years later, the first microprocessor-based desktop computer was launched as a personal computer (PC). By the mid–1980s, the computer field was divided into three categories according to size of computer: mainframes, the descendants of the original large computers, intended for major data processing and computational tasks; PCs, the increasingly ubiquitous desktop computers; and minicomputers/workstations, which were smaller than mainframes but larger than desktop PCs. By the early 1990s, mainframes had become physically smaller and computationally more capable. That trend was reflected strongly in PCs, which by 1990 were outperforming minicomputers built only a few years earlier. This development signalled the demise of the minicomputer, an event that was further hastened by the introduction of PC networks, in which processing and storage capacity may be shared and distributed. The development of powerful workstations in the mid–1980s, however, led to an increasing acceleration in the use of GIS. The overall trend is best illustrated in terms of the costs of computing: a computer's processing and storage capabilities. In other words, cost efficiency increased by a factor of 10 every two to three years.

In the 1970s and 1980s, various systems were evolved to replace manual cartographic computations. Workable production systems became available in the late 1970s and system development continued through the 1980s. Nonetheless, by the mid–1990s, elegant approaches to some cartographic tasks have yet to be found, and computerized cartographic research and development remains a continual challenge. The spread of PCs spurred user-friendly operations and programs capable of processing in ways previously not possible, for example, by considering the logical connections in geographical data.

Increases in microprocessor computing capacity also made the processing of digital and satellite images and other types of raster images commercially available in the mid-1980s. Software systems have developed apace. Relational database systems, such as dBase and Oracle which first appeared in the late 1980s, are particularly useful in processing geographical data. Commercially available relational databases are now used routinely in GIS systems.

In the late 1980s, computing capability became widely accessible as microprocessors were used for a multitude of devices, from household appliances and automobiles to an extensive range of specialized instruments, including those used in GIS. For GIS users, microprocessors have improved such devices as:

- Surveying instruments
- GPS (global positioning system)
- Digitizing table
- Scanners
- Environmental monitoring satellites and data presentation systems, including graphic displays, electrostatic plotters, and laser printers

Recent Developments

The 1990s have produced even faster and more powerful computer equipment and peripherals. However, new developments in the field of data networks and communications are of equal importance, specifically local area network (LAN), wide area network (WAN), and last, but certainly not least, Internet and World Wide Web (WWW). The development of the Internet was initiated by the U.S. Department of Defence as long ago as the late 1960s.

World Wide Web was developed at CERN (European Organization for Nuclear Research) in Switzerland in 1990.

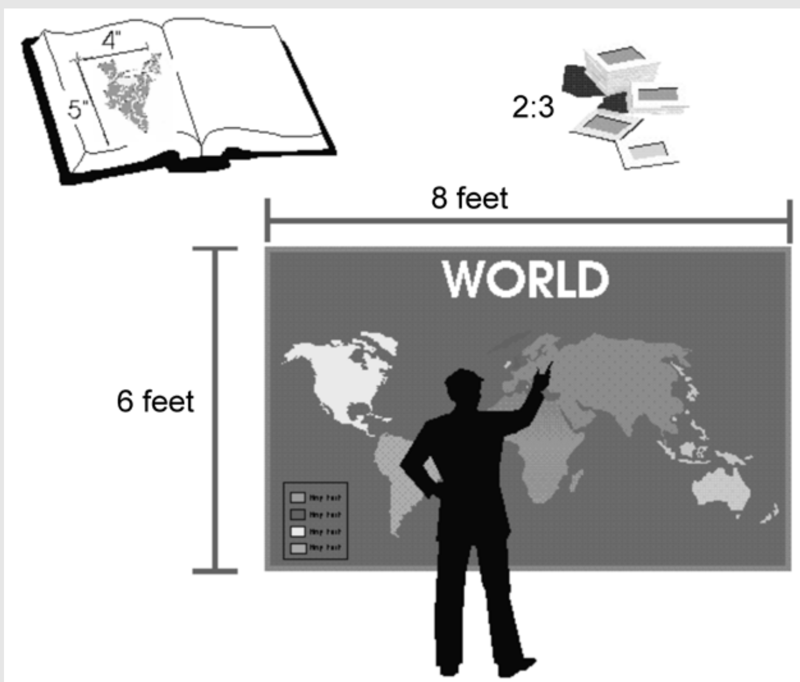
Data networks have opened up a whole new range of opportunities for geographical data search and distribution, thereby considerably increasing the value of GIS, particularly since common data have become more easily accessible. The most spectacular development in the GIS arena has occurred in the field known as multimedia. Multimedia techniques are based on the combination of elements such as figures, text, graphics, pictures, animation, sound, and video. Multimedia brings geospatial information into living maps and makes complex information understandable to those who are not technically sophisticated. Multimedia technology is available on the Internet and has proved to be eminently suitable as an information tool in planning city, roads, tourism, and the distribution of environmental information.

Flight simulators are perhaps the best-known example of the application of data technology to create near real-life situations, thus making them ideal for use in training. The concept behind flight simulators has now been adopted for other activities and is known as virtual reality. Virtual reality is a term used in multimedia; it was marketed as fiction in 1984 and became commercial technology in 1992. The most extreme virtual reality experience is attained by 'dressing up' in a computer and moving into a world where almost all external impulses are artificial. Virtual reality and GIS have many features in common and are becoming more and more integrated.

This page intentionally left blank

Chapter 3

GISs Roots in Cartography



Human Cognition of the Spatial World

At human scales, the world consists of objects, events, processes, and a background environment. The study of cognition is about knowledge: its acquisition, storage and retrieval, manipulation, and use by humans and other intelligent creatures. Cognition includes sensation and perception, thinking, imagery, reasoning and problem-solving, memory, learning, and language. Cognitive structures and process are part of the mind, which emerges from a brain and nervous system inside of a body that exists in a social and physical world.

Spatial cognition deals with the cognition of spatial properties of the world, including location, size, distance, direction, shape, pattern, movement, and inter-object relations.

Sensing and Perceiving the World

Sensation is the first response of the nervous system to stimulation from patterned energy in the world. Sensory systems are organized into *modalities,* including vision, hearing, smelling, tasting, pressure and texture, temperature, kinesthesis (limb position and movement), and vestibular senses (gravity and body acceleration). Perception is the active acquisition of knowledge about the self and the world through the senses.

Characteristics of the perceived world:

- Locational perspective – world perceived from a point-of-view, incomplete access to world
- Redundancy of information (*e.g.*, depth cues of interposition and linear perspective)
- Constancy (colour, size, position, shape) – objects, events, and background maintain many characteristics even as viewing conditions change
- Meaningfulness – tendency to perceive meaningful, familiar objects and events

Cognitive Maps

Cognitive maps are internal representations of the world and its spatial properties stored in memory (also called 'mental maps'). Like – what's out there, what are its attributes, where it is, how to get there. These are both idiosyncratic to individuals, and shared among groups. It is not like a cartographic map in the head. It is not a unitary representation with a constant scale neither completely integrated. It consists of discrete pieces (more vector than raster), *e.g.,* landmarks, routes and regions. The pieces determined by physical, perceptual, or conceptual boundaries. They are hierarchically organized pieces with multiple levels of pieces differing in status (*e.g.,* size). The cognitive maps have distortions, which tell us about properties of cognitive maps and correspondence to physical measurement for example, Sri Lanka is thought to be due south of India where as it is actually southeast or turns are remembered more like right angles and curved lines are often straightened.

Spatial Learning and Development

Learning is a relatively permanent change in cognition or behaviour that results from practice or experience. Spatial knowledge is learned via one or more *media of acquisition*. Direct sensorimotor experience, maps, models, photos and drawings, movies and videos, verbal and written language. Cognitive development is systematic change in the content and process of cognition over time, including learning, maturation, and growth (child or adult). Child psychologist Piaget known for a qualitative 'stage theory' of cognitive development of children suggests of change from concrete sensorimotor space in infancy to abstract spatial reasoning in adolescence. Here 'frame of reference' used to define locations changes from egocentric (self-centered) to allocentric (externally referenced) and geometry of spatial knowledge changes from topological to projective and metric. Information-processing approach provides an alternative theory of continuous and quantitative development.

Traditional theory of developmental sequence in spatial knowledge of the world inspired by Piagetian theory; consists of 3 stages or elements, acquired over time:

- First is 'landmark knowledge': unique patterns of perceptual events that identify a place.
- Second is 'route knowledge': sensorimotor routines that connect ordered sequences of landmarks; little or no metric spatial knowledge.
- Third is 'survey knowledge': two-dimensional layout knowledge of simultaneous interrelations of locations; allows detouring, shortcutting, and creative navigation.

Information-processing approach inspires an alternative sequence of continuous and quantitative increase in extent, accuracy, and completeness of sometimes crude metric spatial knowledge.

Navigation

Navigation is coordinated and goal – directed route following through space. It consists of 2 components: locomotion and way-finding. Locomotion is guidance through space in response to local sensorimotor information in immediate surrounds. It finds support surfaces, avoid obstacles and barriers, follow beacons, move through openings. Way-finding is planning and decision – making in response to non – local information, undertaken to reach goal.

Using and Learning Maps

The main purpose of cartographic maps is to communicate geographical information and support geographical problem-solving. Humans have the ability to quickly extract great amounts of information from spatial depictions (images) like pictures or graphs. Even non-spatial or non-perceivable information can be displayed this way (visualization or spatialization). Maps use convenient scales and viewing perspectives (we can perceive all from a single viewpoint). Maps highlight and clarify relevant properties; omit or downplay irrelevant properties. But projections, generalizations, exaggerations, omissions may mislead or distort knowledge in a map. Similarly, perspective translation from overhead to terrain-level view may be confusing or interpretation of symbols (colours, point symbols, contour lines) may be difficult or misleading. However, training and experience with maps changes the way they are perceived and interpreted.

Spatial Language

Spatial information often expressed verbally, giving verbal directions, spatial descriptions in stories, road signs, and computer queries. Producing spatial language often requires translation of nonverbal spatial knowledge, which can alter the knowledge. Language expresses mostly non – quantitative or imprecise quantitative (fuzzy) information about space; connections and general location more important for example, we say 'turn left at the railway station', not 'turn 80° after you go 1.4 kilometres'. Here quantitative precision usually unnecessary or even confusing for verbal communication but context is critical in interpreting spatial language. Context provided by who is speaking, situation, preceding events, etc.

Relevance to GIS

GISs are frequently difficult to use effectively and efficiently and have not nearly reached their potential, it is more difficult and unpleasant and does not perform all of the tasks that it might. However limitations and problems could be improved with greater attention to cognitive issues in GIS. Cognitive issues touch on all three major functions of GIS: the storage, representation, and analysis of earth-referenced data. Some examples of cognitive issues in GIS:

- How experts and laypeople conceptualize and reason about geographical space, and how GIS can be designed and taught to support both classes of users.
- How people express spatial information in natural language (such as English), and how this can be used to understand communication with a GIS in natural language (such as a navigation computer inside a car).
- How interfaces should be designed to promote accurate and efficient communication of spatial and geographic information, such as scale, uncertainty, and network structure.

GIS and Spatial Cognition

GIS are tools for supporting human decision-making, in applications such as car navigation systems, electronic atlases, GIS are tools to help people acquire spatial information, learn about geography. The interface between the GIS and the user is a filter which determines how successfully information can be transferred. The effective user interfaces depends on how people learn and reason with spatial informations.

Maps are the main source of data for GIS, the traditions of cartography is fundamentally important to GIS. GIS has roots in the analysis of information on maps, and overcomes many of the limitations of manual analysis.

Box 4: What is a map?

Map is a representation, normally to scale and on a flat medium, of a selection of material or abstract features on, or in relation to, the surface of the Earth.

Cartographic Abstraction

Production of a map requires: ***selection*** of the few features in the real world to include, ***classification*** of selected features into groups (*i.e.,* bridges, houses, railways), ***simplification*** of jagged lines like coastlines, ***exaggeration*** of features to be included that are to small to show at the scale of the map and ***symbolization*** to represent the different classes of features chosen.

Role of Maps

Traditionally, maps have four roles today

- Data display-maps provide useful ways of displaying information in a meaningful way.
- Data storage - as a means of storing data.
- Spatial indexes - a map can show *the boundaries* of areas (*e.g.,* land use zones, soil or rock types) and identify each area with a label.
- Data analysis tool - maps are used in analysis to make or test hypotheses and examine the relationship between two distributions using simple transparent overlays.

Changeover to Computer Mapping

Impetus for change began in two communities

1. Scientists wishing to make maps quickly to see the results of modelling, or to display data from large archives already in digital form, *e.g.,* census tables.
2. Cartographers seeking to reduce the cost and time of map production and editing.

GIS and Computer Cartography

Computer cartography has a primary goal of producing maps, systems have advanced tools for map layout, placement of labels, large symbol and font libraries, interfaces for expensive, high quality output devices. However, it is not an analytical tool, therefore, unlike data for GIS, cartographic data does not need to be stored in ways which allow, for example, analysis of relationships between different themes such as population density and housing prices or the routing of flows along connecting highway or river segments.

Contd...

GIS's Advantage over Maps

Data Storage

- Spatial data stored in digital format in a GIS allows for rapid access for traditional as well as innovative purposes.
- The nature of maps creates difficulties when used as sources for digital data.
- Most GIS take no account of differences between datasets derived from maps at different scales.
- Idiosyncrasies (*e.g.,* generalization procedures) in maps become "locked in" to the data derived from them.

Data Indexes

- This function can be performed much better by GIS due to the ability to provide multiple and efficient cross-referencing and searching.

Data Analysis Tool

- GIS is a powerful tool for map analysis.
- Traditional impediments to the accurate and rapid measurement of area or to map overlay no longer exist.

Data Display Tool

Electronic display offers significant advantages over the paper map

- Ability to browse across an area without interruption by map sheet boundaries.
- Ability to zoom and change scale freely.
- Potential for the animation of time dependent data.
- Display in "3 dimensions" (perspective views), with "real-time" rotation of viewing angle.
- Potential for continuous scales of intensity and the use of colour and shading independent of the constraints of the printing process, ability to change colours as required for interpretation.

Defining A Map

According to the International Cartographic Association, a map is a representation, normally to scale and on a flat medium, of a selection of material or abstract features on, or in relation to, the surface of the Earth. The term 'map' is often used in mathematics to convey the notion of transferring information from one form to another, just as cartographers transfer information from the surface of the Earth to a sheet of paper. The term 'map' is used loosely to refer to any visual display of information, particularly if it is abstract, generalized or schematic. Cartography is very much a process of abstraction in which features of the real world are generalized or simplified to meet the demands of the theme and audience. Not all elements or details have a bearing on the pattern or process being studied and so some are eliminated to draw the reader's attention to those facts that are relevant. Too much detail can even hide or disguise the message of a map. The amount of detail that can be included is very much dependent on the scale at which the map will be produced, as the following examples demonstrate. A small-scale map of an area must, almost of necessity, be more generalized.

Map show only a static situation and easy to use to answer certain types of questions: like, how do I get there from here or what is at this point? But it is difficult or time-consuming to answer other types: like, what is the area of this lake or what does that thematic map show at the point I'm interested in on this topographic map?

Production of a map requires selection of the few features in the real world to include, classification of selected features into groups (*i.e.,* roads, houses, railways), simplification of jagged lines like river meandering, exaggeration of features to be included that are to small to show at the scale of the map and symbolization to represent the different classes of features chosen. Maps provide useful ways of displaying information in a meaningful way. In practice, the cost of making and printing a map is high, so its contents are often a compromise between different needs.

Types of Maps

Generally, in practice normally there are two types of maps

- *Topographic map:* These maps are a reference tool, showing the outlines of selected natural and man-made features of the Earth, often acts as a frame for other information. 'Topography' refers to the shape of the surface, represented by contours and/or shading, but topographic maps also show roads and other prominent features.
- *Thematic map:* These maps are a tool to communicate geographical concepts such as the distribution of population densities, climate, land use etc. Thematic maps are important in GIS. An area class map shows zones of constant attributes, such as vegetation, soil type, or forest species. The boundaries are different for each map as they are determined by the variation of the attribute being mapped, *e.g.,* breaks of soil type may occur independently of breaks of vegetation.

Map type is not just characteristics of the map but can be determined by use, *e.g., can* look at *distribution* of major roads on a general-reference topographic map or *can* find specific *location* of observation units (like district) on a thematic map. The classification of maps can be made on the basis of content of the map (climate, socio-economic...), form of the map (dot, choropleth, animated...), display technology used (electronic, paper,...), production technology used (manual, automated,...), scale of the map (large, medium, small), resolution of the map (country, state,...) etc.

Other Representations of the World

Maps are not the only representation of the world; others include: air photos, satellite imagery, drawings and artwork, verbal description, tables etc. But the uniqueness of the map among representations of the world lies in content as well as area shown is selective (unlike air photos, satellite imagery, snapshots), maker has control over emphasis, (unlike air photos, satellite imagery, snapshots), emphasis is on spatial relations (unlike drawings and artwork, in which spatial relations support some other message), it is an analogue of what is represented (unlike words, tables, and digital data). However, maps still remain an excellent way of compiling spatial information, it can be designed to be easy to convert to digital form, *e.g.,* by the use of different colours which have distinct signatures when scanned by electronic sensors. However, consistent, accurate retrieval of data from maps is difficult and only limited amounts of data can be shown due to constraints of the paper medium.

GIS and Computer Cartography

Computer cartography has a primary goal of producing maps. Systems have advanced tools for map layout, placement of labels, large symbol and font libraries, interfaces for expensive, high quality output devices. However, it is not an analytical tool, therefore, unlike data for GIS, cartographic data does not need to be stored in ways which allow, for example, analysis of relationships between different themes such as population density and housing prices.

In GIS spatial data stored in digital format allows for rapid access for traditional as well as innovative purposes. The nature of maps creates difficulties when used as sources for digital data but most GIS take no account of differences between datasets derived from maps at different scales. Idiosyncrasies (*e.g.,* generalization procedures) in maps become 'locked in' to the data derived from them. The prime differences between a GIS and computer cartography are in their functional components:

A GIS contains these four components:

a. Input b. Database c. Analysis d. Output

In contrast, a mapping (cartographic) system can be described in three components:

i. Input ii. Map design iii. Output

This difference is best shown in a software query that lists element attributes

Table 3.1: Comparison between computer cartography and GIS.

Computer Cartography	GIS
Feature type	Area (m^2)
Boundary colour	Perimeter (m)
Pattern	Land use/Land cover
Fill colour	Residential
Design level	Average plot area

...in other words, a cartographic query gives information on design features, while a GIS query yields details or parameters about the features themselves, where the data are stored in a GIS database. We can't 'ask' a cartographic map to display where residential areas are or the houses with area of more than 500 square meters. However, we can ask a GIS the same question and it will display where the query is true.

Table 3.2: Capabilities of different mapping softwares.

Mapping Software	Capabilities			
	Means of data input	*Database management system (DBMS)*	*Analysis capability*	*Graphics output*
CAD Computer Assiting Drawing *e.g.* AutoCad, Microstation	√	X	X	√
GDS Graphic Design System *e.g.* CorelDraw, Illustrator	√	X	X	√
DBMS Database Management System *e.g.* Oracle, Sybase	√	√	X	X
DIPS Digital Image Processing System *e.g.* PCI, Erdas	√	√	√	√
GIS Geographical Information System *e.g.* ArcGis, Ilwis, MapInfo	√	√	√	√

Each of these may perform better than a GIS at their specialty, but only a GIS has all four components, *e.g.,* mapping software may be better for map production, databases for database management.

Table 3.3: Major advantages and disadvantages of computer cartography.

Advantages	**Disadvantages**
Lower cost for simple maps, faster production	Relatively few full - scale systems have been shown to be truly cost - effective in practice.
Greater flexibility in output easy scale or projection change maps can be tailored to user needs	Computer methods do not ensure production of maps of high quality. There is a perceived loss of regard for the 'cartographic tradition' with the consequent production of 'cartojunk'.
Other uses for digital data	High capital cost, though this is now much reduced.

Mapping Concepts, Features & Properties

A map represents geographic features or other spatial phenomena by graphically conveying information about locations and attributes. Locational information describes the position of particular geographic features on the Earth's surface, as well as the spatial relationship between features, such as the shortest path from a bus station to a hospital, the proximity of competing businesses, and so on. Attribute information describes characteristics of the geographic features represented, such as the feature type, its name or number and quantitative information such as its area or length.

Thus the basic objective of mapping is to provide

- Descriptions of geographic phenomenon
- Spatial and non spatial information
- Map features like Point, Line, & Polygon

MAP FEATURES: Locational information is usually represented by points for features such as wells and schools, lines for features such as streams, roads and contour lines and areas for features such as lakes, cultivated lands and census tracts.

SCALE IN DIGITAL MAPS: With digital maps, the traditional concept of scale in terms of distance does not apply because digital maps do not remain fixed in size. They can be displayed or plotted at any possible magnification. Yet we still speak of the scale of a digital map.

In digital mapping, the term scale is used to indicate the scale of the materials from which the map was made. For example, if a digital map is said to have a scale of 1:100,000, it was made from a 1:100,000-scale paper map.

However, a digital map's scale still allows us to make some educated guesses about its contents because, generally, digital maps retain the same accuracy and characteristics as their source maps. So it is still true that a large-scale digital map will usually be more accurate and less general than a small-scale digital map. Because the display size of a computer-based map is not fixed, users are often tempted to blow up maps to very large sizes. For example, a 1:100,000-scale map can easily be plotted at a size of 1:24,000 or even 1:2,000-but it usually is not a good idea to do so.

Data collected at a specific scale are suitable for mapping and analysis only at similar scales

- At smaller scales, large scale data are too complex (but could be generalised).
- At larger scales, small scale data are too generalized (detail cannot be 'added').

As scale is reduced – fewer elements, fewer details can be displayed. Area features in large scale become points and lines in small scale map, like a city or river in large scale becomes point and line respectively in small scale maps.

MAP RESOLUTION: Map resolution refers to how accurately the location and shape of map features can be depicted for a given map scale. Scale affects resolution. In a larger-scale map, the resolution of features more closely matches real-world features because the extent of reduction from ground to map is less. As map scale decrease, the map resolution diminishes because features must be smoothed and simplified, or not shown at all.

MAP ACCURACY: Many factors besides resolution, influence how accurately features can be depicted, including the quality of source data, the map scale, draftsman skills and the width of lines drawn on the ground. A fine drafting pen will draw line's 1/100 of an inch wide. Such a line represents a corridor on the ground, which is almost 53 feet wide.

In addition to this, human drafting errors will occur and can be compounded by the quality of our source maps and materials. A map accurate for one purpose is often inaccurate for others since accuracy is determined by the needs of the project as much as it is by the map itself.

Some measurements of a map's accuracy are discussed below

- ***Absolute accuracy*** of a map refers to the relationship between a geographic position on a map (a street corner, for instance) and its real-world position measured on the surface of the earth. Absolute accuracy is primarily important for complex data requirements such as those for surveying and engineering-based applications.
- ***Relative accuracy*** refers to the displacement between two points on a map (both distance and angle), compared to the displacement of those same points in the real world. Relative accuracy is often more important and easier to obtain than absolute accuracy because users rarely need to know absolute positions. More often, they need to find a position relative to some known landmark, which is what relative accuracy provides. Users with simple data requirements generally need only relative accuracy.
- ***Attribute accuracy*** refers to the precision of the attribute database linked to the map's features. For example, if the map shows road classifications, are they correct? If it shows street addresses, how accurate are they? Attribute accuracy is most important to users with complex data requirements.
- A map's ***Currency*** refers to how up-to-date it is. Currency is usually expressed in terms of a revision date, but this information is not always easy to find.
- A map is ***Complete,*** if it includes all the features a user would expect it to contain. For example, does a street map contain all the streets? Completeness and currency usually are related because a map becomes less complete as it gets older.

The most important issue to remember about map accuracy is that the more accurate the map, the more it costs in time and money to develop. For example, digital maps with coordinate accuracy of about 100 feet can be purchased inexpensively. If 1-foot accuracy is required, a custom survey is often the only way to get it, which drives up data-acquisition costs by many orders of magnitude and can significantly delay project implementation – by months or even years.

Therefore, too much accuracy can be as detrimental to the success of a GIS project as too

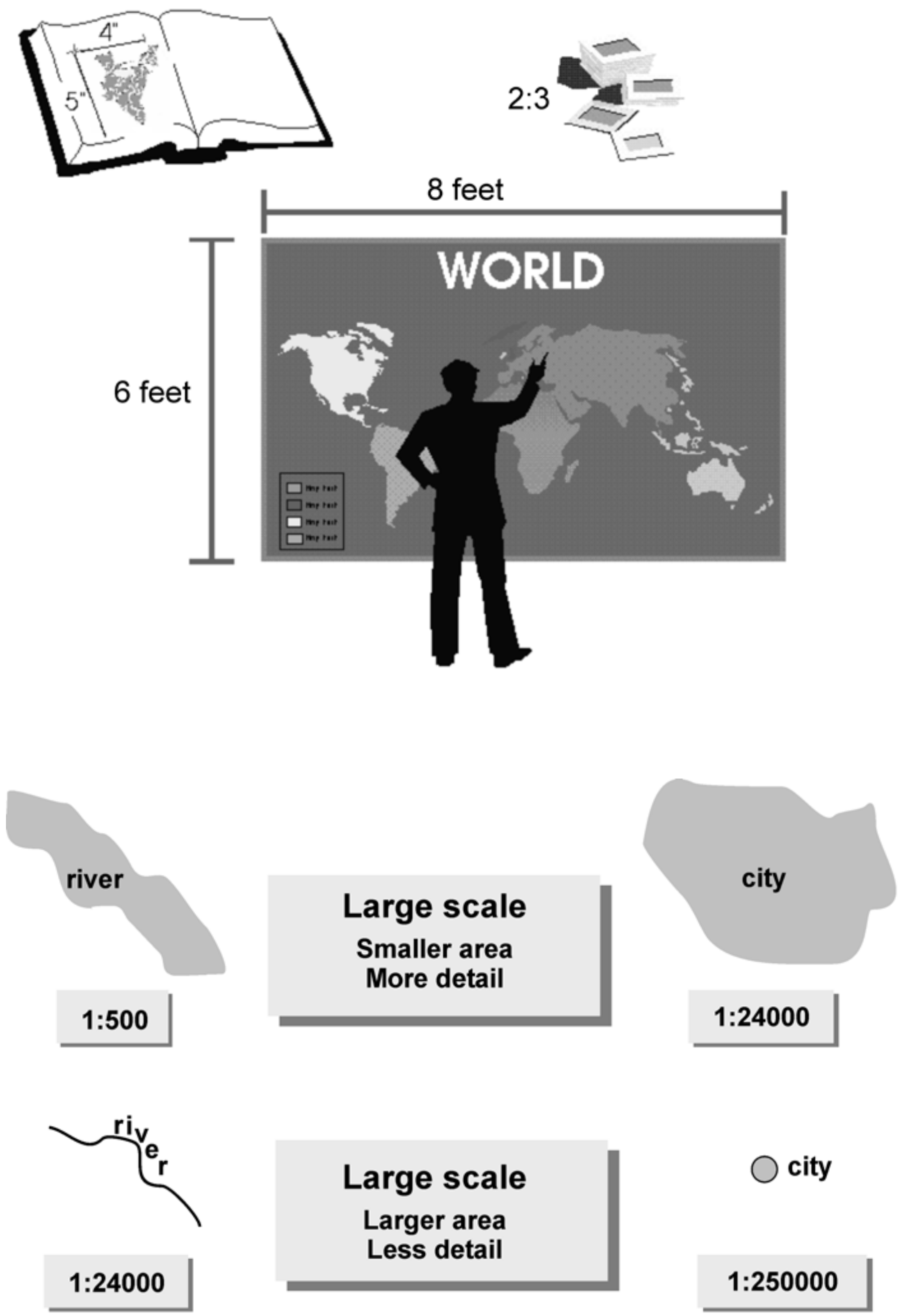


Figure 3.1: Map scale determines the size and shape of features.

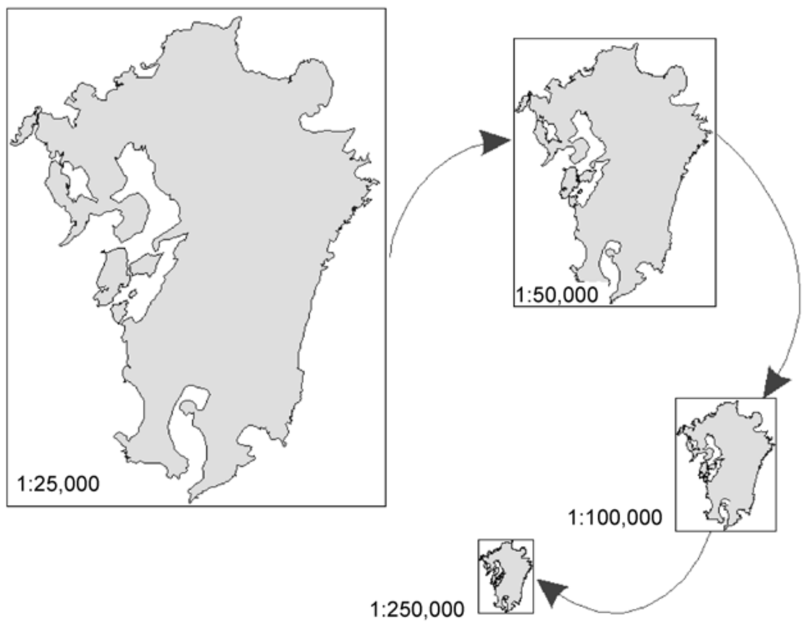


Figure 3.2: The details are blurred as the scale decreases.

little. Rather than focusing on the project's benefits, a sponsoring organization may focus on the costs that result from a level of accuracy not justified for the project. Project support inevitably erodes when its original objectives are forgotten in a flurry of cost analyses.

A far better strategy is to start the project with whatever data is readily available and sufficient to support initial objectives. Once the GIS is up and running, producing useful results, project scope can be expanded. The quality of its data can be improved as required.

Even though no maps are entirely accurate, they are still useful for decision-making and analysis. However, it is important to consider map accuracy to ensure that our data is not used inappropriately.

Any number of factors can cause error. Note these sources can have at cumulative effect.

$$E = f(f) + f(I) + f(e) + f(d) + f(a) + f(m) + f(rms) + f(mp) + u$$

Where,

f = flattening the round Earth onto a two-dimensional surface (transformation from spherical to planar geometry)

I = accurately measuring location on Earth (correct project and datum information)

c = cartographic interpretation (correct interpretation of features)

d = drafting error (accuracy in tracing of features and width of drafting pen)

a = analog to digital conversion (digitizing board calibration)

m = media stability (warping and stretching, folding. Wrinkling of map)
p = digitizing processor error (accuracy of cursor placement)
rms = Root Mean Square (registration accuracy of ties)
mp = machine precision (coordinate rounding by computer in storing and transforming)
u = additional unexplained source error.

MAP EXTENT: The aerial extent of map is the area on the Earth's surface represented on the map. It is the limit of the area covered, usually defined by rectangle just large enough to include all mapped features. The size of the study area depends on the map scale. The smaller the scale the larger the area covered.

DATABASE EXTENT: A critical first step in building a geographic database is defining its extent. The aerial extent of a database is the limit of the area of interest for the GIS project. This usually includes the areas directly affected by the organization's responsibility (such as assigned administrative units) as well as surrounding areas that either influence or are influenced by relevant activities in the administrative area.

DATA AUTOMATION: Map features are logically organized into a set of layers or themes of information. A base map can be organized into layers such as roads, soils, land use/land cover or state boundaries. Map data, regardless of how a spatial database will be applied, is collected, automated and updated as series of adjacent map sheets or aerial photograph. Here each sheet is mounted on the digitizer and digitized, one sheet at a time. In order to be able to combine these smaller sheets into larger units or study areas, the co-ordinates of coverage must be transformed into a single common co-ordinate system. Once in a common coordinate system, attributes are associated with features. Then as needed map sheets for layer are edge matched and joined into a single coverage for our study area.

TYPES OF INFORMATION IN A DIGITAL MAP

Any digital map is capable of storing much more information than a paper map of the same area, but it's generally not clear at first glance just what sort of information the map includes. For example, more information is usually available in a digital map than what we see on-screen. And evaluating a given data set simply by looking at the screen can be difficult: What part of the image is contained in the data and what part is created by the GIS program's interpretation of the data? We must understand the types of data in our map to be used it appropriately.

Three general types of information can be included in digital maps

- Geographic information, which provides the position and shapes of specific geographic features.
- Attribute information, which provides additional non-graphic information about each feature.
- Display information, which describes how the features will appear on the screen.

Some digital maps do not contain all three types of information. For example, raster maps usually do not include attribute information, and many vector data sources do not include display information.

The fundamental characteristic of GIS is its ability to handle spatial data. GIS not only analyse and display spatial data but also the relationship among spatial data are analysed. The analysis of spatial data are possible only when we transform the real world data into GIS, using precisely defined *coordinate system* and a *map projection*. But before discussing these features, we will look into basic properties of the earth – shape, size, geometry etc. and know how earth is measured and modelled for the purpose of positioning in GIS.

The Shape of the Earth

From the early civilizations, the effort for determining shape and size of earth was a major challenge to humans. Eratosthenes, a Greek geographer, gave the notion of spherical earth in second century B.C. But now researchers have confirmed that earth's surface is not spherical or flat rather it is oblate ellipsoidal, which means all points on the surface of the earth are not equidistant from the geometric centre. The radius to the poles is slightly less to equator (approximately 21 kilometres lesser). The flattening of the ellipse for the earth is only 1/297, but it necessary to take care in calculations for plotting accurate maps on large scales. On small scales, this oblateness is negligible but even then for making transformations and geometric relations in GIS, necessary adjustments are essential.

DATUMS: Datums define the reference systems that describe the size and shape of the earth, and the origin and orientation of the coordinate systems used to map the earth. Hundreds of different datums have been used to frame position descriptions since the first estimates of the earth's size were made by Aristotle. Datums have evolved from those describing a spherical earth to ellipsoidal models derived from years of satellite measurements.

Modern geodetic datums range from flat-earth models used for plane surveying to complex models used for international applications which completely describe the size, shape, orientation, gravity field, and angular velocity of the earth. While cartography, surveying, navigation, and astronomy all make use of geodetic datums, the science of geodesy is the central discipline for the topic. Referencing geodetic coordinates to the wrong datum can result in position errors of hundreds of meters. Different nations and agencies use different datums as the basis for coordinate systems used to identify positions in geographic information systems, precise positioning systems, and navigation systems. The diversity of datums in use today and the technological advancements that have made possible global positioning measurements with sub-meter accuracies requires careful datum selection and careful conversion between coordinates in different datums.

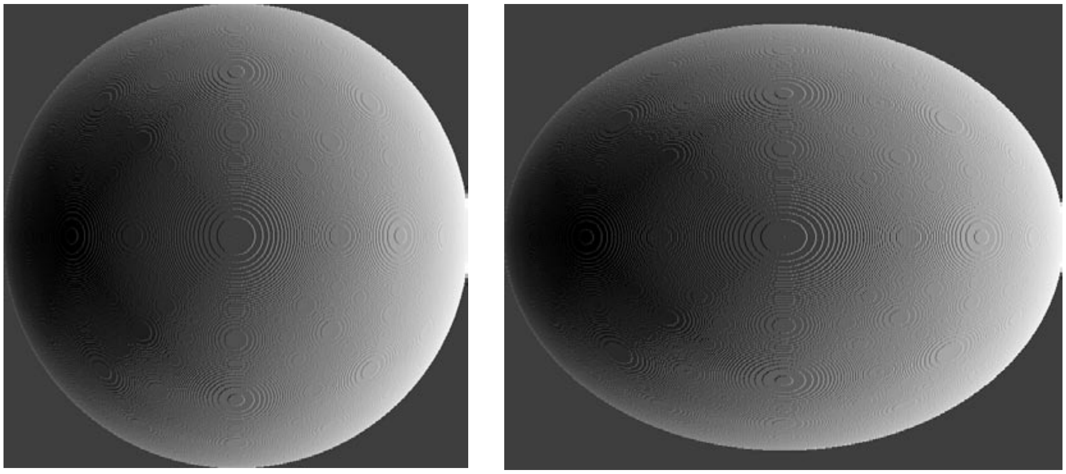


Figure 3.3: Earth shape: sphere or ellipsoid.

Geodetic datums and the coordinate reference systems based on them were developed to describe geographic positions for surveying, mapping, and navigation. Through a long history, the 'figure of the earth' was refined from flat-earth models to spherical models of sufficient accuracy to allow global exploration, navigation and mapping. True geodetic datums were employed only after the late 1700s when measurements showed that the earth was ellipsoidal in shape.

Datum Types

1. *Horizontal:* Datums that define the relationship between the physical earth and horizontal coordinates such as latitude and longitude. Examples include the North American Datum of 1927 (NAD27) and the European Datum 1950 (ED50).
2. *Vertical:* Datums that define level surfaces. Examples include the National Geodetic Vertical Datum of 1929 (NGVD29) and the North American Vertical Datum of 1988 (NAVD88). Some are based on sea-level measurements and levelling networks (NGVD29), others on gravity measurements (NAVD88).
3. *Complete:* Datums that describe both vertical and horizontal systems. Some, such as World Geodetic System 1984 (WGS-84), also describe other parameters such as the rotation rate of the earth and various physical constants such as the angular velocity of the earth and the earth's gravitational constant.

Reference Ellipsoids

Reference ellipsoids are defined by either *semi-major* (equatorial radius) and *semi-minor* (polar radius) axes, or the relationship between the semi-major axis and the flattening of the ellipsoid (expressed as its *eccentricity*). Many reference ellipsoids are in use by different nations and agencies. Reference ellipsoids are identified by a name and often by a year for example, the Clarke 1866 ellipsoid is different from the Clarke 1858 and the Clarke 1880 ellipsoids.

Geodetic Datums

Precise positioning must also account for irregularities in the earth's surface due to factors in addition to polar flattening. Topographic and sea-level models attempt to model the physical variations of the surface:

- The *topographic surface* of the earth is the actual surface of the land and sea at some moment in time. Aircraft navigators have a special interest in maintaining a positive height vector above this surface.
- *Sea level* can be thought of as the average surface of the oceans, though its true definition is far more complex. Specific methods for determining sea level and the temporal spans used in these calculations vary considerably. Tidal forces and gravity differences from location to location cause even this smoothed surface to vary over the globe by hundreds of meters.

Gravity models and ***geoids*** are used to represent local variations in gravity that change the local definition of a level surface. *Gravity models* attempt to describe in detail the variations in the gravity field. The importance of this effort is related to the idea of *levelling.* Plane and geodetic surveying uses the idea of a plane perpendicular to the gravity surface of the earth which is the direction perpendicular to a plumb bob pointing toward the center of mass of the earth. Local variations in gravity, caused by variations in the earth's core and surface

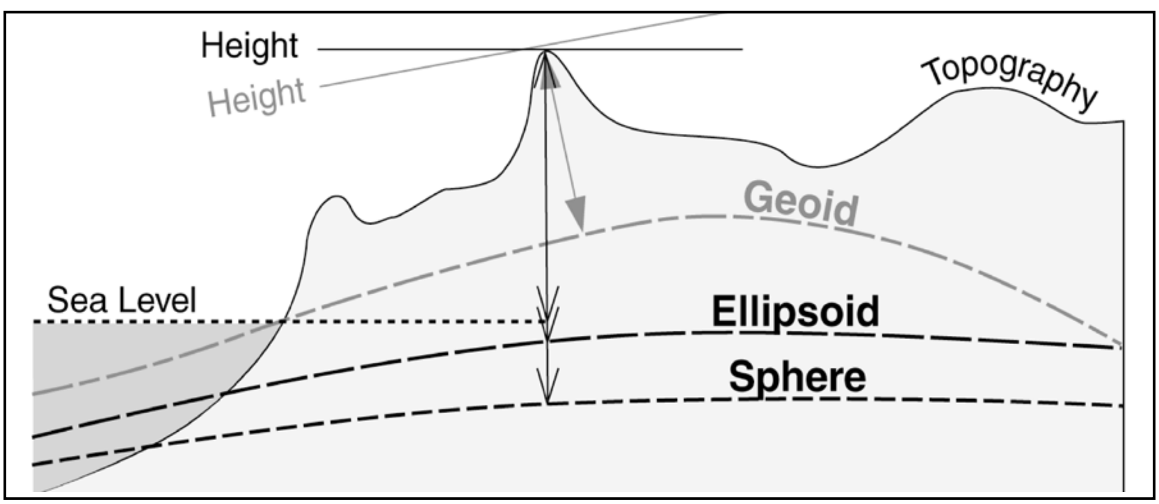


Figure 3.4: Elevations defined with reference to a sphere, ellipsoid, geoid, or local sea level will all be different. Even location as latitude and longitude will vary somewhat.

materials, cause this gravity surface to be irregular. *Geoid model*s attempt to represent the surface of the entire earth over both land and ocean as though the surface resulted from gravity alone.

Geodetic datums define reference systems that describe the size and shape of the earth based on these various models. While cartography, surveying, navigation, and astronomy all make use of geodetic datums, they are the central concern of the science of ***geodesy.*** Hundreds of different datums have been used to frame position descriptions since the first estimates of the earth's size were made by the ancient Greeks. Datums have evolved from those describing a spherical earth to ellipsoidal models derived from years of satellite measurements. Modern geodetic datums range from *flat-earth* models, used for plane surveying to *complex* models, used for international applications, which completely describe the size, shape, orientation, gravity field, and angular velocity of the earth.

Different nations and international agencies use different datums as the basis for coordinate systems in geographic information systems, precise positioning systems, and navigation systems. Linking geodetic coordinates to the wrong datum can result in position errors of hundreds of meters. The diversity of datums in use today and the technological advancements that have made possible global positioning measurements with sub-meter accuracies requires careful datum selection and careful conversion between coordinates in different datums. For the purposes of this unit, reference system can be divided into two groups:

- *Global systems* can refer to positions over much of the Earth.
- *Regional systems* have been defined for many specific areas, often covering national, state, or provincial areas.

General Coordinate Systems

Coordinates are used to identify locations on the earth's surface. Locations may be relative to the earth's surface, the image or map display. Choice depends on size of area of interest. Standardized coordinate systems use absolute locations. To compare or edge-match maps in a GIS, both maps MUST be in the same coordinate system. They are based on measurements of displacement from a given location. They are of two types:

- Plane
- Global

PLANE COORDINATE SYSTEM – CARTESIAN COORDINATES: Cartesian coordinates are determined by locating an origin there after setting two axes through origin in fixed directions, at right angles to each other. By convention these are usually identified as x and y, where x is horizontal and y vertical (x is east, y is north). To measure linear displacement from the origin in directions defined by the two axes produces an ordered (x, y) pairs.

Table 3.4: Selected reference ellipsoids.

Ellipse	Semi-major axis	Flattening
Airy 1830	6377563.396	299.3249646
Bessel 1841	6377397.155	299.1528128
Clarke 1866	6378206.4	294.9786982
Clarke 1880	6378249.145	293.465
Everest 1830	6377276.345	300.8017
Fischer 1960 (Mercury)	6378166	298.3
Fischer 1968	6378150	298.3
G R S 1967	6378160	298.247167427
G R S 1975	6378140	298.257
G R S 1980	6378137	298.257222101
International	6378388	297.0
Krassovsky 1940	6378245	298.3
WGS 60	6378165	298.3
WGS 66	6378145	298.25
WGS 72	6378135	298.26
WGS 84	6378137	298.257223563

STORING COORDINATES: In a GIS, coordinates must be stored in the computer as numbers, there are two important concepts that need to be considered:

1. *Integer vs real numbers:* Integers are whole numbers, optionally preceded by ' – ' to indicate negation. They are discrete since mathematically there is a distance of 1 between consecutive numbers. Real numbers can be expressed as decimal numbers and are continuous. Real numbers are often expressed as floating point numbers, usually expressed as two sets of digits (a,b). Here, the first set gives the significant digits and the second set gives the exponent, which determines the position of the decimal place. The number is the product (a × 10b), *e.g.,* +1234 +2 would indicate 0.1234 × 10^2 or 12.34.
2. *Computer precision:* In the computer, the number of digits which can be stored for each value is limited by the hardware, integers are normally stored using 16 bits of memory and can have a range from – 32767 to +32767. Floating point numbers can use single or double precision. Single and double precision are used to refer to the number of digits that can be stored for a single value. Single precision commonly allocates 32 bits, or 4 bytes, of memory for each value, equivalent to 7 significant decimal digits. Actual numeric ranges vary between computer implementations, in QuickBasic the exponent range is – 45 to +38.

Double precision commonly allocates 64 bits or 8 bytes, equivalent to 15 or 16 significant decimal digits and, in QuickBasic, an exponent range from – 324 to +308. Questions of

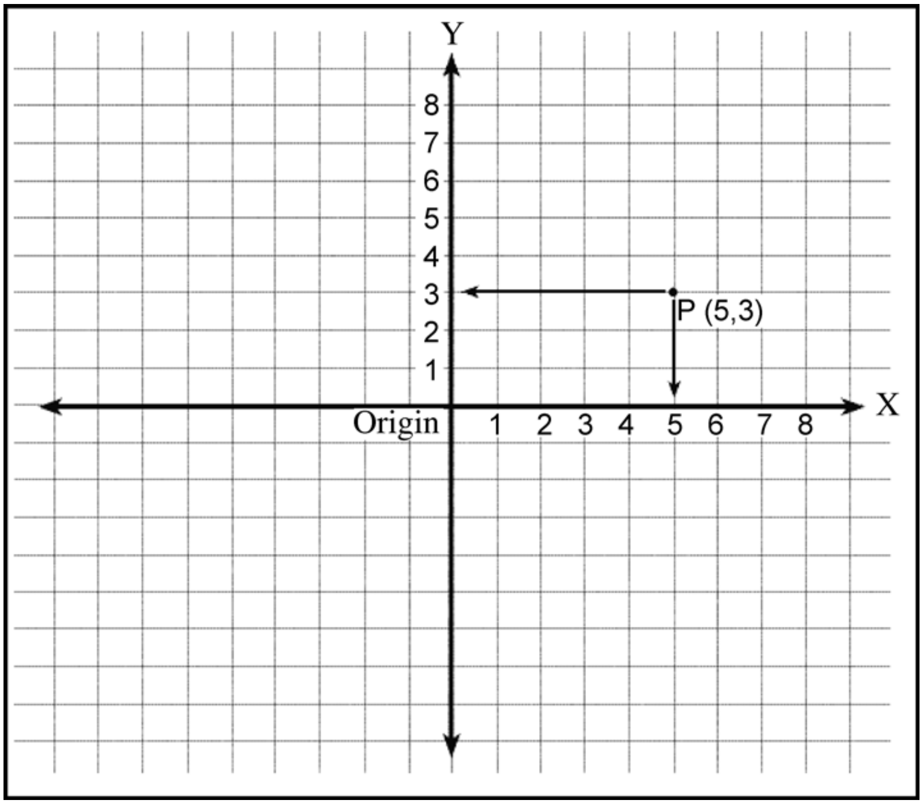


Figure 3.5: The Cartesian coordinate system.

precision are important when doing calculations since extra digits produced by division and multiplication operations may exceed the precision capacity of our system.

PRECISION OF CARTESIAN COORDINATES: The number of significant digits required for a specific project when using Cartesian coordinates depends on two measures:

- Size of the study area
- Resolution (accuracy) of measurement

For example, if the study area is 10 km across and the resolution of measurement is 10 cm, this would create a range of values from 0 to 10^5 and requires 5 significant decimal digits or approximately 15 binary digits. It can calculate approximate number of binary digits by multiplying number of decimal digits by 3 (log210). Since the computer system usually offers more resolution than needed by the data. The data is stored at higher precision than is justified by its accuracy. GIS designers are reluctant to throw away extra significant digits because designers may not be aware of the resolution of the data that will be used. Coordinate systems based on a global scale where the size of the area is 10,000 km and the resolution is 1 mm would need 10 decimal digits or 30 binary digits, this will require double precision coordinates, which few GIS systems offer.

PLANE COORDINATE SYSTEM – POLAR COORDINATES: Polar coordinates use distance from origin (r) and angle from fixed direction (q), usually fixed direction is north and angle is measured clockwise from it. Polar coordinates are useful for measuring from some fixed point such as the center of the city or when using data from sources such as ground surveys and radar.

To translate from (r, q) to (x,y)

$x = r \sin(q)$ $y = r \cos(q)$

$r = (x^2 + y^2)$ $q = \arctan(x/y)$

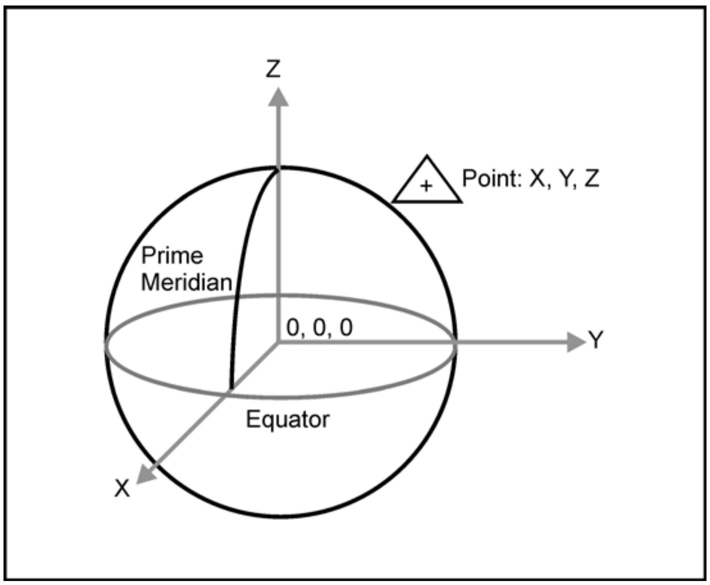


Figure 3.6: Earth Centered, Earth Fixed (ECEF) Cartesian coordinates can also be used to define three dimensional positions.

EARTH COORDINATE GEOMETRY

The earth's spherical shape is more difficult to describe than a plane surface. Concepts from Cartesian coordinate geometry have been incorporated into the earth's coordinate system.

ROTATION OF THE EARTH: The spinning of the earth on its imaginary axis is called *rotation.* Aside from the cultural influences of rotation, this spinning also has a physical influence. The spinning has led to the creation of a system to determine points and directions on the sphere. The North and South poles represent the axis of spin and are fixed reference points. If the North Pole was extended, it would point to a fixed star, the North Star (Polaris). Any point on the earth's surface moves with the rotation and traces imaginary curved lines are *Parallel of Latitude.*

THE EQUATOR: If a plane bisected the earth midway between the axis of rotation and perpendicular to it, the intersection with the surface would form a circle. This unique circle is the *equator.* The equator is a fundamental reference line for measuring the position of points around the globe. The equator and the poles are the most important parts of the earth's coordinate system.

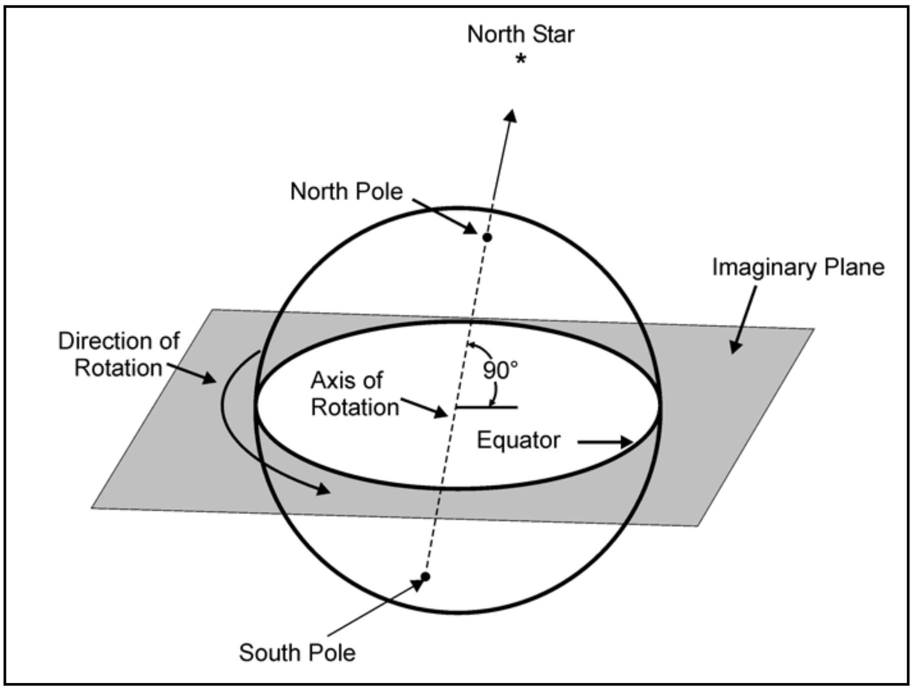


Figure 3.7: Location of the equator, north and south poles, and the imaginary axis of rotation.

THE GEOGRAPHIC GRID: The spherical coordinate system with latitudes and longitudes used for determining the locations of surface features.

- Parallels: east-west lines parallel to the equator.
- Meridians: north-south lines connecting the poles.
- Parallels are constantly parallel, and meridians converge at the poles.
- Meridians and parallels always intersect at right angles.

PARALLELS OF LATITUDE: *Parallels of latitude* are all small circles, except for the equator. They are true east-west lines, always parallel, any two are always equal distances apart and an infinite number can be created. Parallels are related to the horizontal x-axes of the Cartesian coordinate system.

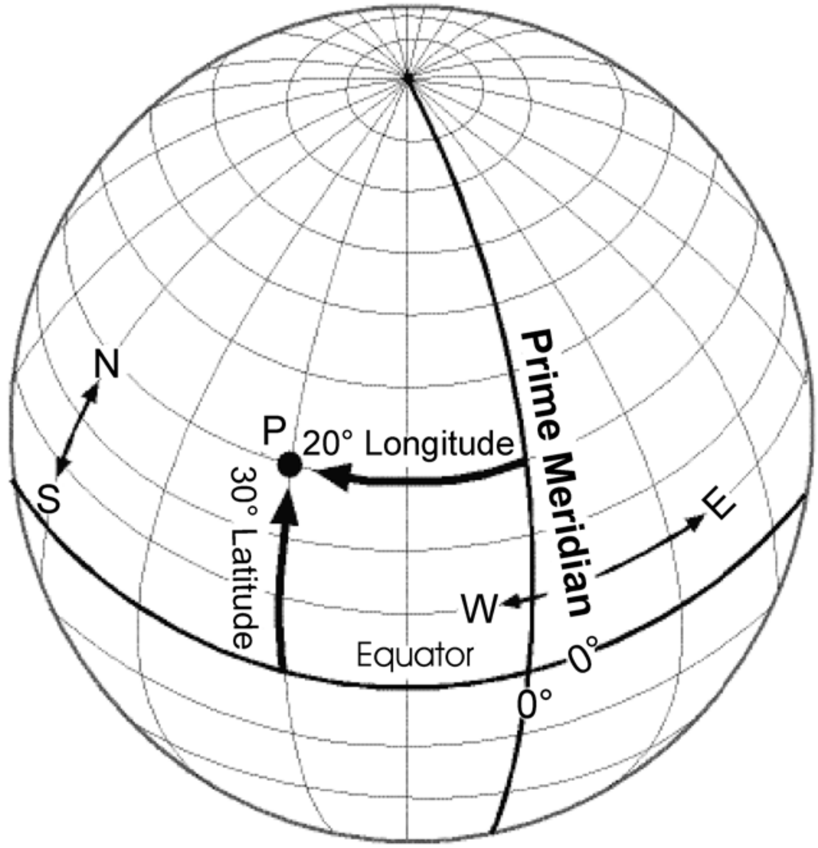


Figure 3.8: The Geographic grid.

The Point P has a latitude of 30 degrees North and a longitude of 20 degrees West.

MERIDIANS OF LONGITUDE: *Meridians of longitude* are halves of great circles, connecting one pole to the other. All run in a true north-south direction, spaced farthest apart at the equator and converge to a point at the poles, an infinite number can be created on a globe. Meridians are similar to the vertical y-axes of the Cartesian coordinate system.

DEGREES, MINUTES, AND SECONDS: Angular measurement is used in addition to simple plane geometry to specify location on the earth's surface. This is based on a *sexagesimal scale:* A circle has 360 degrees, 60 minutes per degree, and 60 seconds per minute. There are 3,600 seconds per degree. For example, 45° 33′ 22″ (45 degrees, 33 minutes, 22 seconds).

It is often necessary to convert this conventional angular measurement into decimal degrees. To convert 45° 33′ 22″, first multiply 33 minutes by 60, which equals 1,980 seconds. Next add 22 seconds to 1,980: 2,002 total seconds. Now the ratio: 2,002/3,600 = 0.55. Adding this to 45 degrees, the answer is 45.55°.

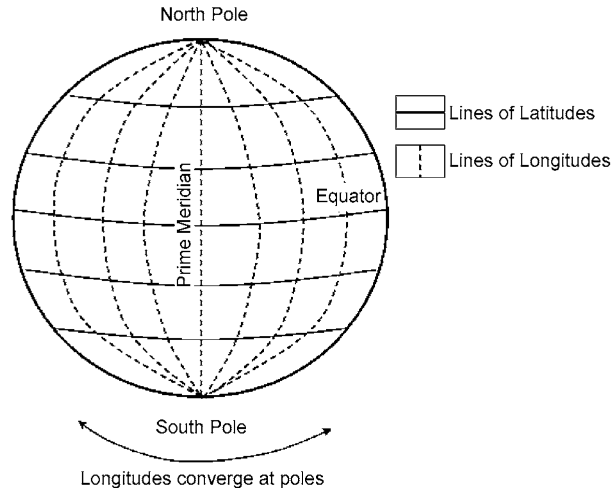


Figure 3.9: Parallels of latitude and Meridians of longitude.

The earth rotates on its axis once every 24 hours, therefore, any point moves through 360° a day, or 15° per hour.

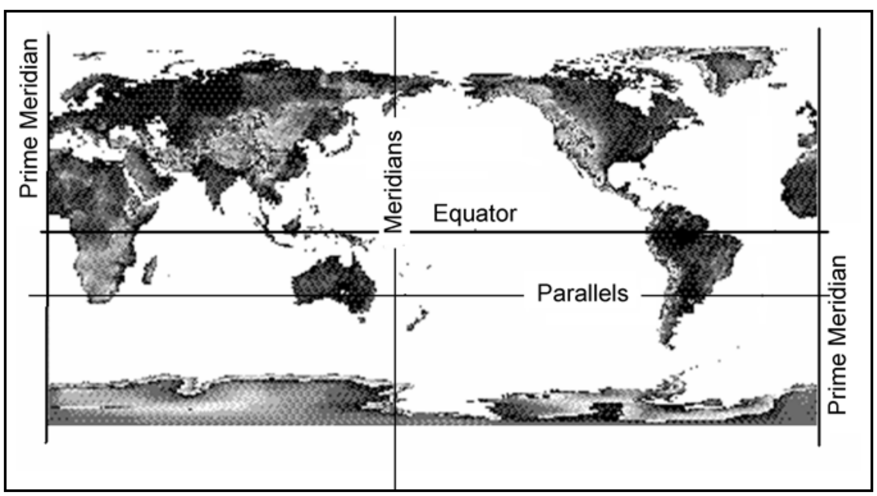


Figure 3.10: Geographic coordinates.

GREAT AND SMALL CIRCLES: A *great circle* is a circle formed by passing a plane though the exact center of a sphere. It is the largest circle that can be drawn on a sphere's surface. An infinite number of great circles can be drawn on a sphere. Great circles are used in the calculation of distance between two points on a sphere. A *small circle* is produced by passing a plane through any part of the sphere other than the center.

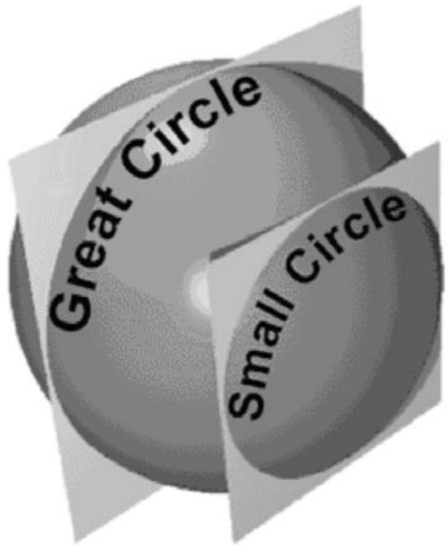


Figure 3.11: Great and small circles.

Latitude and Longitude and Locations

Latitude

- ***Authalic Latitude*** is based on a spherical earth. It measures the position of a point on the earth's surface in terms of the angular distance between the equator and the poles. It indicates how far north or south of the equator a particular point is situated. *North latitude:* all points north of the equator in the northern hemisphere. *South latitude:* all points south of the equator in the southern hemisphere. Latitude is measured in angular degrees from 0° at the equator to 90° at either of the poles. A point in the northern hemisphere 28 degrees north of the equator is labelled Lat. 28° N. The north or south measurement of latitude is actually measured along the meridian which passes through that location. It is known as an *arc* of the meridian.
- ***Geodetic Latitude*** is based on an ellipsoidal earth. The ellipsoid is a more accurate representation of the earth than a sphere since it accounts for polar flattening. Modern large-scale mapping, GIS, and GPS technology all require the higher accuracy of an ellipsoidal reference surface. When the earth's shape is based on the WGS 84 Ellipsoid, the length of 1° of latitude is not the same everywhere as it is on the sphere.

  At the equator, 1° of latitude is 110.57 kilometers (68.7 miles).

  At the poles, 1° of latitude is 111.69 kilometers (69.4 miles).

LATITUDE AND DISTANCE: Parallels of latitude decrease in length with increasing latitude. Length of parallel at latitude x = (cosine of x) * (length of equator). The length of each degree is obtained by dividing the length of that parallel by 360°. For example, the cosine of 60° is 0.5, so the length of the parallel at that latitude is one half the length of the equator. Since the variation in lengths of degrees of latitude varies by only 1.13 kilometers (0.7 mile), the standard figure of 111.325 kilometers (69.172 miles) can be used. For example, anywhere on the earth, the length represented by 3° of latitude is

$$(3 \times 111.325) = 333.975 \text{ kilometers.}$$

Longitude

Longitude measures the position of a point on the earth's surface east or west from a specific meridian, the *prime meridian.* The longitude of a place is the arc, measured in degrees along a parallel of latitude from the prime meridian. The most widely accepted prime meridian is based on the *Bureau International de l'Heure (BIH) Zero Meridian.* It passes through the old Royal Observatory in Greenwich, England. The prime meridian has the angular designation of 0° longitude. All other points are measured with respect to their position east or west of this meridian. Longitude ranges from 0° to 180°, either east or west. For the purposes of measurement, no one prime meridian is better than another. Having a widely accepted meridian allows comparison between maps published in different areas. The distance represented by a degree of longitude varies upon where it is measured. The length of a degree of longitude along a meridian is not constant because of polar flattening. At the equator, the approximate length is determined by dividing the earth's circumference (24,900 miles) by 360 degrees *i.e.,* 111.05 kilometers (69 miles). The meridians converge at the poles, and the distance represented by one degree decreases. At 60° N latitude, one degree of longitude is equal to about 55.52 kilometers (34.5 miles).

Longitude and Distance

The earth is not a perfect sphere, thus, the equatorial circumference does not equal that of the meridians. On a perfect sphere, each meridian of longitude equals one-half the circumference of the sphere. The length of each degree is equal to the circumference divided by 360°. Each degree is equal to every other degree. Measurement along meridians of longitude accounts for the earth's polar flattening and degree lengths along meridians are not constant. For example, 111.325 kilometers (69.172 miles) per degree at the equator, while 16.85 kilometers (10.47 miles) per degree at 80° North and 0 kilometers at the poles. The distance between meridians of longitude on a sphere is a function of latitude. The *Mathematical expression* is: Length of a degree of longitude = cos (latitude) × 111.325 kilometers. For example, 1° of longitude at 40° N = cos (40°) × 111.325. Since the cosine of 40° is 0.7660, the length of one degree is 85.28 kilometers.

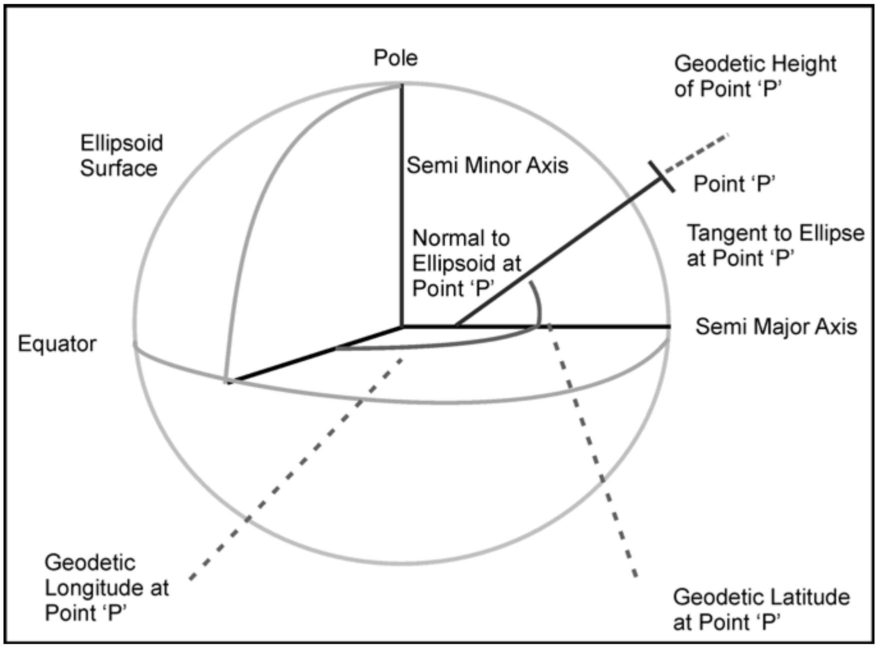


Figure 3.12: Geodetic latitude, longitude, and height.

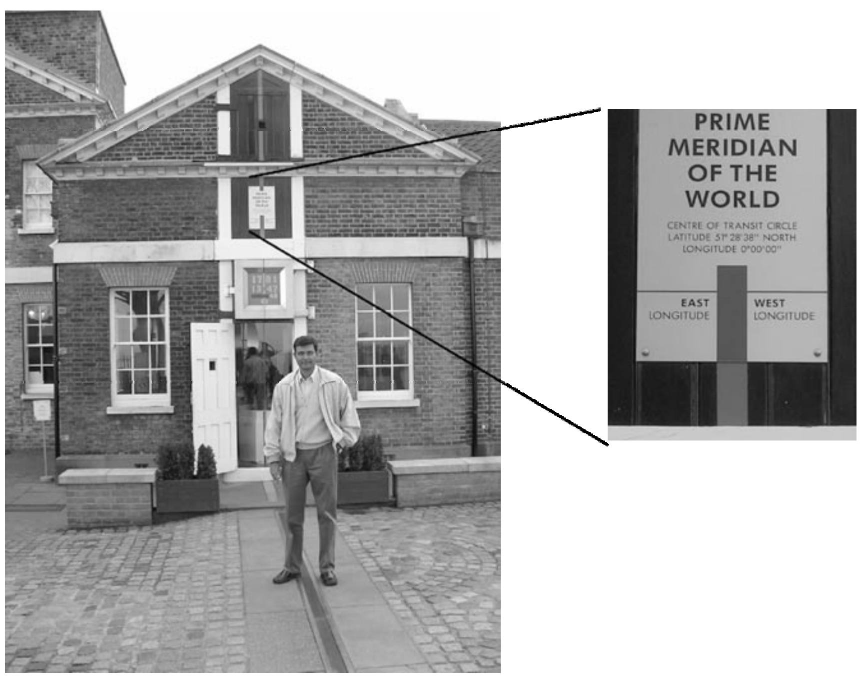


Figure 3.13: The author at Royal Observatory in Greenwich, England.

Table 3.5: Length of a degree of geodetic latitude and geodetic longitude.

Latitude (º)	Length of a Degree of Geodetic Latitude		Length of a Degree of Geodetic Longitude	
	Miles	Kilometers	Miles	Kilometers
0°	68.71	110.57	69.17	111.32
10°	68.73	110.61	68.13	109.64
20°	68.79	110.70	65.03	104.65
30°	68.88	110.85	59.95	96.49
40°	68.99	111.04	53.06	85.39
50°	69.12	111.23	44.55	71.70
60°	69.23	111.41	34.67	55.80
70°	69.32	111.56	23.73	38.19
80°	69.38	111.66	12.05	19.39
90°	69.40	111.69	0.00	0.00

Earth-Based Locational Reference Systems

Reference systems and map projections extend the ideas of Cartesian and polar coordinate systems over all or part of the earth. Map projections portray the nearly spherical earth in a two-dimensional representation. Earth-based reference systems are based on various *models* for the size and shape of the earth. Earth shapes are represented in many systems by a *sphere.* However, precise positioning reference systems are based on an *ellipsoidal earth* and *complex gravity models*.

Map Projections

A map projection is a system in which locations on the curved surface of the earth are displayed on a flat sheet or surface according to some set of rules. Mathematically, projection is a process of transforming global location to a planar position.

MAP PROJECTIONS AND GIS: Maps are a common source of input data for a GIS. Generally input maps collected from different sources are in different projections, requiring transformation of one or all maps to make coordinates compatible, thus, mathematical functions of projections are needed in a GIS. Often GIS are used for projects of global or regional scales so consideration of the effect of the earth's curvature is necessary. Monitor screens are analogous to a flat sheet of paper; thus, need to provide transformations from the curved surface to the plane for displaying data. Angles, areas, directions, shapes and distances become distorted when transformed from a curved surface to a plane. All these properties cannot be kept undistorted in a single projection. Usually the distortion in one property will be kept to a minimum while other properties become much distorted.

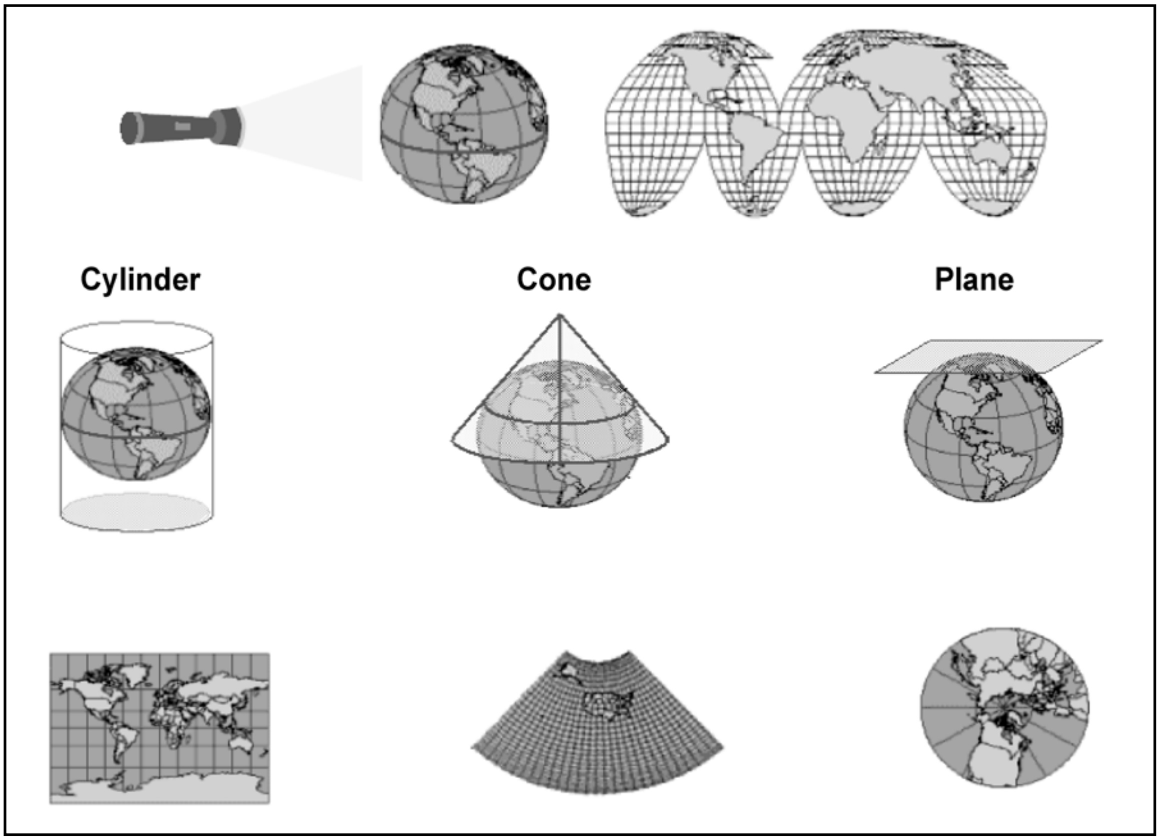


Figure 3.14: Map projections convert curved surface of the earth into a flat surface.

Tissot's Indicatrix: This is a convenient way of showing distortion. If a tiny circle drawn on the surface of the globe, on the distorted map the circle will become an ellipse, squashed or stretched by the projection. The size and shape of the Indicatrix will vary from one part of the map to another, the Indicatrix is used to display the distorting effects of projections.

Figure of the Earth: The figure of the earth is a geometrical model used to generate projections; a compromise between the desire for mathematical simplicity and the need for accurate approximation of the earth's shape. The common types are:

a. *Plane:* It assume the earth is flat (use no projection) and used for maps only intended to depict general relationships or for maps of small areas. At scales larger than 1:10,000 planar representations has little effect on accuracy. Planar projections are usually assumed when working with air photos.
b. *Sphere:* It assumes the earth is perfectly spherical thus does not truly represent the earth's shape.

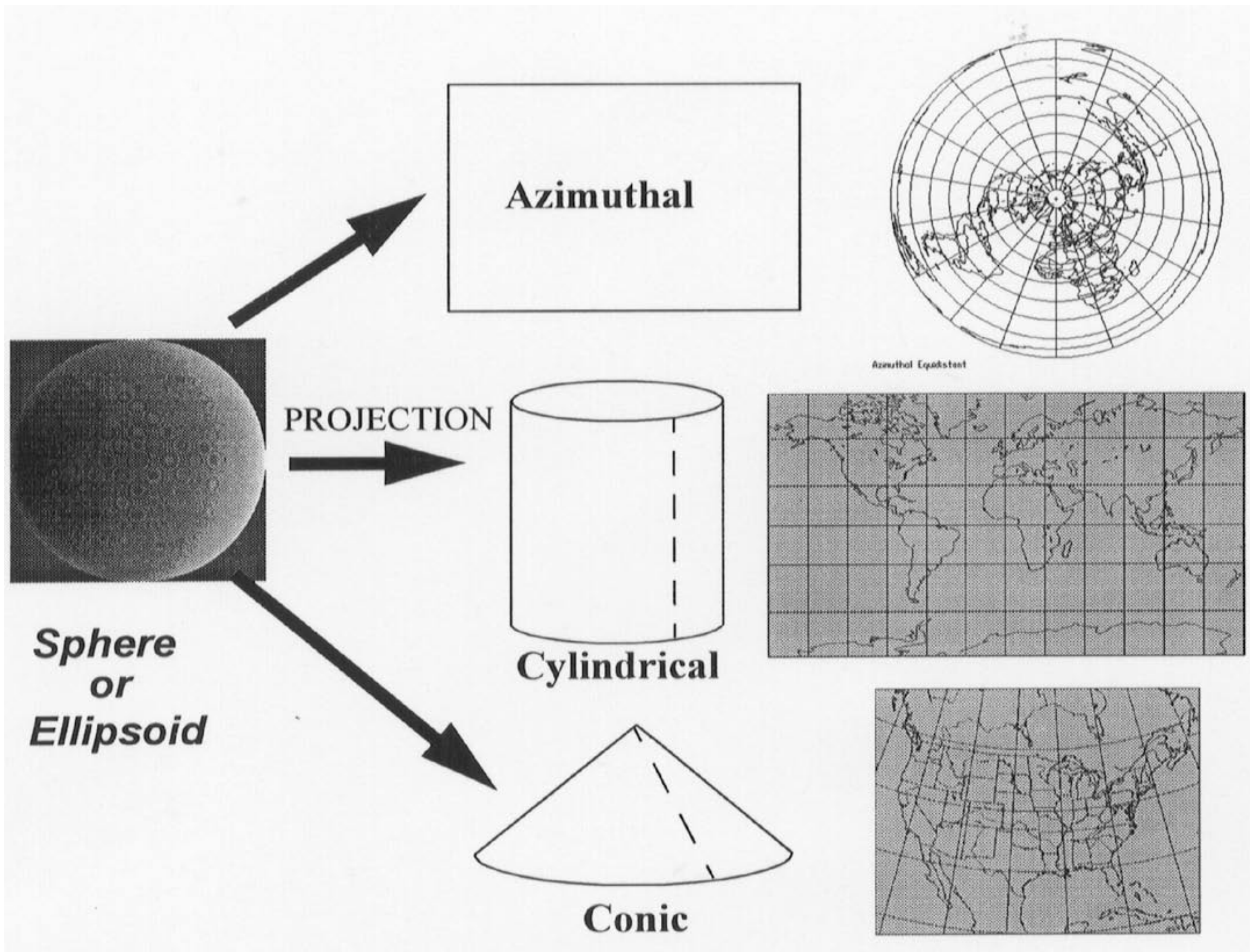


Figure 3.15: The earth can be projected in many ways, but basically into three shapes that can be unrolled into a flat map. A flat plane, a cylinder and a cone.

ELLIPSOID: This is the figure created by rotating an ellipse about its minor axis. The ellipsoid models the fact that the earth's diameter at the equator is greater than the distance between poles, by about 0.3 %. At global scales, the difference between the sphere and ellipsoid are small, about equal to the topographic variation on the earth's surface. A line width of 0.5 mm the earth would have to be drawn with a radius of 15 cm before the two models would deviate. The difference is unlikely to affect mapping of the globe at scales smaller than 1:10,000,000.

The ellipsoid is still an approximation to the actual shape, the earth is actually slightly pear shaped, slightly larger in the southern hemisphere, and has other smaller bulges. Therefore, different ellipsoids are used in different regions, each chosen to fit the observed datum of each region. Accurate conversion between latitude and longitude and projected coordinates requires knowledge of the specific figures of the earth that have been used. The actual shape of the earth can be determined quite accurately by observing satellite orbits. Satellite systems, such as GPS, can determine latitude and longitude at any point on the earth's surface to accuracies of fractions of a second, thus, it is now possible to observe otherwise unapparent errors introduced by the use of an approximate figure for map projections.

PLANAR OR AZIMUTHAL PROJECTIONS: A flat sheet is placed in contact with a globe, and points are projected from the globe to the sheet. Mathematically, the projection is easily expressed as mappings from latitude and longitude to polar coordinates with the origin located at the point of contact with the paper. The examples are:

- Stereographic projection
- Gnomic projection
- Lambert's azimuthal equal-area projection
- Orthographic projection

CONIC PROJECTIONS: The transformation is made to the surface of a cone tangent at a small circle (tangent case) or intersecting at two small circles (secant case) on a globe. Mathematically, this projection is also expressed as mappings from latitude and longitude to polar coordinates, but with the origin located at the apex of the cone. The examples are:

- Alber's conical equal area projection with two standard parallels
- Lambert conformal conic projection with two standard parallels
- Equidistant conic projection with one standard parallel

CYLINDRICAL PROJECTIONS: These projections are developed by transforming the spherical surface to a tangent or secant cylinder. Mathematically, a cylinder wrapped around the equator is expressed with x equal to longitude, and the y coordinates some function of latitude. The Example is Mercator projection.

NON-GEOMETRIC PROJECTIONS: Some projections cannot be expressed geometrically, they have only mathematical descriptions. The examples are Molleweide and Eckert etc.

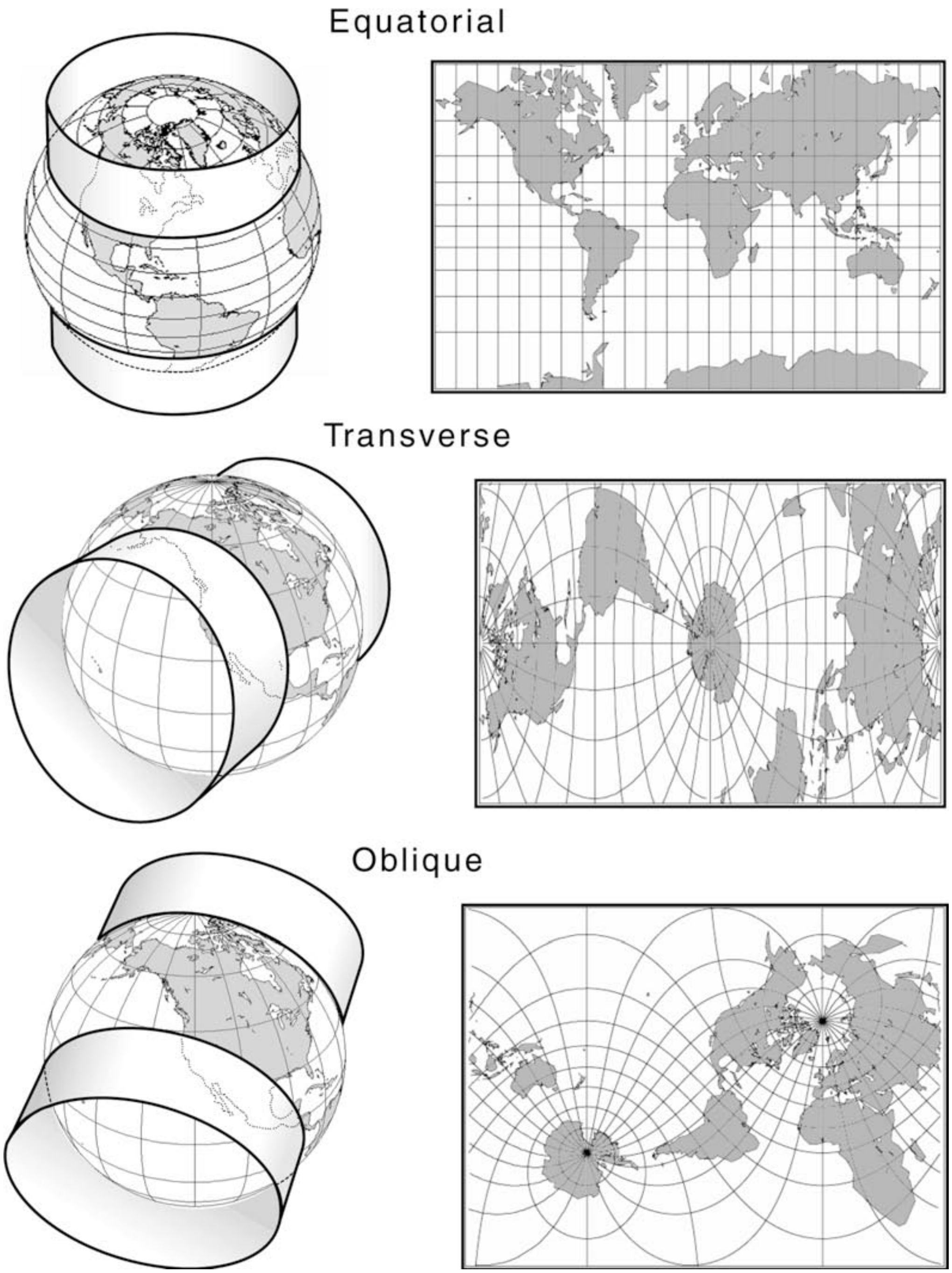


Figure 3.16: Variations on the Mercator projection shown as **Secant**.

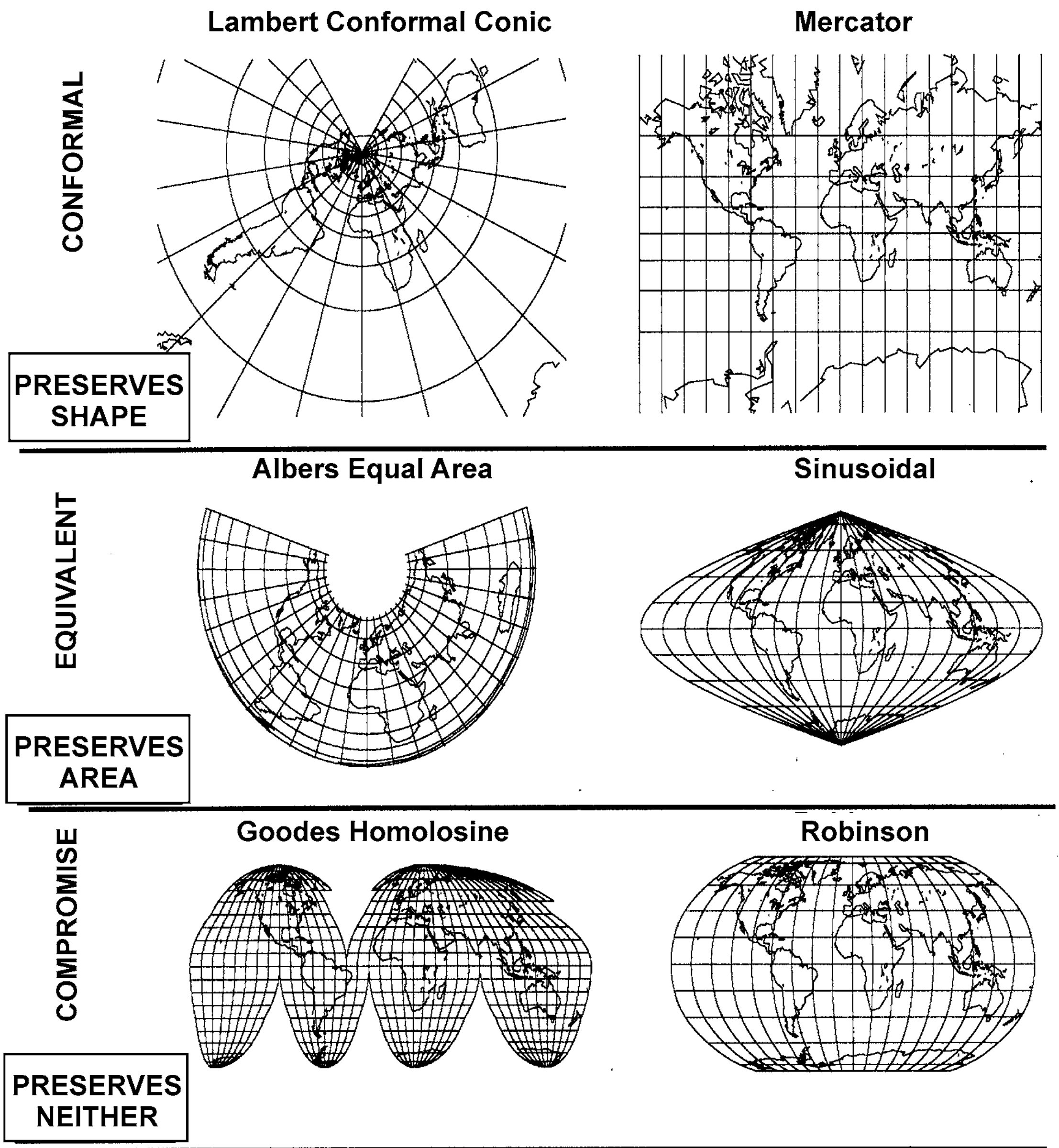


Figure 3.17: Examples of projections classified by their distortions.

Geometric Analogy

The most common methods of projection can be conceptually described by imagining the developable surface, which is a surface that can be made flat by cutting it along certain lines and unfolding or unrolling it. The points or lines where a developable surface touches the globe in projecting from the globe are called standard points and lines, or points and lines of zero distortion. At these points and lines, the scale is constant and equal to that of the globe, no linear distortion is present.

If the developable surface touches the globe, the projection is called **tangent** and if the surface cuts into the globe, it is called **secant**. Where the surface and the globe intersect, there is no distortion while where the surface is outside the globe, objects appear bigger than in reality-scales are greater than 1 and where the surface is inside the globe, objects appear smaller than in reality and scales are less than 1.

Conformal (Orthomorphic) Projections: A projection is conformal if the angles in the original features are preserved, over small areas the shapes of objects will be preserved. Preservation of shape does not hold with large regions (*i.e.,* Greenland in Mercator projection). A line drawn with constant orientation (*e.g.,* with respect to north) will be straight on a conformal projection, is termed a rhumb line or loxodrome. Parallels and meridians cross each other at right angles (note: not all projections with this appearance are conformal). The Tissot Indicatrix is a circle everywhere, but its size varies. Conformal projections cannot have equal area properties, so some areas are enlarged, generally, areas near margins have a larger scale than areas near the center.

Equal Area (Equivalent) Projections: The representation of areas is preserved so that all regions on the projection will be represented in correct relative size. Equal area maps cannot be conformal, so most earth angles are deformed and shapes are strongly distorted. The Indicatrix has the same area everywhere, but is always elliptical, never a circle (except at the standard parallel).

Equidistant Projections: We cannot make a single projection over which all distances are maintained. Thus, equidistant projections maintain relative distances from one or two points only, *i.e.,* in a conic projection all distances from the center are represented at the same scale.

Universal Transverse Mercator (UTM)

UTM provides georeferencing at high levels of precision for the entire globe. Established in 1936 by the International Union of Geodesy and Geophysics, it is adopted by many national and international mapping agencies. It is commonly used in topographic and thematic mapping, for referencing satellite imagery and as a basis for widely distributed spatial databases. Universal Transverse Mercator (UTM) coordinates define two dimensional, horizontal, positions. Each UTM zone is identified by a number. UTM zone *numbers* designate individual 6° wide longitudinal strips extending from 80° South latitude to 84° North latitude as distortions at the poles is too large. Each zone has a *central meridian*. For

Table 3.6: Common map projections: Their properties and their application areas.

Projection	Properties	Application
Albers equal area	Equal area: conformal along standard parallel	Small regional and national maps
Azimuth equidistant	Equidistant: true direction from centre	Air and sea navigation, large scale maps in the equatorial areas.
Lambert conformal conical	Conformal: locally true direction	Navigation, US – state plane system
Mercator	Conformal: true dirction	Navigation, world maps
Equidistant conical	Equidistant along standard parallel and central meridian	Mid latitude areas with east – west extent, atlas mapping for smaller countries
Polyconic-conical	Equidistant along each parallel and central meridian	Topographic maps, Survey of India maps, USGS
Sinusoidal-cylindrical	Equal area, true direction along central meridian and equator	World maps
Stereographic-planar	Conformal: true direction	Navigational maps
Transverse Mercator–cylindrical	Conformal: locally true direction	Topographic mapping for areas with north south extents

example, Zone 14 has a central meridian of 99° west longitude. The zone extends from 96° to 102° west longitude. Locations within a zone are measured in meters eastward from the central meridian and northward from the equator. However, eastings increase eastward from the central meridian which is given a *false* easting of 500 km so that only positive eastings are measured anywhere in the zone. Northings increase northward from the equator with the equator's value differing in each hemisphere. In the Northern Hemisphere, the Equator has a northing of 0, while for Southern Hemisphere locations, the Equator is given a false northing of 10,000 km.

COORDINATES: They are expressed in meters, eastings (x) are displacements eastward while northings (y) express displacement northward. The central meridian is given an easting of 500,000 m. The northing for the equator varies depending on hemisphere, when calculating coordinates for locations in the northern hemisphere, the equator has a northing of 0 m while in the southern hemisphere, the equator has a northing of 10,000,000 m.

DISTORTIONS: To reduce the distortion across the area covered by each zone, scale along the central meridian is reduced to 0.9996. This produces two parallel lines of zero distortion approximately 180 km away from the central meridian.

World Geographic Reference System (GEOREF)

The World Geographic Reference System is used for aircraft navigation. GEOREF is based

Table 3.7: UTM zones and their extents

Zone no.	Central meridian	Bounding meridians	Zone no.	Central meridian	Bounding meridians	Zone no.	Central meridian	Bounding meridians
1	177°W	180° - 174°W	21	57°W	60° - 54°W	41	63°E	60° - 66°E
2	171°W	174° - 16°W	22	51°W	54° - 48°W	42	69°E	66° - 72°E
3	165°W	168° - 162°W	23	45°W	48° - 42°W	43	75°E	72° - 78°E
4	159°W	162° - 156°W	24	39°W	42° - 36°W	44	81°E	78° - 84°E
5	153°W	156° - 150°W	25	33°W	36° - 30°W	45	87°E	84° - 90°E
6	147°W	150° - 144°W	26	27°W	30° - 24°W	46	93°E	90° - 96°E
7	141°W	144° - 138°W	27	21°W	24° - 18°W	47	99°E	96° - 102°E
8	135°W	138° - 132°W	28	15°W	18° - 12°W	48	105°E	102° - 108°E
9	129°W	132° - 126°W	29	09°W	12° - 06°W	49	111°E	108° - 114°E
10	123°W	126° - 120°W	30	03°W	06° - 00°W	50	117°E	114° - 120°E
11	117°W	120° - 114°W	31	03°E	00° - 06°E	51	123°E	120° - 126°E
12	111°W	114° - 108°W	32	09°E	06° - 12°E	52	129°E	126° - 132°E
13	105°W	108° - 102°W	33	15°E	12° - 18°E	53	135°E	132° - 138°E
14	99°W	102° - 96°W	34	21°E	18° - 24°E	54	141°E	138° - 144°E
15	93°W	96° - 90°W	35	27°E	24° - 30°E	55	147°E	144° - 150°E
16	87°W	90° - 84°W	36	33°E	30° - 36°E	56	153°E	150° - 156°E
17	81°W	84° - 78°W	37	39°E	36° - 42°E	57	159°E	156° - 162°E
18	75°W	78° - 72°W	38	45°E	42° - 48°E	58	165°E	162° - 166°E
19	69°W	72° - 66°W	39	51°E	48° - 54°E	59	171°E	166° - 172°E
20	63°W	66° - 60°W	40	57°E	54° - 60°E	60	177°E	172° - 180°E

on latitude and longitude. The globe is divided into twelve bands of latitude and twenty-four zones of longitude, each 15° in extent. These 15° areas are further divided into one degree units identified by 15 characters.

REGIONAL SYSTEMS: Several different systems are used regionally to identify geographic location. Some of these are true coordinate systems, such as those based on UTM and UPS systems. Others, such as the Public Land Survey systems are simply partition space. Many nations have defined grid systems based on Transverse Mercator coordinates that cover their territory.

The British National Grid (BNG)

The British National Grid (BNG) is based on the National Grid System of England, administered by the British Ordnance Survey. The BNG has been based on a Transverse Mercator projection since the 1920s. The modern BNG is based on the Ordnance Survey of Great Britain Datum 1936. The true origin of the system is at 49° north latitude and 2 degrees west longitude. The false origin is 400 km west and 100 km north. Scale factor at the

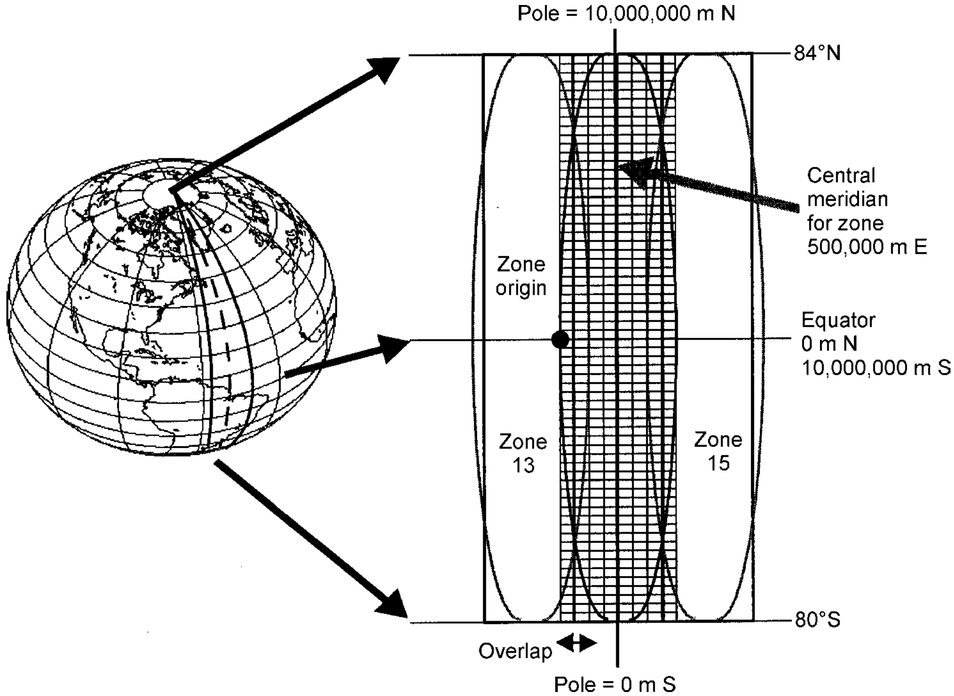


Figure 3.18: Overlap in UTM projection.

central meridian is 0.9996012717. The first BNG designator defines a 500 km square. The second designator defines a 100 km square. The remaining digits define 10 km, 1 km, 100 m, 10 m, and 1 m eastings and northings.

Indian Grid System

The Indian system follows almost the same as British system. The Indian system has eight grid zones named as 00, 0I, IIA, IIB, IIIA, IIIB, IVA, IVB based on Lamberts conical orthomorphic projection with two standard parallels covering India, Pakistan, Myanmar, Afghanistan, parts of Iran, China, Tibet and Thailand. Each zone has a belt of 8° latitude. The false origin for all the zones, except of zone 00 is 3000000 yards easting and 1000000 yards northing. The origin of grid 00 is 2355000 yards easting and 2590000 yards northing. The grid lines are drawn at 1000 yards apart, on 1 inch to 1 mile and larger, whereas on 1 inch to 4 miles and smaller, the grid lines are 10000 yards. The topographical maps in India are not based on lamberts projection but on polyconic projections, due to this the grid squares are not perfect squares.

State Plane Coordinates (SPC)

SPCs are individual coordinate systems adopted by U.S. state agencies. State plane systems were developed in order to provide local reference systems that were tied to a national datum. In the United States, the **State Plane System 1927** was developed in the 1930s and was based on the North American Datum 1927 (NAD-27). NAD-27 coordinates are in

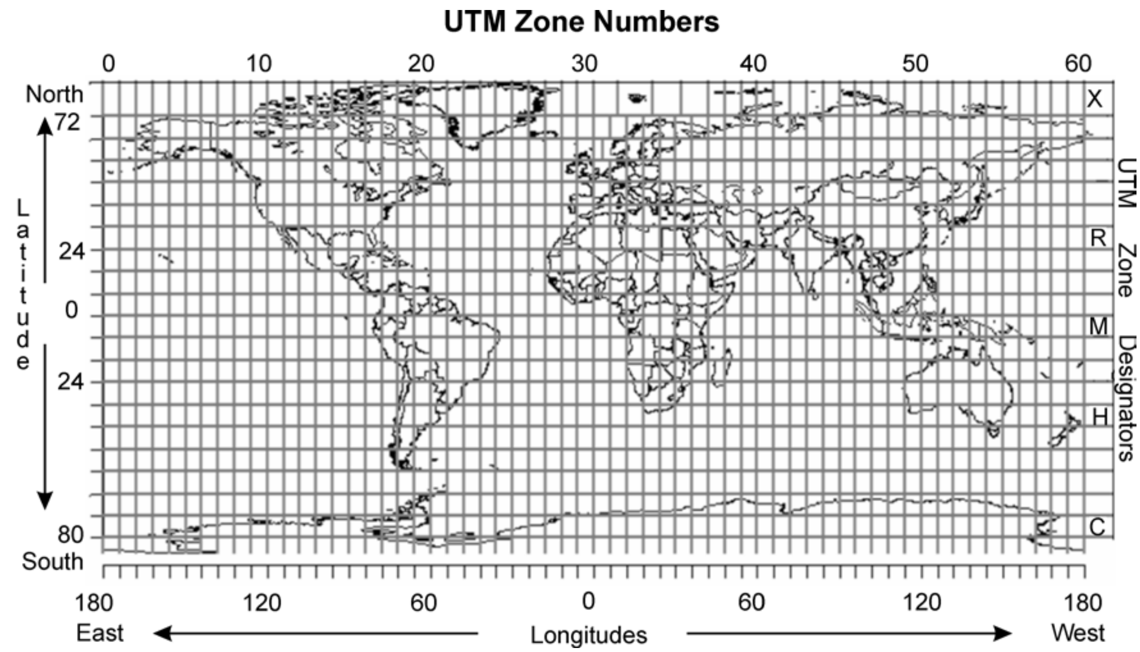


Figure 3.19: Universal transverse Mercator system.

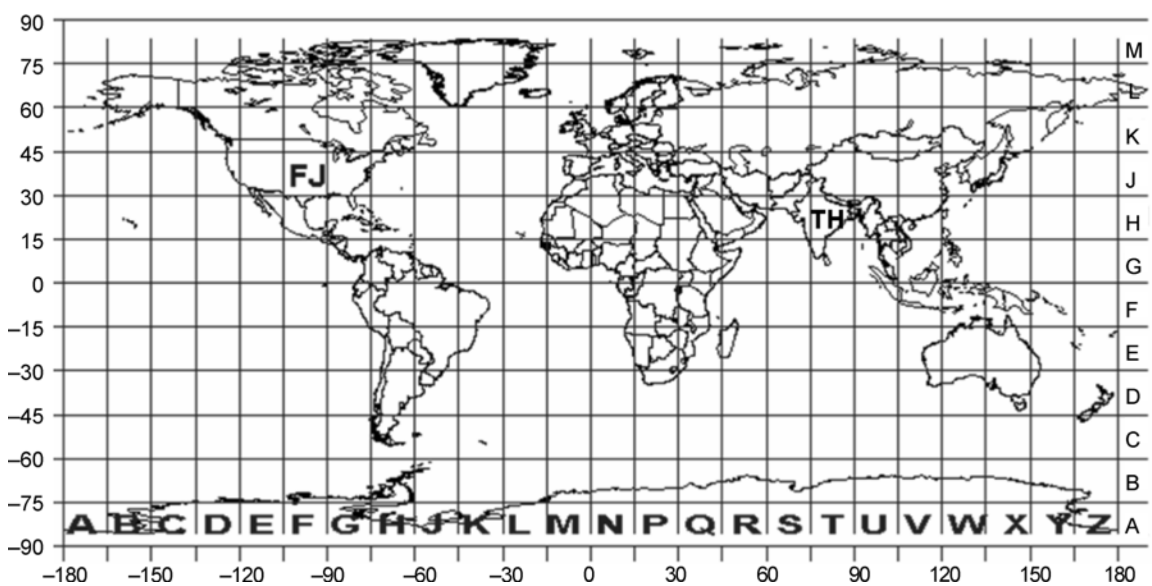


Figure 3.20: World geographic reference system (GEOREF).

English units (feet). The *State Plane System 1983* is based on the North American Datum 1983 (NAD-83). NAD-83 coordinates are metric. While the NAD-27 State Plane System has been superceded by the NAD-83 System, maps in NAD-27 coordinates are still in use. Each state's shape determines which projection is chosen to represent that state, *e.g.* a state extended N/S may use a Transverse Mercator projection while a state extended E/W may use a Lambert Conformal Conic projection (both of these are conformal). Projections are chosen to minimize distortion over the state and a state may have 2 or more overlapping zones, each with its own projection system and grid. The measuring units are generally in feet. The advantages of SPC coordinates are simpler than that of UTM and it gives a better representation than the UTM system for a state's area. However, SPC are not universal from state to state and problems arises at the boundaries of projections.

Georeferencing

Geographic location is the element that distinguishes spatial data with non spatial data. Methods for specifying location on the earth's surface for geographical data in a map is called as georeferencing. The primary requirements of a georeference are that it should be **unique,** so that there is only one location associated with a given georeference (*e.g.,* Hyderabad – one in India another in Pakistan). It should stay ***constant*** through time, because it could create confusion if it changes (*e.g.,* Madras – Chennai).

Box 5 : Commonly used systems of georeferencing

System	Domain of uniqueness	Metric / Non-metric	Example	Spatial resolution
Place name	Varies	Non metric	Hyderabad – India or Pakistan?	Varies
Postal address	Country	Non metric	11, Rose Apartments, Marris Road, Aligarh	Size of one mailbox
Postal code	Country	Non metric	202002(Aligarh, India) or WC1H OPF (London, U.K.)	Area occupied by a defined number of mailbox
Telephone code	Country	Non metric	011 (New Delhi, India)	Varies
Latitude/Longitude	Global	Metric	27°53′ North Latitude and 78°35′ East Longitude.	Infinitely fine
UTM	Zones of six degrees of longitude wide	Metric	1393267 & 3117373	Infinitely fine
State plane coordinates	USA only	Metric	55046.37 E & 75246.64 N	Infinitely fine

Data in a GIS must contain a geographic reference to a map, such as latitude and longitude. The GIS cross-references the attribute data with the map data.

Discrete Georeferencing

The georeferencing methods covered so far (latitude–longitude, Cartesian, projections from latitude/longitude to the plane) are continuous this means that there is no effective limit to precision, as coordinates are measured on continuous scales. The discrete methods – systems of georeferencing for discrete units on the earth's surface are indirect, this means that the method provides a key or index, which can then be used with a table to determine latitude/longitude or coordinates. For example, a Zip code is an indirect georeference, where instead assigning latitude/longitude for a place directly, it provides a unique number which can be looked up on a map if coordinates are needed. Since these methods are indirect, it is important to consider the precision of these systems. Precision is related directly to the size of the discrete unit which forms the basis of the georeferencing system.

STREET ADDRESS: This is a common discrete method of georeferencing, here the precision of street addresses as georeferences varies greatly. It is better for cities but poor for rural areas, where the address may indicate only that the place is somewhere in the area served by the post office. In GIS general approach is to match address to a list of streets (called address matching or 'addmatch'). Here, spelling and punctuation variations make this difficult *e.g.,* Ave. or Avenue, apartment number before or after street number

POSTAL CODE SYSTEM: Postal code systems have been set up in many countries, these often provide a high level of spatial precision. In India, zip codes are designed to assist with mail sorting and delivery. The codes are hierarchically nested, states are uniquely identified by one or more sets of the first 2 numbers. The 6 digit ZIP potentially provides a much higher level of spatial resolution, but problems exist with overlapping and fragmented boundaries.

US PUBLIC LAND SURVEY SYSTEM: PLSS is the basis for land surveys and legal land description over much of the US. Unlike the previous systems, it is designed to reference land parcels, because it is a comprehensive, systematic approach it is possible to use it as a georeference in GIS with ease. It is commonly used by agencies such as the Bureau of Land Management and the US Forest Service, and within the oil and gas industry.

Affine and Curvilinear Transformations

Coordinate transformations are required when we need to register different sets of coordinates for objects in the same area that may have come from maps of different (and sometimes unknown) projections. In this case we need to transform one or more sets of coordinates so that they are represented in the same coordinate system as other sets. There are two ways to look at coordinate transformations:

i. move objects on a fixed coordinate system so that the coordinates change
ii. hold the objects fixed and move the coordinate system, this is the more useful way to consider transformations for GIS purposes.

There are two major groups of transformations

I. **Affine transformations** are those which keep parallel lines parallel and they are a class of transformations which have 6 coefficients.

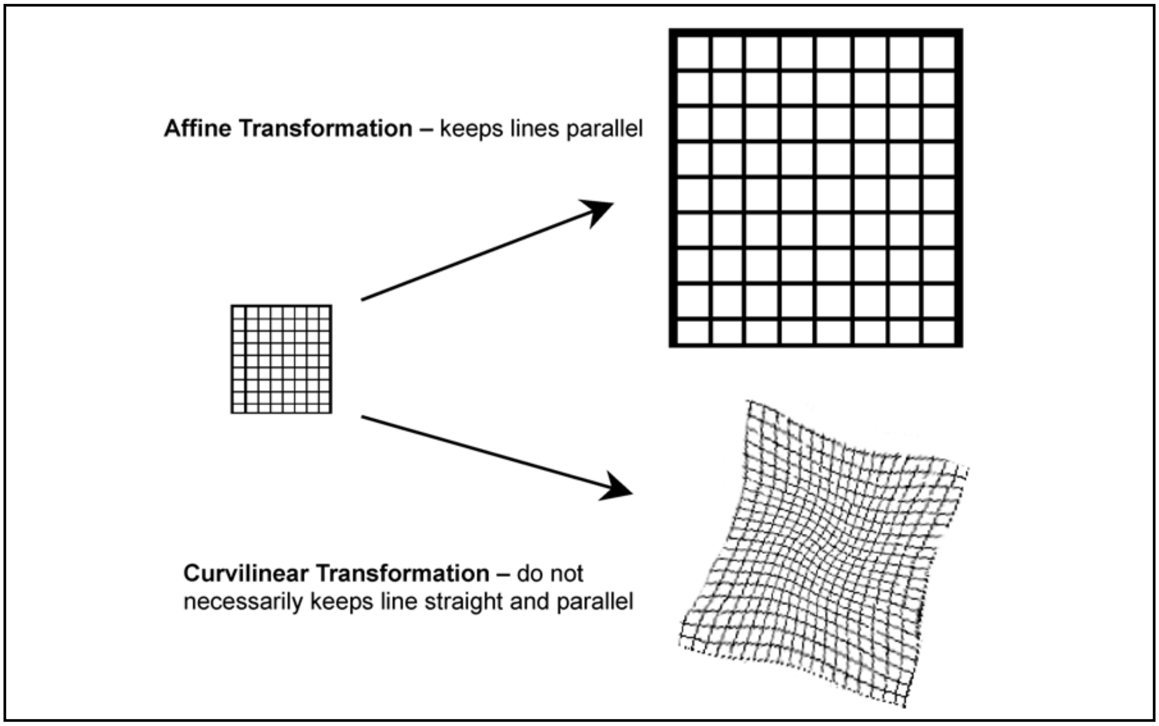


Fig. 3.21: Coordinate transformation: affine and curvilinear.

II. **Curvilinear transformations** are higher order transformations that do not necessarily keep lines straight and parallel and these transformations may require more than 6 coefficients.

Affine Transformation Primitives: affine transformations keep parallel lines parallel and there are four different types (primitives):

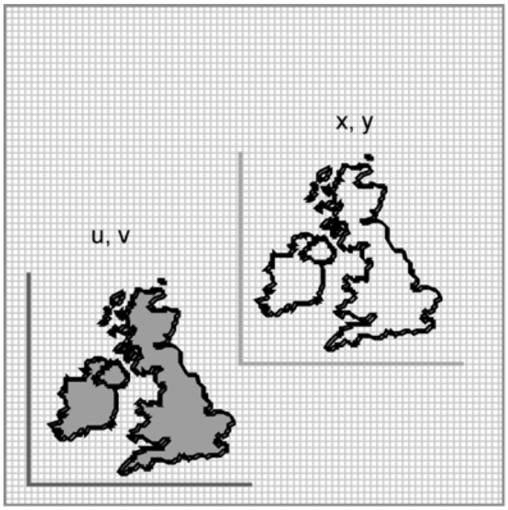
a. **Translation**-origin is moved, axes do not rotate
$u = x - a$ $v = y - b$
here, origin is moved a units parallel to x and b units parallel to y

b. **Scaling**-both origin and axes are fixed, scale changes
$u = cx$ $v = dy$
here, scaling of x and y may be different, if the scaling is different, the shape of the object will change

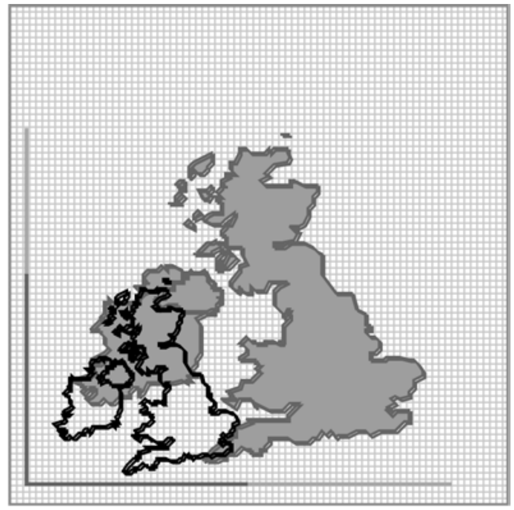
c. **Rotation**-origin fixed, axes move (rotate about origin)
$u = x\cos(a) + y\sin(a)$ $v = -x\sin(a) + y\cos(a)$
(here a is measured counterclockwise)

d. **Reflection**-coordinate system is reversed, objects appear in mirror image to reverse y, but not x: $u = x$ $v = c - y$
here, this transformation is important for displaying images on video monitors

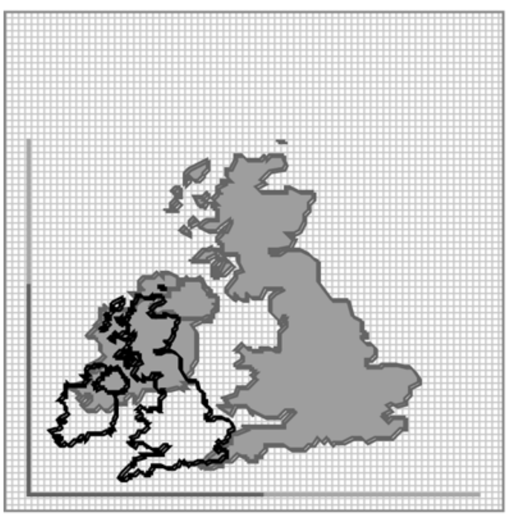
as the default coordinate system has the origin in the upper left corner and coordinates which run across and down.



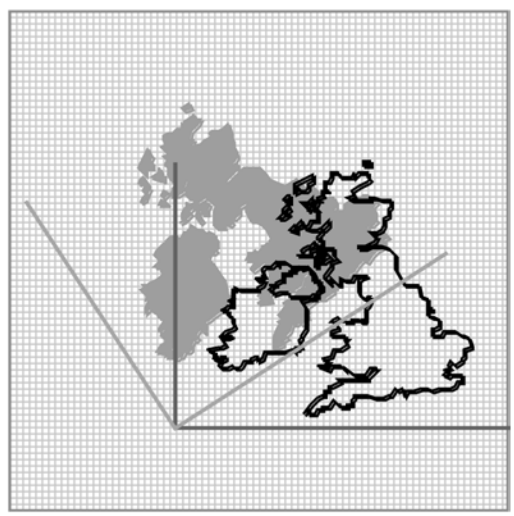
Translation (a & B)
origin is moved, axes do not rotate
u = x – a
= yv – b



Scaling (c & f)
both origin and axes are fixed,
scale changes
u = cx
v = cy



Rotation (a & d)
origin fixed, axes move
$u = x \cos(a) + y \sin(a)$
$v = -x \sin(a) + y \cos(a)$
(a = angle measured counter-clockwise)



Reflection
(b & e)
co-ordinate system is reversed, objects
appear in mirror image
u = x
v = c – y

Figure 3.22: Affine transformations.

COMPLEX AFFINE TRANSFORMATIONS: Usually a combination of these transformations will be needed because often we cannot actually separate the needed transformations into one or more of the primitives defined above as. One transformation will cause changes that appear to be caused by another transformation, and the order is important the combined equations are:

$$u = a + bx + cy \qquad v = d + ex + fy$$

AFFINE TRANSFORMATIONS IN GIS: Developing spatial databases for use in GIS, the data we use is generally on map sheets which use unknown or inaccurate projections and in order to register two data sets, a set of control points or tics must be identified that can be located on both maps. Here, it is necessary to have at least 3 control points since 3 points provide 6 values which can be used to solve for the 6 unknown points. Another precaution which is important that control points must not be on a straight line (not collinear).

Curvilinear Transformations

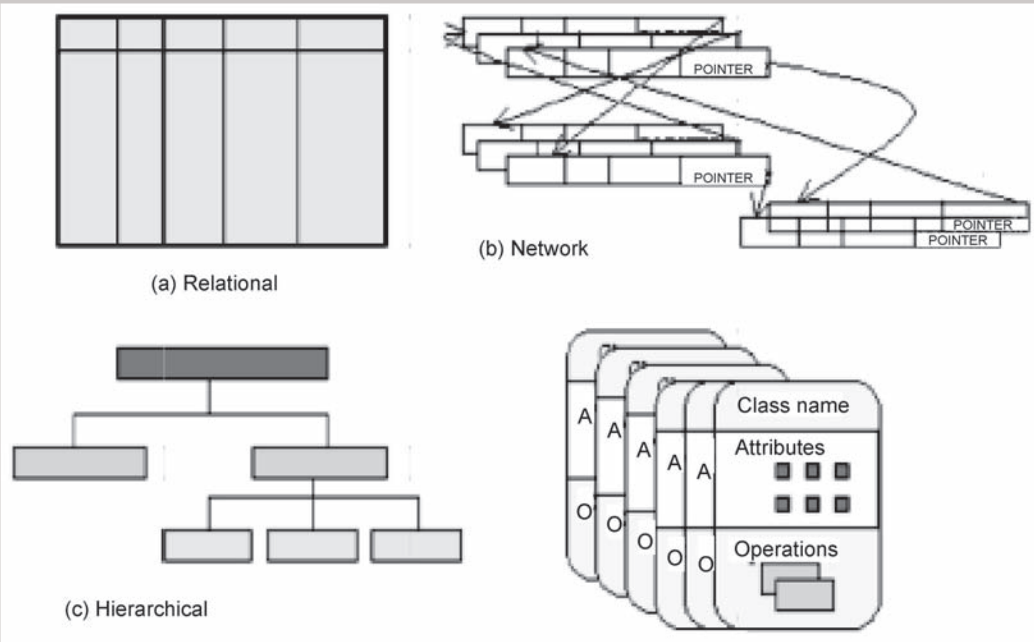
Simple linear affine transformation equations can be extended to higher powers:

$u = a + bx + cy + gxy$ or $u = a + bx + cy + gx^2$ or $u = a + bx + cy + gx^2 + hy^2 + ixy$ equations of this form create curved surfaces, provides rubbersheeting in which points are not transformed evenly over the sheet, transformations are not affine (parallel lines become non-parallel, possibly curved). Rubber-sheet transformations may also be piecewise, map divided into regions, each with its own transformation equations and equations must satisfy continuity conditions at the edges of regions. Curvilinear transformations usually give greater accuracy; accuracy means that when used to transform the control points or tics, the equations faithfully reproduce the known coordinates in the other system. However, if error in measurement is present, and it always is to some degree, then greater accuracy may not be desirable. A curvilinear transformation may be more accurate for the control points, but less accurate on average.

This page intentionally left blank

CHAPTER 4

SPATIAL DATA STRUCTURE AND MODELS



(a) Relational (b) Network (c) Hierarchical (d) Object-oriented

Information Organization and Data Structure

DATA AND INFORMATION: We use the terms 'data' and 'information' as synonyms but these two terms actually convey very distinct concepts. 'Data' is defined as a body of facts or figures, which have been gathered systematically for one or more specific purposes. Data can exist in the forms of:

- linguistic expressions (*e.g.,* name, age, address, date, ownership)
- symbolic expressions (*e.g.,* traffic signs)
- mathematical expressions (*e.g.,* $E = mc^2$)
- signals (*e.g.,* electromagnetic waves)

'Information' is defined as data which have been processed into a form that is meaningful to a recipient and is of perceived value in current or prospective decision making. Although data are ingredients of information, not all data make useful information. Data not properly collected and organized are a burden rather than an asset to an information user. Data that make useful information for one person may not be useful to another person. Information is only useful to its recipients when it is:

- relevant (to its intended purposes and with appropriate level of required detail)
- reliable, accurate and verifiable (by independent means)
- up-to-date and timely (depending on purposes)
- complete (in terms of attribute, spatial and temporal coverage)
- intelligible (*i.e.,* comprehensible by its recipients)
- consistent (with other sources of information)
- convenient/easy to handle and adequately protected.

The function of an *information system* is to change 'data' into 'information', using the following processes:

- conversion – transforming data from one format to another, from one unit of measurement to another, and/or from one feature classification to another
- organization – organizing or re-organizing data according to database management rules and procedures so that they can be accessed cost-effectively
- structuring – formatting or re-formatting data so that they can be acceptable to a particular software application or information system
- modelling – including statistical analysis and visualization of data that will improve user's knowledge base and intelligence in decision making.

The concepts of 'organization' and 'structure' are crucial to the functioning of information systems-without organization and structure it is simply impossible to turn data into information.

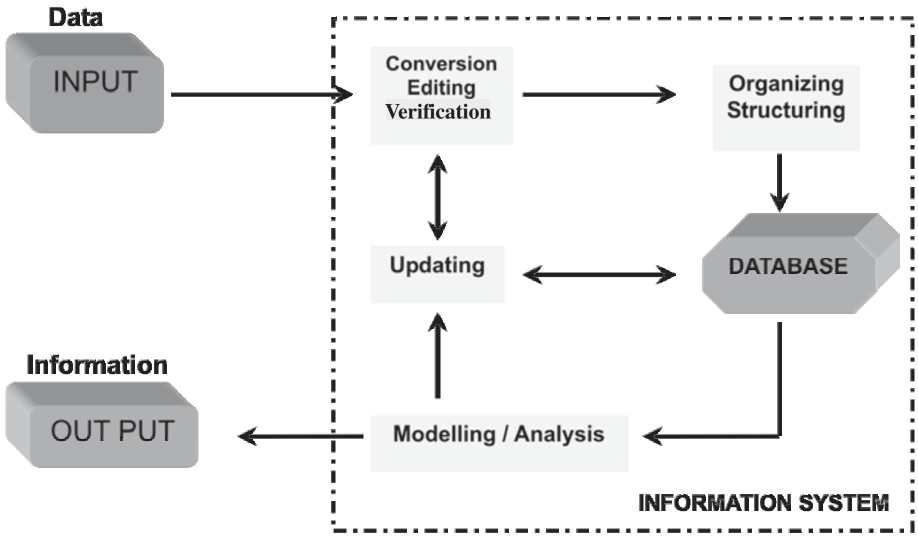


Figure 4.1: Changing data into information in an information system.

Geographic Data and Geographic Information

Geographic data are a special type of data; by 'geographic', it means that

- The data are pertinent to features and resources of the Earth, as well as the human activities based on or associated with these features and resources.
- The data are collected and used for problem solving and decision making associated with geography, *i.e.,* location, distribution and spatial relationships within a particular geographical framework.
- Geographic data are different from other types of data in that they are geographically referenced, *i.e.,* they can be identified and located by coordinates. They are made up of a *descriptive element* (which tells what they are) and a *graphical element* (which tells what they look like, where they are found and how they are spatially related to one another). The descriptive element is also commonly referred to as *non-spatial data* while the graphical element is also commonly referred to as *spatial data*.
- Geographic information is obtained by processing geographic data, the aim of which is to improve the user's knowledge about the geography of the Earth's features and resources, as well as human activities associated with these features and resources. It enable the user's to develop spatial intelligence for problem solving and decision making concerning the occurrence, utilization and conservation of the Earth's features and resources, as well as the impacts and consequences of human activities associated with them.
- Since the special nature and characteristics of geographic data, generic concepts of information organization and data structure cannot be applied directly to them.
- Geographic data have three dimensions:

a. Temporal – *e.g.,* 26th December 2004,
b. Thematic – *e.g.,* occurrence of tsunami in Indian Ocean,
c. Spatial – *e.g.,* affected area included south east coast of India.

GIS emphasizes on the use of the spatial dimension for turning data in to information, which assist our understanding of geographic phenomena.

Information Organization

Information organization can be understood from four perspectives:

- a data perspective
- a relationship perspective
- an operating system (OS) perspective
- an application architecture perspective

THE DATA PERSPECTIVE OF INFORMATION ORGANIZATION: The information organization of geographic data are considered in terms of their descriptive elements and graphical elements because these two types of data elements have distinctly different characteristics, they have different storage requirements and also they have different processing requirements.

INFORMATION ORGANIZATION OF DESCRIPTIVE DATA: The descriptive data, *data item* is the most basic element of information organization. A data item represents an *occurrence* or *instance* of a particular characteristic pertaining to an entity (which can be a person, thing, event or phenomenon). It is the smallest unit of stored data in a database, commonly referred to as an *attribute*. In database terminology, an attribute is also referred to as a *stored field.* The value of an attribute can be in the form of a number (integer or floating-point), a character string, a date or a logical expression (*e.g.,* T for 'true' or 'present'; F for 'false' or 'absent'). Some attributes have a definite set of values known as *permissible values* or *domain of values* (*e.g.,* age of people from 1 to 70; the categories in a land use classification scheme; and the academic departments in a university).

A group of related data items form a *record* (figure 4.2). Related data items, means that the items are occurrences of different characteristics pertaining to the same person, thing, event or phenomenon (*e.g.,* in a land resource inventory, a record may contain related data items such as identification number, owner, size of land holding and use of land etc.). A record may contain a combination of data items having different types of values (*e.g.,* in the above example, a record has two character strings representing the identification number and dominant use of land; an integer representing the average size of land holding rounded to the nearest meter; and a floating-point number representing identification). In database terminology, a record is always formally referred to as a *stored record* while in relational database management systems, records are called *tuples.*

A set of related records constitutes a *data file* (figure 4.2). Related records, means that the records represent different occurrences of the same type or class of people, things, events and phenomena. A data file made up of a single record type with single-valued data items is called a *flat file* (table 4.1). A data file made up of a single record type with nested repeating groups of items forming a multi-level organization is called a *hierarchical file* (table 4.1)

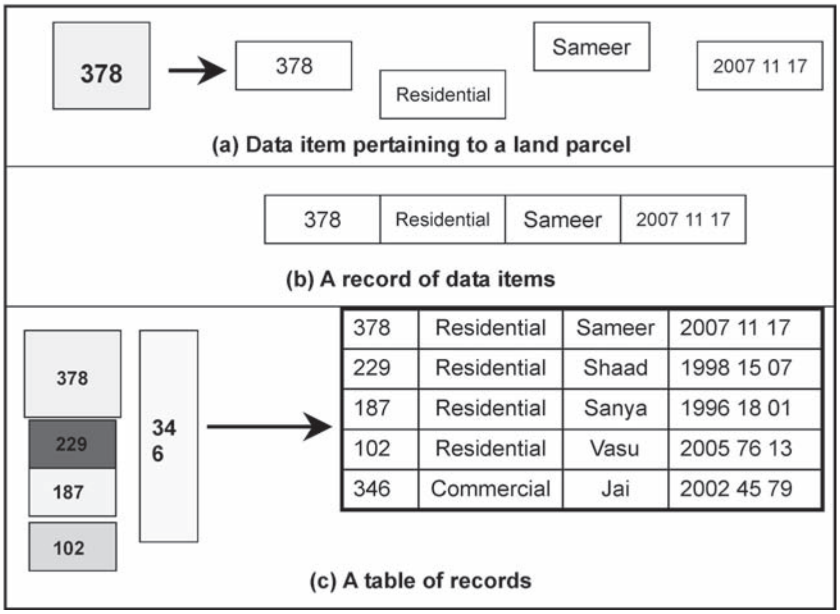


Figure 4.2: Data item, record, date file.

Table 4.1: Flat file and hierarchical file.

- **A Flat file**

Ward no.	Population	No. of households	Average monthly income
14	2431	654	Rs. 10,500
21	1740	389	Rs. 15,000
56	1985	557	Rs. 12,000

- **A Hierarchical file**

Ward no.	Population		No. of households		Average monthly income	
	1991	2001	1991	2001	1991	2001
14	1434	2431	568	654	Rs. 8,000	Rs. 10,500
21	1047	1740	307	389	Rs. 13,500	Rs. 15,000
56	1286	1985	489	557	Rs. 9,000	Rs. 12,000

A data file is individually identified by a *filename*. A data file may contain records having different types of data values or having a single type of data value. A data file containing records made up of character strings is called a *text file* or *ASCII file*. A data file containing records made up of numerical values in binary format is called a *binary file*. Data processing literature, collections of data items or records are sometimes referred to by other terms other than 'data file' according to their characteristics and functions. An *array* is a collection of data items of the same size and type (although they may have different values)

- a one-dimensional array is called a *vector*
- a two-dimensional array is called a *matrix*

A table is a data file with data items arranged in rows and columns. Data files in relational databases are organized as tables. Such tables are also called *relations* in relational database terminology. A *list* is a finite, ordered sequence of data items (known as *elements*). Here 'ordered', means that each element has a position in the list. An ordered list has elements positioned in ascending order of values; while an unordered list has no permanent relation between element values and position. Each element has a data type, in the simple list implementation, all elements must have the same data type but there is no conceptual objection to lists whose elements have different data types.

A *tree* is a data file in which each data item is attached to one or more data items directly beneath it (figure 4.3). The connections between data items are called *branches*. Trees are often called *inverted trees* because they are normally drawn with the root at the top.

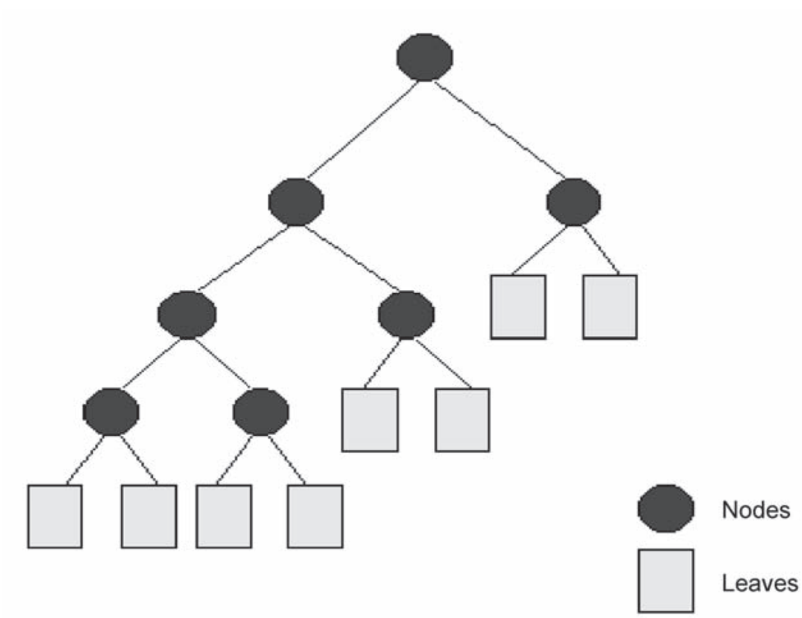


Figure 4.3: The tree data structure.

The data items at the very bottom of an inverted tree are called *leaves*; other data items are called *nodes*. A *binary tree* is a special type of inverted tree in which each element has only two branches below it. A *heap* is a special type of binary tree in which the value of each node is greater than the values of its leaves. Heap files are created for sorting data in computer processing-the *heap sort algorithm* works by first organizing a list of data into a heap.

The concept of *database* is the approach to information organization in computer-based data processing today. A database is defined as an automated, formally defined and centrally controlled collection of persistent data used and shared by different users in an enterprise. This definition excludes the informal, private and manual collection of data. 'Centrally controlled' means databases today tend to be physically distributed in different computer systems, at the same or different locations. A database is set up to serve the information needs of an organization and data sharing is key to the concept of database. Data in a database are described as 'permanent' in the sense that they are different from 'transient' data such as input to and output from an information system. The data usually remain in the database for a considerable length of time, although the actual content of the data can change very frequently. The use of database does not mean the demise of data files; data in a database are still organized and stored as data files. The use of database represents a *change* in the *perception* of data, the mode of data processing and the purposes of using the data, rather than physical storage of the data.

Table 4.2: Distinction between a data file and a database.

Characteristics of a data file	Characteristics of a database
A collection of records usually of the same data type and format description	A collection of interrelated records, organized in one or more data files, that may have different data types and format descriptions
Data file processing is usually associated with computer programming that aims at solving a particular problem, *i.e.,* it stops when an answer is obtained	Database processing is always associated with database management systems that aim at solving the operation or production needs of an organization, *i.e.,* it involves routine, largely repetitive applications executed over and over again
Mainly used in support of the information need of an *ad hoc* application	Mainly used in support of the day to day operation of business (transaction processing) but increasingly used in decision support (management decision making)

Databases can be organized in different ways known as *database models*. The conventional database models are: *relational, network, hierarchical and object-oriented* (figure 4.4).

- relational-data are organized by records in relations which resemble a table
- network-data are organized by records which are classified into record types, with 1: n pointers linking associated records
- hierarchical-data are organized by records on a parent-child one-to-many relations
- object oriented-data are uniquely identified as individual objects that are classified into object types or classes according to the characteristics (attributes and operations) of the object.

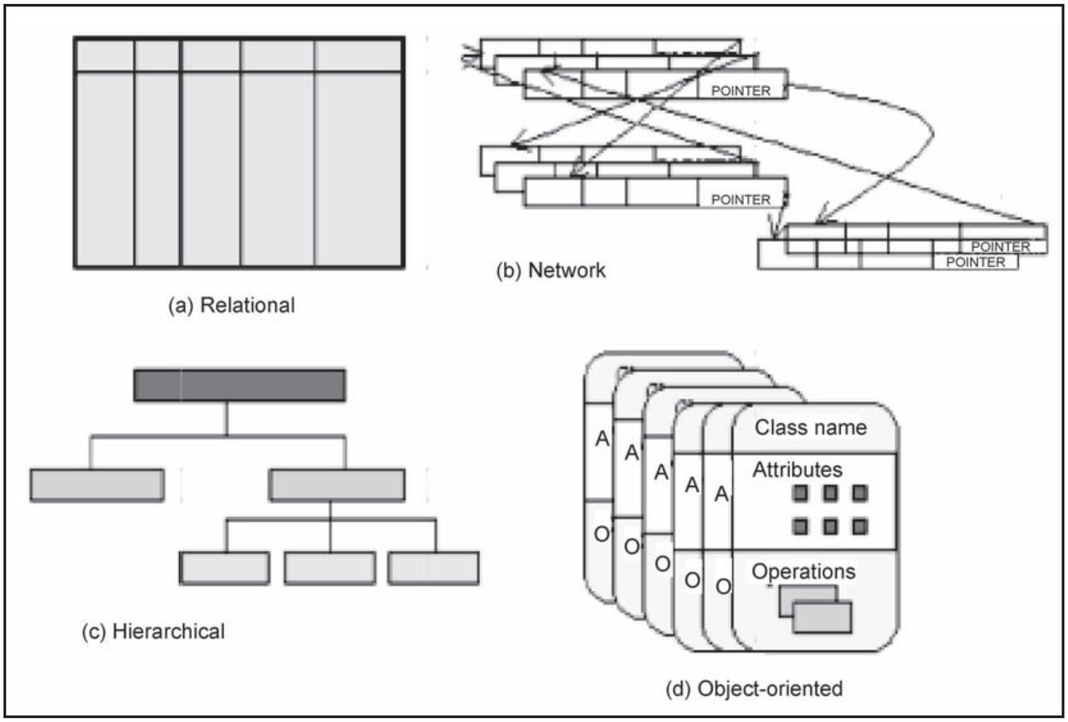


Figure 4.4: Database models.

Information Organization of Graphical Data

The graphical data, where the most basic element of information organization is called as *basic graphical element.* There are three basic graphical elements (figure 4.5):

- *point*
- *line*, also referred to as *arc*
- *polygon*, also referred to as *area*

These basic graphical elements can be individually used to represent geographic features or entities for example, point for a well; line for a road segment and polygon for a lake. They can also be used to construct complex features. For example, the geographic entity 'India' on a map is represented by a group of polygons of different sizes and shapes.

POINT FEATURE: A point has neither length nor breadth and hence is said to be of dimension 0. A point feature represents as single location. A point is the simplest graphical representation of an object. Points have no dimensions but may be indicated on maps or displayed on screens by using symbols. The corner of a property boundary is a typical point, as is the representative coordinate of a building. It is, of course, the scale of viewing that determines whether an object is defined as a point or an area. In a large-scale representation a building may be shown as an area, whereas it may only be a point (symbol) if the scale is reduced.

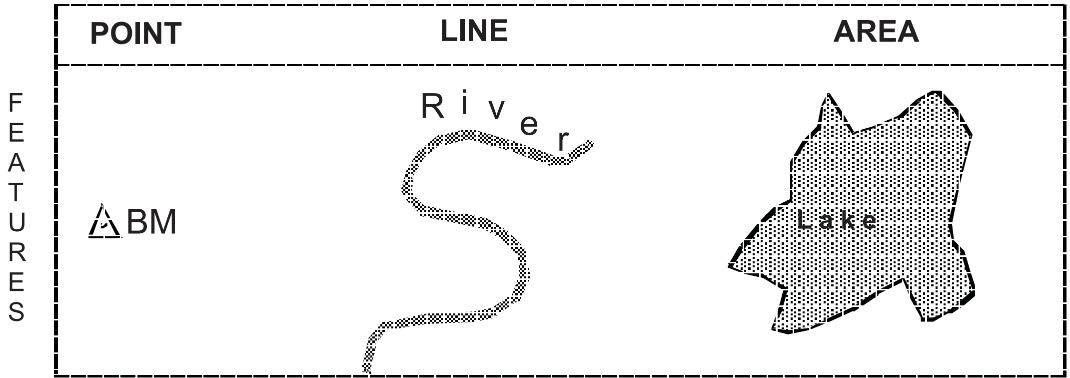


Figure 4.5: The feature model: Examples of a point feature (elevation bench mark), a line feature (river) and an area feature (lake).

LINE FEATURE: Lines have length, but not breadth hence is of dimension 1. They are used to represent linear entities such as rivers, roads, pipelines, and cables etc. A line feature is a set of connected, ordered coordinates representing the linear shape of a map object that may be too narrow to display as an area such as a road or feature with no width such as a contour line.

AREA FEATURE: Area objects have the two dimensions of length and breadth. An area feature is a closed figure whose boundary encloses homogeneous characteristics, such as a state boundary, soil type or lake. Again, physical size in relation to the scale determines whether an object is represented by an area or by a point. An area is delineated by at least three connecting lines, each of which comprises points. In databases, areas are represented by polygons (*i.e.,* plane figures enclosed by at least three straight lines intersecting at a like number of points). Therefore, the term polygon is often used instead of area.

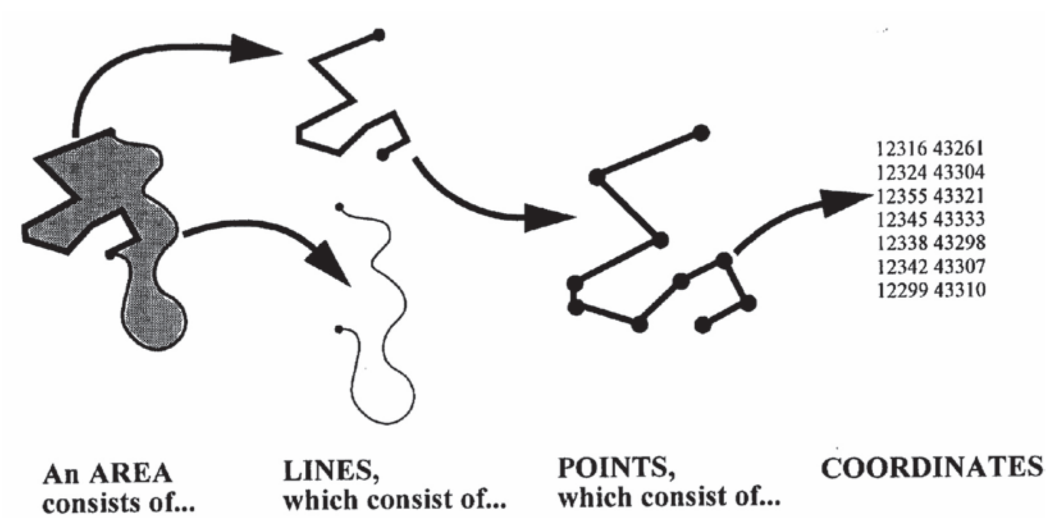


Figure 4.6: Geographic information has dimensions, areas are two dimensional and consists of lines, which are one dimensional and consists of points, which are zero dimensional and consist of a coordinate pair.

The method of representing geographic features by the basic graphical elements of points, lines and polygon is said to be the *vector method* or *vector data model*, and the data are called *vector data*. Related vector data are always organized by *themes*, which are also referred to as *layers* or *coverages*, examples of themes: base map, soil, vegetation cover, land use, transportation, drainage and hydrology, political boundaries, land parcel and others. For themes covering a very large geographic area, the data are always divided into *tiles* so that they can be managed more easily. A tile is the digital equivalent of an individual map in a map series, it is uniquely identified by a file name. A collection of themes of vector data covering the same geographic area and serving the common needs of a multitude of users constitutes the *spatial component* of a *geographical database*. The vector method of representing geographic features is based on the concept that these features can be identified as discrete entities or objects, this method is therefore based on the *object view of the real world* (Goodchild, 1992).

The object view is the method of information organization in conventional mapping and cartography. Graphical data captured by imaging devices in remote sensing and digital cartography (such as multi-spectral scanners, digital cameras and image scanners) are made up of a matrix of picture elements (pixels) of very fine resolution. Geographic features in such form of data can be visually recognized but not individually identified in the same way that geographic features are identified in the vector method. They are recognizable by differentiating their spectral or radiometric characteristics from pixels of adjacent features, for example, a lake can be visually recognized on a satellite image because the pixels forming it are darker than those of the surrounding features; but the pixels forming the lake are not identified as a single discrete geographic entity, *i.e.,* they remain individual pixels or a highway can be visually recognized on the same satellite image because of its particular shape; but the pixels forming the highway do not constitute a single discrete geographic entity as in the case of vector data.

The method of representing geographic features by pixels is called the *raster method* or *raster data model*, and the data are described as *raster data*. A raster pixel represents the generalized characteristics of an area of specific size on or near the surface of the Earth. The actual ground size depicted by a pixel is dependent on the resolution of the data, which may range from smaller than a square meter to several square kilometers. Raster data are organized by themes, which are also referred to as layers for example; a raster geographic database may contain the following themes: bed rock geology, vegetation cover, land use, topography, hydrology, rainfall, temperature. Raster data covering a large geographic area are organized by *scenes* (for remote sensing images) of by *raster data files* (for images obtained by map scanning). The raster method is based on the concept that geographic features are represented as surfaces, regions or segments, this method is therefore based on the *field view of the real world* (Goodchild, 1992).

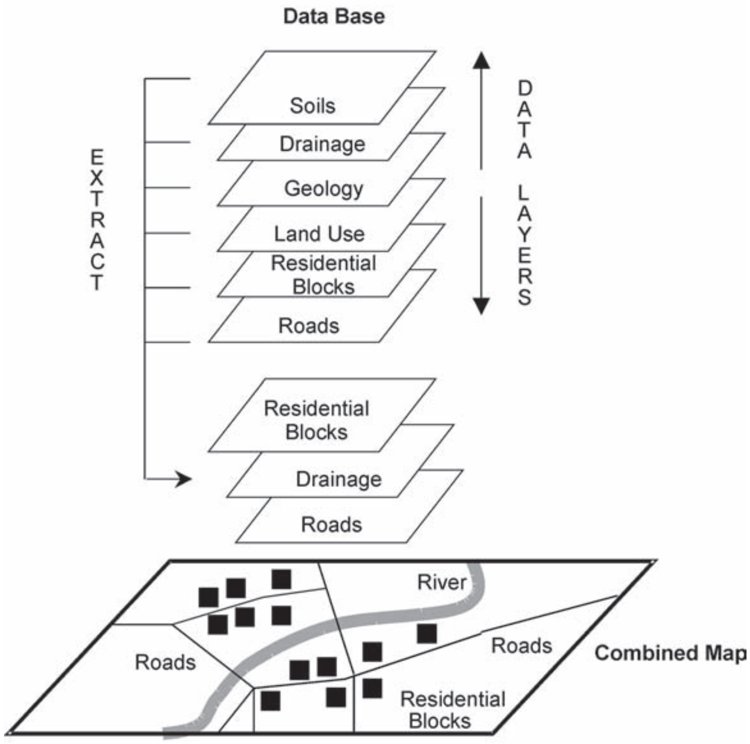


Figure 4.7: The layer based approach.

Levels of data abstraction

Information organization is concerned with the internal organization of data. It represents the user's view of data, *i.e., c*onceptualization of the real world. It is the lowest level of data abstraction, which can be done with or without any intent for computer implementation and it is expressed in terms of *data models* (Peuquet, 1991).

The difference between "data models" and "database models" is:

The vector and raster methods of representing the real world are "data models" and,

The relational, network, hierarchical and object-oriented databases are "database models" --- they are the software implementation of data models

Data structure represents a higher level of data abstraction than information organization in the sense that it is concerned with the design and implementation of information organization. It represents the human implementation-oriented view of data and expressed in terms of database models, this implies that data structure is software-dependent but hardware is not yet a consideration. Data structure forms

Contd...

Box 6: Data structure

the basis for the next level of data abstraction in information system: *file structure* or *file format.* File structure is the hardware implementation-oriented view of data, which reflects the physical storage of the data on some specific computer media such as magnetic tapes or hard disk. This implies that file structure is hardware-dependent.

Descriptive data structures

Descriptive data structures describe the design and implementation of the information organization of non-spatial data. As most commercial implementations of information systems today are based on the relational and object-oriented database models.

Relational data structure: The relational data structure is the table which is formally called a relation.

Object-oriented data structure: Unlike the relational data structure, there is not a formalized object-oriented data structure, this means that different object-orientation implementations have different data structures.

Graphical data structures

Raster data structure: In the raster data structure space is subdivided into regular grids of square grid cells or other forms of polygonal meshes known as picture elements (pixels). There are several variants to the regular grid raster data structure, including: *irregular tessellation* (*e.g.,* triangulated irregular network (TIN)), *hierarchical tessellation* (*e.g.,* quad tree) and *scan-line* (Peuquet, 1991)

Vector data structure: there are many implementations of vector data structures, including: spaghetti - a direct line-for-line unstructured translation of the paper map, hierarchical - a vector data structure developed to facilitate data retrieval by separately storing points, lines and areas in a logically hierarchical manner and topological - a vector data structure that aims at retaining spatial relationship by explicitly storing adjacency information.

The georelational data structure

The georelational data structure was developed to handle geographic data. It allows the association between spatial (graphical) and non-spatial (descriptive) data. Both spatial and non-spatial data are stored in relational tables and entities in the spatial and non-spatial relational tables are linked by the common FIDs of entities.

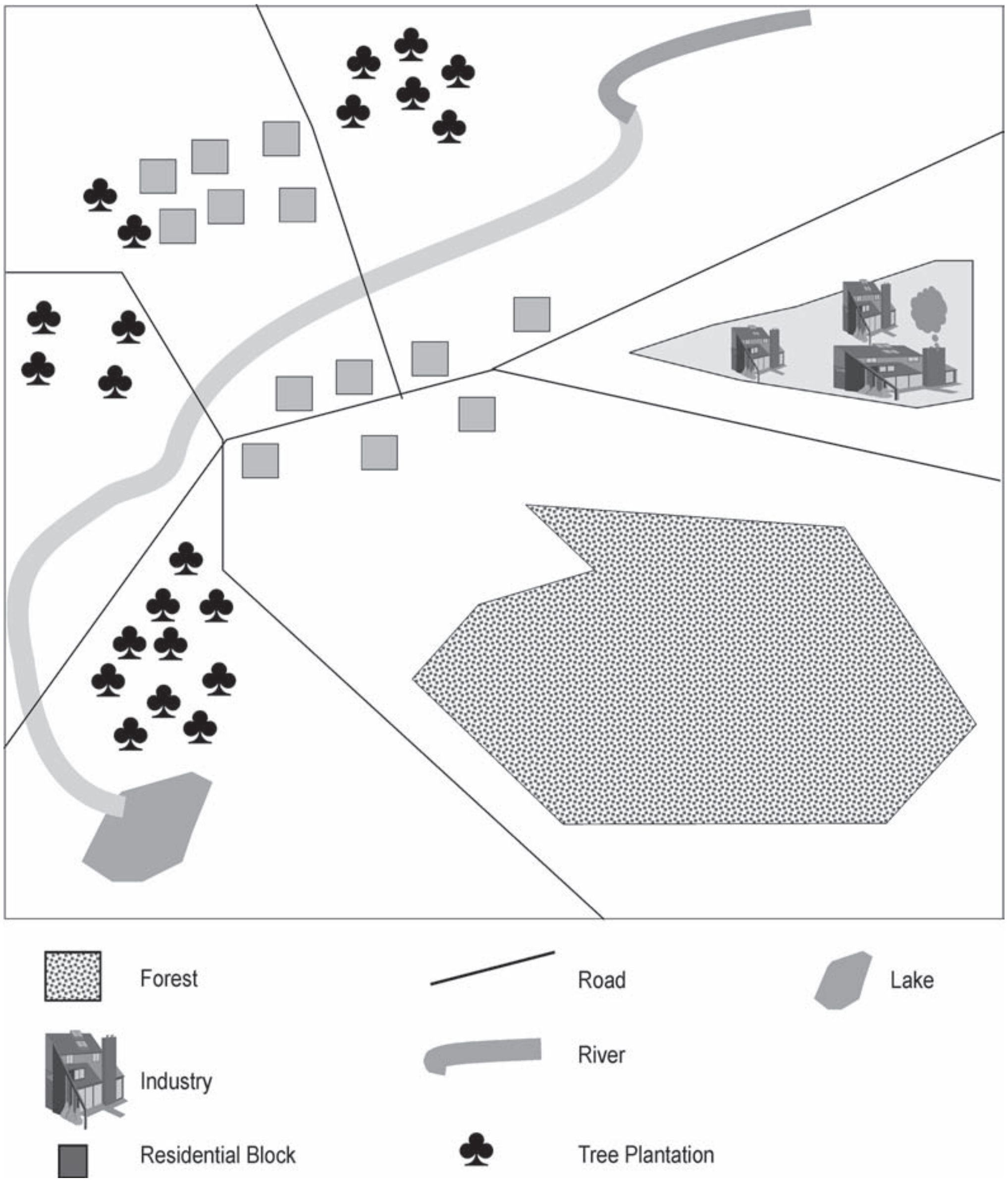


Figure 4.8: The object-oriented approach.

The Relationship Perspective of Information Organization

Relationships represent an important concept in information organization – it describes the logical association between entities. Relationships can be *categorical* or *spatial,* depending on whether they describe location or other characteristics.

Categorical relationships: Categorical relationships describe the association among individual features in a classification system. The classification of data is based on the concept of *scale of measurement*:

There are four scales of measurement:

- Nominal – a qualitative, non-numerical and non-ranking scale that classifies features on intrinsic characteristics for example, in a land use classification scheme, polygons can be classified as industrial, commercial, residential, agricultural, public and institutional.
- Ordinal – a nominal scale with ranking which differentiates features according to a particular order for example, in a land use classification scheme, residential land can be denoted as low density, medium density and high density.
- Interval – an ordinal scale with ranking based on numerical values that are recorded with reference to an arbitrary datum for example, temperature readings in degrees centigrade are measured with reference to an arbitrary zero (*i.e.,* zero degree temperature does not mean no temperature).
- Ratio – an interval scale with ranking based on numerical values that are measured with reference to an absolute datum for example, rainfall data are recorded in mm with reference to an absolute zero (*i.e.,* zero mm rainfall mean no rainfall).

Categorical relationships based on ranking are hierarchical or taxonomic in nature which means that data are classified into progressively different levels of detail. Data in the top level are represented by limited broad basic categories. Data in each basic category are then classified into different sub-categories, which can be further classified into another level if necessary. The classification of descriptive data is typically based on categorical relationships.

Table 4.3: Example of a classification scheme of descriptive data.

Level I	Level II
1. Built-up Land	1.1 Residential 1.2 Commercial 1.3 Industrial 1.4 Services 1.5 Transportation
2. Agricultural Land	2.1 Crop Land 2.2 Orchards, Vineyards, Nurseries 2.3 Pastures
3. Forest Land	3.1 Mixed Forest 3.2 Evergreen Forest 3.3 Deciduous Forest
4. Water Bodies	4.1 Rivers 4.2 Pond/Lake 4.3 Water Logged Area

SPATIAL RELATIONSHIPS: Spatial relationships describe the association among different features in space. Spatial relationships are visually obvious when data are presented in the graphical form. However, it is difficult to build spatial relationships into the information organization and data structure of a database. There are numerous types of spatial relationships possible among features. Recording spatial relationships implicitly demands considerable storage space. Computing spatial relationships on-the-fly slows down data processing particularly if relationship information is required frequently.

There are two types of spatial relationships (figure 4.9)

- topological – describes the property of adjacency, connectivity and containment of contiguous features.
- proximal – describes the property of closeness of non-contiguous features.

Spatial relationships are very important in geographical data processing and modelling. The objective of information organization and data structure is to find a way that will handle spatial relationships with the minimum storage and computation requirements.

Table 4.4: Point-line-area relationship matrix

	Point	**Line**	**Area**
Point	Is nearest to Is neighbour of	Ends at Is nearest to Lies on	Is within Outside of Can be seen from
Line		Crosses Joins Flows into Comes within Is parallel to	Crosses Borders Intersects
Area			Overlaps Is nearest to Is adjacent to Is contained in

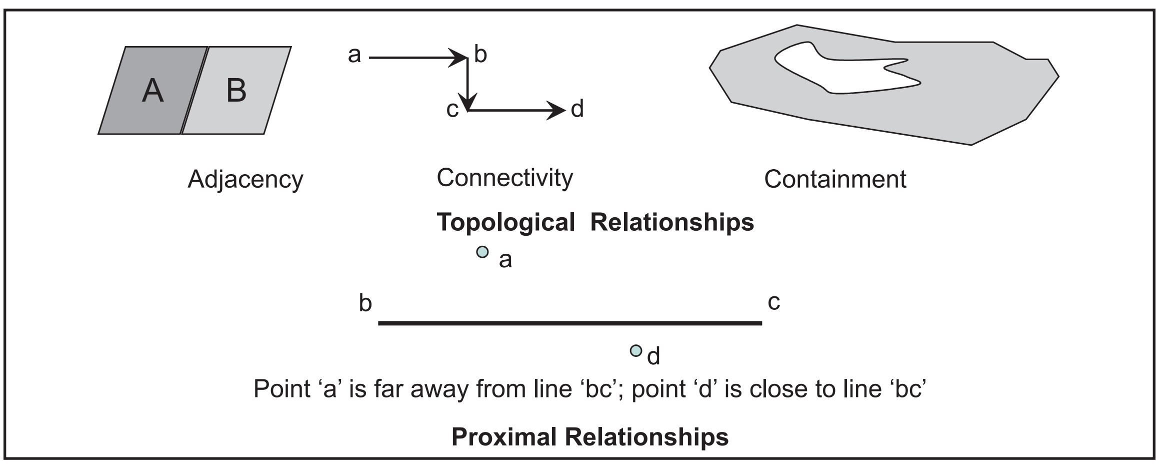


Figure 4.9: Topological and proximal relationships.

The Operating System (OS) Perspective of Information Organization

From the operating system perspective, information is organized in the form of *directories*. Directories are a special type of computer files used to organize other files into a hierarchical structure (figure 4.10). Directories are also referred to as *folders*, particularly in systems using graphical user interfaces. A directory may also contain one of more directories:

- the topmost directory in a computer is called the root directory
- a directory that is below another directory is referred to as a *sub-directory*
- a directory that is above another directory is referred to as a *parent directory*

Directories are designed for bookkeeping purposes in computer systems and a directory is identified by a unique directory name. Computer files of the same nature are usually put under the same directory. A data file can be accessed in a computer system by specifying a *path* that is made up of the device name, one or more directory names and its own file name. The concept of *workspace* used by many geographic information system software packages is based on the directory structure of the host computer. A workspace is a directory under which all data files relating to a particular project are stored (figure 4.11).

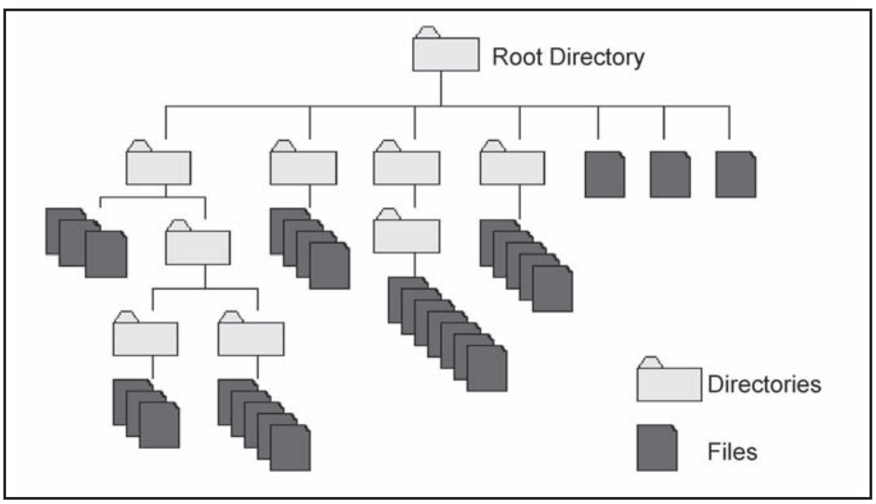


Figure 4.10: Information organization by directories.

The Application Architecture Perspective of Information Organization

Computer applications nowadays tend to be constructed on the *client/server* systems architecture. Client/server is primarily a relationship between processes running in the same computer or, more commonly, in separate computers across a telecommunication network. The *client* is a process that requests services, the dialog between the client and the server is always initiated by the client and client can request services from many servers at the same time. The *server* is a process that provides the service, a server is primarily a passive service provider and a server can service many clients at the same time.

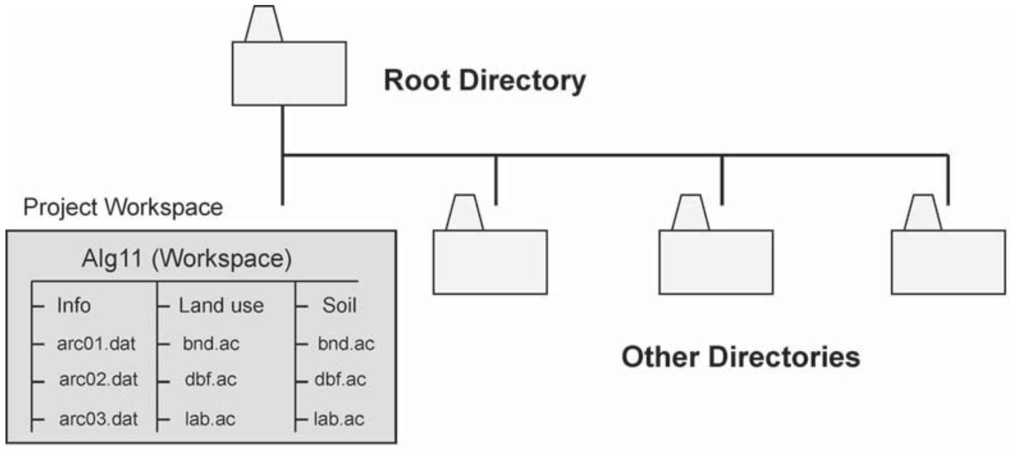


Figure 4.11: Example of a GIS project workspace.

There are many ways of implementing a client/server architecture but from the perspective of information organization, the following five are most important:

- *file servers* – the client requests specific records from a file; and the server returns these records to the client by transmitting them across the network
- *database servers* – the *client* sends *structured query language* (SQL) requests to the server; the server finds the required information by processing these requests and then passes the results back to the client
- *transaction servers* – the client invokes a remote procedure that executes a transaction at the server side; the server returns the result back to the client via the network
- *Web server* – communicating interactively by the *Hypertext Transfer Protocol* (HTTP) over the Internet, the Web server returns documents when clients ask for them by name
- *groupware servers* – this particular type of servers provides a set of applications that allow clients (and their users) to communicate with one another using text, images, bulletin boards, video and other forms of media.

From the application architecture perspective, the objective of information organization and data structure is to develop a data design strategy that will optimize system operation by balancing the distribution of data resources between the client and the server. The databases are typically located on the server to enable data sharing by multiple users. Static data that are used for reference are usually allocated to the client, ensuring the logical allocation of data resources among different servers. Data that are commonly used together should be placed in the same server while data that have common security requirements should be placed in the same server. Data intended for a particular purpose (file service, database query, transaction processing, Web browsing or groupware applications) should be placed in the appropriate server.

Data – Fundamental Concepts

DATA: Data are facts. Some facts are more important to us than others. Some facts are important enough to warrant keeping track of them in a formal, organized way. Important data are like the books we keep in a library almirah. They are a small subset of our total collection but they are so important that we protect them by putting them in a special, safe place. 'Data' is a plural and is a broad concept that can include things such as pictures (binary images), programs, and rules. Informally, *data* are the things we want to store in a *database*.

Spatial – Non-Spatial Data

Spatial data includes location, shape, size, and orientation. For example, a particular square: its center (the intersection of its diagonals) specifies its location, its shape is a square, the length of one of its sides specifies its size and the angle its diagonals, say, the x-axis specifies its orientation.

Spatial data includes spatial relationships. For example, the arrangement of three stumps in a cricket ground is spatial data.

Non-spatial data (also called *attribute* or *characteristic* data) is that information which is independent of all geometric considerations. For example, a person's height, mass, and age are non-spatial data because they are independent of the person's location. It is possible to ignore the distinction between spatial and non-spatial data. However, there are fundamental differences between them:

- spatial data are generally multi-dimensional and auto-correlated.
- non-spatial data are generally one-dimensional and independent.

These distinctions put spatial and non-spatial data into different philosophical camps with far-reaching implications for conceptual, processing, and storage issues. For example, sorting is perhaps the most common and important non-spatial data processing function that is performed. It is not obvious how to even sort locational data such that all points end up 'nearby' their nearest neighbours. These distinctions justify a separate consideration of spatial and non-spatial data models.

DATABASES FOR SPATIAL DATA: A database is a collection of facts, a set of data. The information in a phone book is an example of a database. The book itself is not the database, rather, the database is the information stored on the pages of the book, not the pieces of paper with ink on them.

Many different data types are encountered in geographical data, *e.g.,* pictures, words, coordinates, complex objects, but very few database systems have been able to handle textual data, *e.g.,* descriptions of soils in the legend of a soil map can run to hundreds of words. This is the primary reason why some GIS designers have chosen not to use standard database solutions for coordinate data, but only for attribute tables. Because variable length records are needed, often not handled well by standard systems, *e.g.,* number of coordinates in a line can vary.

Standard database systems assume the order of records is not meaningful. In geographical data the positions of objects establish an implied order which is important in many operations and often need to work with objects that are adjacent in space, thus it helps to have these objects adjacent or close in the database. This is a problem with standard database systems since they do not allow linkages between objects in the same record type (class). There are so many possible relationships between spatial objects, that not all can be stored explicitly, however, some relationships must be stored explicitly as they cannot be computed from the geometry of the objects, *e.g.,* existence of grade separation at street crossing. The integrity rules of geographical data are too complex, *e.g.,* the arcs forming a polygon must link into a complete boundary.

Database has been an important issue in GIS, initially attempts to build GIS began using very limited tools like operating systems and compilers. More recently, GIS have been built around existing database management systems (DBMS). The DBMS handles many functions which would otherwise have to be programmed into the GIS. Any DBMS makes assumptions about the data which it handles and to make effective use of a DBMS it is necessary to fit those assumptions. Certain types of DBMS are more suitable for GIS than others because their assumptions fit spatial data better. There are two ways to use DBMS in a GIS:

I. Total DBMS solution: All data are accessed through the DBMS, so must fit the assumptions imposed by the DBMS designer.

II. Mixed solution: Some data (usually attribute tables and relationships) are accessed through the DBMS because they fit the model well, while some data (usually locational) are accessed directly because they do not fit the DBMS model.

REPOSITORY: A repository is a structure that stores and protects data. Repositories provide the following functionality:

- add (insert) data to the repository
- retrieve (find, select) data in the repository
- delete data from the repository

Some repositories allow data to be changed, to be updated. This is not strictly necessary because an update can be accomplished by retrieving a copy of the datum from the repository, updating the copy, deleting the old datum from the repository, and inserting the updated datum into storage. Repositories are like a bank vault. They exist mainly to protect their contents from theft and accidental destruction.

- *Security:* Repositories are typically password protected, many have much more elaborate security mechanisms.
- *Robustness:* Accidental data loss is safeguarded against via the *transaction* mechanism.

A *transaction* is a sequence of database manipulation operations. Transactions have the property that, if they are interrupted before they complete, the database will be restored to a self-consistent state, usually the one before the transaction began. If the transaction completes, the database will be in a self-consistent state. Transactions protect the data from power failures, system crashes, and concurrent user interference.

ADVANTAGES OF A DATABASE APPROACH: The advantages of this approach include:

- reduction in data redundancy
- shared rather than independent databases, which reduces problem of inconsistencies in stored information, *e.g.,* different addresses in different wards for a postman
- maintenance of data integrity and quality
- data are self-documented or self-descriptive, where information on the meaning or interpretation of the data can be stored in the database, *e.g.,* names of items, metadata
- avoidance of inconsistencies, which means data must follow prescribed models, rules, standards
- reduced cost of software development
- security restrictions, which means database includes security tools to control access, particularly for writing.

Database Management System (DBMS)

A *database management system* is a data repository along with a user interface providing for the manipulation and administration of a database. A phone book is an example of a DBMS. A DBMS is like a full-service bank, providing many features and services missing from the comparatively Spartan repository. It is a software system, a program (or suite of programs) that is run on a digital computer. A few examples of commercially available DBMSs include Codasyl, Sybase, Oracle, DB2, Access, and dBase.

Queries: Many DBMSs provide a user interface consisting of some sort of formal language.

- ***A data definition language*** (DDL) is used to specify which data will be stored in the database and how they are related.
- ***A data manipulation language*** (DML) is used to add, retrieve, update, and delete data in the DBMS.
- A ***query*** is often taken as a statement or group of statements in either a DDL or a DML or both. Some researchers view queries as read-only operations, no data modifications are allowed.
- ***A query language*** is a formal language that implements a DDL, a DML, or both. Examples of query languages include SQL (Structured Query Language), QUEL, ISBL, and Query-by-Example.

Data Models

A *data model* is mathematical formalism consisting of two parts A notation for describing data, and A set of operations used to manipulate that data. A data model is a way of organizing a collection of facts pertaining to a system under investigation. Data models provide a way of thinking about the world, a way of organizing the phenomena that interest us. They can be thought of as an abstract language, a collection of words along with a grammar by which we describe our subject. By choosing a language, words are limited to those in the language only and whose sentence structure is governed by the language's grammar. We are not free to use random collections of symbols for words nor can we put the words together in any *ad hoc* fashion.

A major benefit we receive by following a data model stems from the theoretical foundation of the model. From the theory emerges the power of analysis, the ability to extract inferences and to create deductions that emerge from the raw data. Different models provide different conceptualizations of the world; they have different outlooks and different perspectives.

DBMSs are seen to be composed of three levels of abstraction:

- ***Physical:*** This is the implementation of the database in a digital computer. It is concerned with things like storage structures and access method data structures.
- ***Conceptual:*** This is the expression of the database designer's model of the real world in the language of the data model.
- ***View:*** Different user groups can be given access to different portions of the database. A user groups portion of the database is called their view.

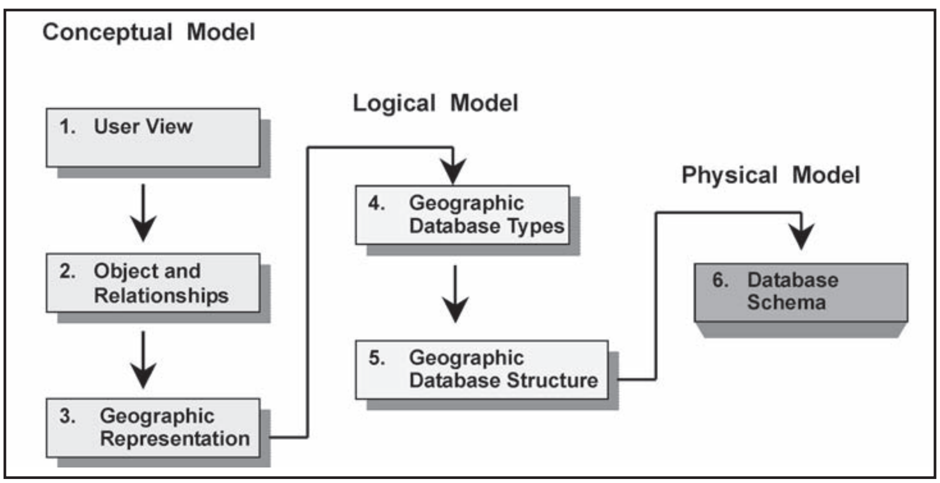


Figure 4.12: Stages in database design.

DATA MODELLING

Data modelling is the process of defining real world phenomena or geographic features of interest in terms of their characteristics and their relationships with one another. It is concerned with different phases of work carried out to implement information organization and data structure. There are three steps in the data modelling process, resulting in a series of progressively formalized data models as the form of the database becomes more and more rigorously defined

- Conceptual data modelling–Defining in broad and generic terms the scope and requirements of a database.

- ***Logical data modelling*** – Specifying the user's view of the database with a clear definition of attributes and relationships.
- ***Physical data modelling*** – Specifying internal storage structure and file organization of the database.

Data modelling is closely related to the three levels of data abstraction in database design:

- conceptual data modelling ⇒ data model
- logical data modelling ⇒ data structure
- physical data modelling ⇒ file structure

a. Conceptual data modelling: *Entity-relationship* (E-R) modelling is probably the most popular method of conceptual data modelling. It is sometimes referred to as a method of *semantic data modelling* because it used a human language-like vocabulary to describe information organization, involving four aspects of work:
  - identifying entities – defined as a person, a place, an event, a thing, etc.
  - identifying attributes
  - determining relationships
  - drawing an *entity-relationship diagram* (E-R diagram)

b. Logical data modelling: Logical data modelling is a comprehensive process by which the conceptual data model is consolidated and refined. The proposed database is reviewed in its entirety in order to identify potential problems such as: irrelevant data that will not be used; omitted or missing data; inappropriate representation of entities; lack of integration between various parts of the database; unsupported applications; and potential additional cost to revise the database. The end product of logical data modelling is a **logical schema** which is developed by mapping the conceptual data model (such as the E-R diagram) to a software – dependent design document.

c. Physical data modelling: Physical data modelling is the database design process by which the actual tables that will be used to store the data are defined in terms of:

- data format – the format of the data that is specific to a database management system (DBMS).
- storage requirements – the volume of the database.
- physical location of data – optimizing system performance by minimizing the need to transmit data between different storage devices or data servers.

The end product of physical data modelling is a *physical schema,* which is also variably known as *data dictionary, item definition table, data specific table* or *physical database definition*. It is both software – and hardware specific, this means the physical schemas for different systems look different from one another.

d. Process modelling: Process modelling is the process-oriented approach, as opposed to the data-oriented approach, of information system design. The objective is to identify the processes that the information system will perform. It also aims at identifying how information is transformed from one process to another. The end product of process modelling is a *data flow diagram* (DFD), this implies that process modelling is by no means only concerned with process, it also deals with information organization and data structure. In the context of information system design, process modelling is one of the methods of *structured business function decomposition* used to determine user requirements in conceptual modelling.

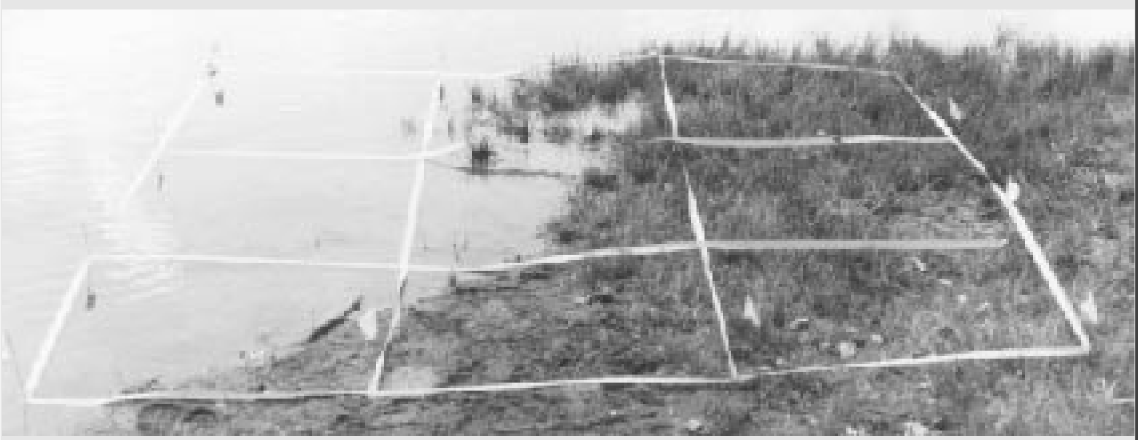
Data flow diagram is the principal modelling tool which is constructed using four basic symbols to represent *process, data stores, entities* and *data flow* in a business function:

- process – it represents the transformation of data as they flow through the system: data flow into a process, are changed, and then flow out to another process or a data store.
- entity – the basic definition of an entity is similar to that for E-R modelling and it represents the initial source and final destination of data in a DFD.
- data store – a temporary or permanent holding area for data.
- data flow – the connection between processes and data stores along which individual entities or collection of entities flow.

This page intentionally left blank

Chapter 5

The Nature and Source of Geographic Data



Much of GIS analysis and description consists of investigating the properties of geographic features and determining the relationships between them. The chosen way of representing phenomena in GIS not only defines the apparent nature of geographic variation, but also the way in which geographic variation may be analyzed. Some objects, such as agricultural fields or digital terrain models, are represented in their natural state. Others are transformed from one spatial object class to another, as in the transformation of population data from individual points to census tract areas, for reasons of confidentiality, convenience, or convention. The classification of spatial phenomena into object types is fundamentally dependent upon scale. For example, on a less-detailed map of the world, New Delhi is represented as a zero-dimensional point. On a more-detailed map such as a road atlas it will be represented as a two-dimensional area. Yet if we visit the city, it is very much experienced as a three-dimensional entity, and virtual reality systems seek to represent it as such. These features are represented through coordinates, so areas are lines are points are coordinates.

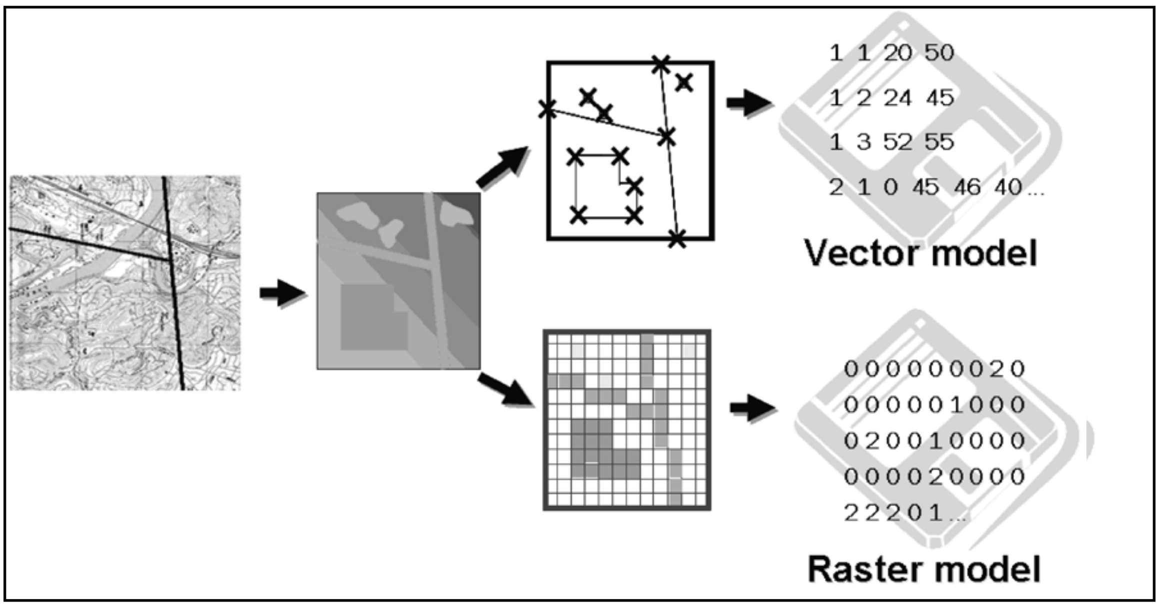


Figure 5.1: Modelling the real world.

SPATIAL DATA FORMATS: Raster Data Format: Raster data represents a graphic object as a pattern of dots, whereas vector data represents the object as a set of lines drawn between specific points. Consider a line drawn diagonally on a piece of paper. A raster file would represent this image by subdividing the paper into a matrix of small rectangles-similar to a sheet of graph paper-called cells. Each cell is assigned a position in the data file and given a value based on the attribute at that position. Its row and column co-ordinates may identify any individual pixel. This data representation allows the user to easily reconstruct or visualize the original image.

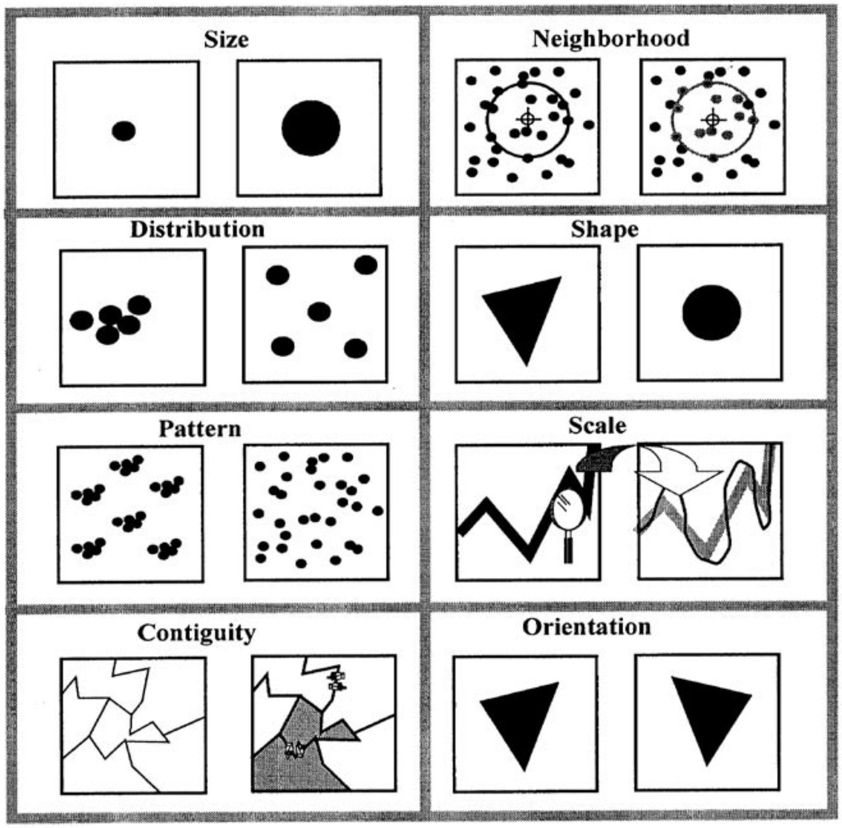


Figure 5.2: Basic properties of geographic features.

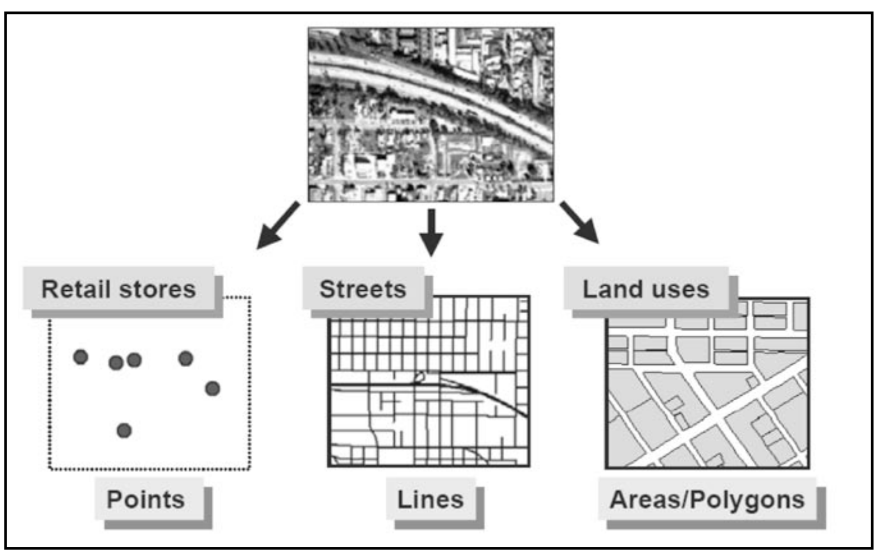


Figure 5.3: Representation of geographic details, point, line and area features.

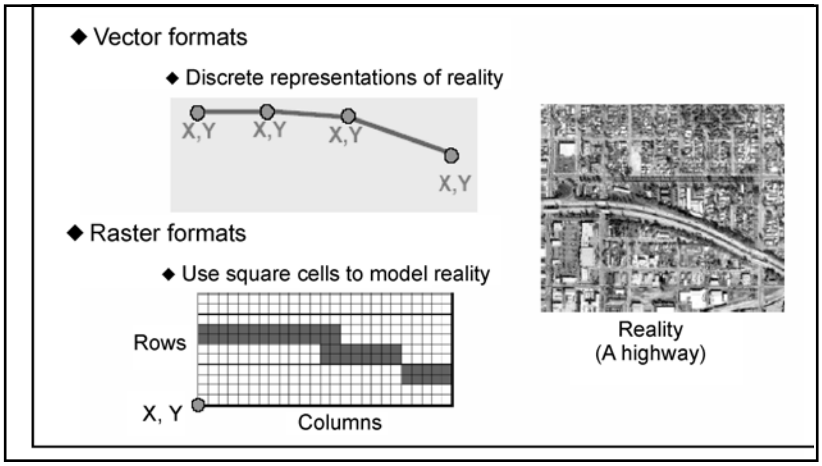


Figure 5.4: Storing of spatial data, vector and raster data formats.

Raster files are most often used:

- For digital representations of aerial photographs, satellite images, scanned paper maps, and other applications with very detailed images.
- When costs need to be kept down.
- When the map does not require analysis of individual map features.
- When 'backdrop' maps are required.

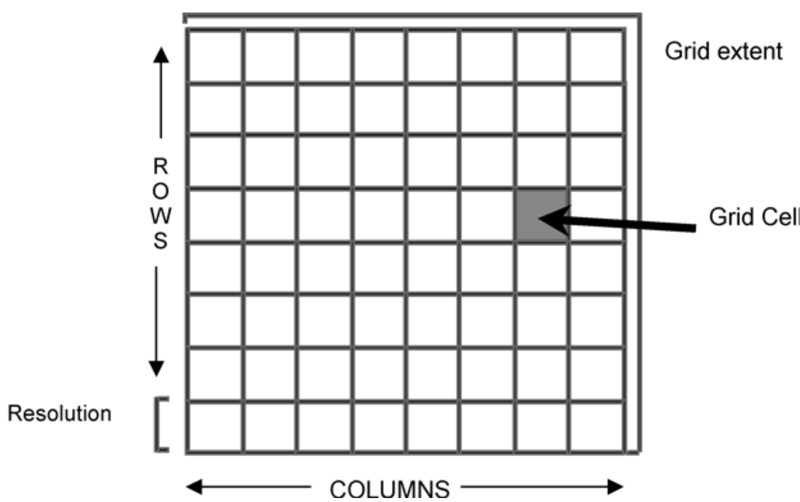


Figure 5.5: Generic structure for a grid.

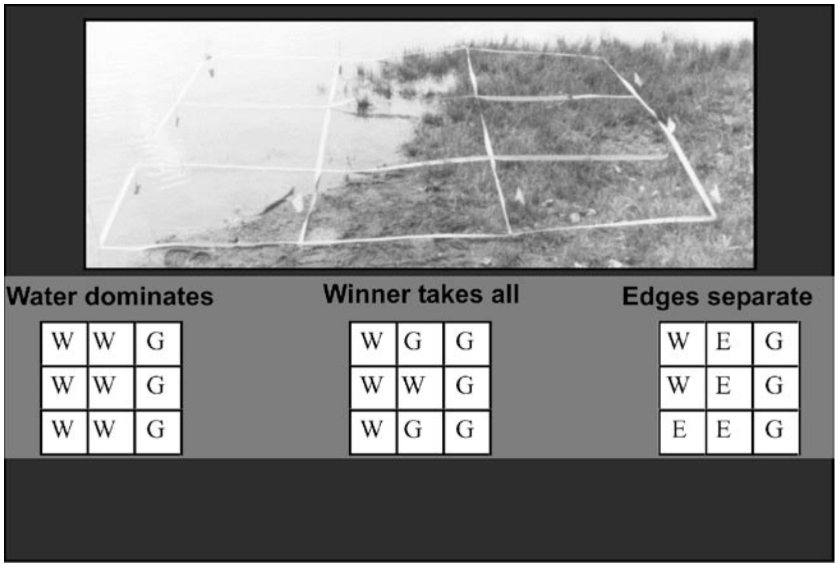


Figure 5.6: The mixed pixel problem.

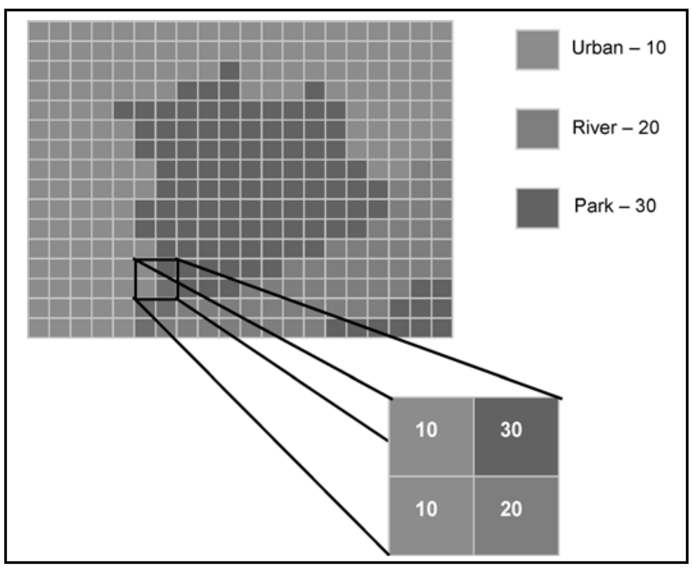


Figure 5.7: Attribute handling in raster data. Each pixel is assigned a single value which represents a real world object. Pixels can only hold numeric data; each pixel value in the raster here represents a feature class.

The relationship between cell size and the number of cells is expressed as the RESOLUTION of the raster.

A finer RESOLUTION gives a more accurate and better quality image.

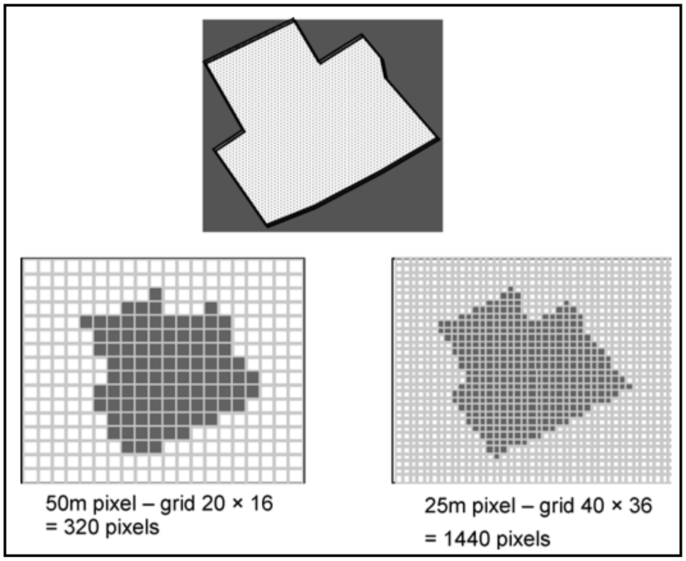


Figure 5.8: Effect of grid size on data in raster format.

VECTOR DATA FORMAT: A vector representation of the same diagonal line would record the position of the line by simply recording the coordinates of its starting and ending points. Each point would be expressed as two or three numbers (depending on whether the representation was 2D or 3D, often referred to as X,Y or X,Y,Z coordinates. The first number, X, is the distance between the point and the left side of the paper; Y, the distance between the point and the bottom of the paper; Z, the point's elevation above or below the paper. Joining the measured points forms the vector.

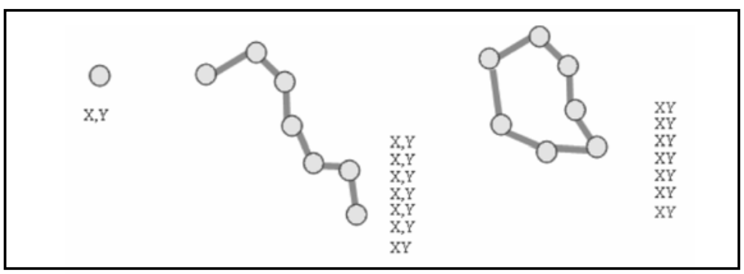


Figure 5.9: The vector data model is based around the storage of coordinate pairs.

A vector data model uses points stored by their real (earth) coordinates. Here lines and areas are built from sequences of points in order. Lines have a direction to the ordering of the points. Polygons can be built from points or lines. vectors can store information about topology. Manual digitizing is the best way of vector data input.

Vector files are most often used:

- Highly precise applications.
- When file sizes are important.

- When individual map features require analysis.
- When descriptive information must be stored.

Box 7: Comparison of raster and vector data formats.

Raster Model	Vector Model
Advantages	***Advantages***
• Simple data structure	• Compact data structure
• Easy and efficient overlaying	• Efficient for network analysis
• Compatible with RS imagery	• Efficient projection transformation
• High spatial variability is efficiently represented	• Accurate map output
• Simple for own programming	
• Same grid cells for several attributes	
Disadvantages	***Disadvantages***
• Inefficient use of computer storage	• Complex data structure
• Errors in perimeter, and shape	• Difficult overlay operations
• Difficult network analysis	• High spatial variability is inefficiently represented
• Inefficient projection transformations	• Not compatible with RS imagery
• Loss of information when using large cells Less accurate (although interactive) maps	

The method of representing geographic features by the basic graphical elements of points, lines and polygon is said to be the *vector method* or *vector data model*, and the data are called *vector data*. Related vector data are always organized by themes, which are also referred to as *layers* or *coverages*. Examples of themes: geodetic control, base map, soil, vegetation cover, land use, transportation, drainage and hydrology, political boundaries, land parcel and others.

For themes covering a very large geographic area, the data are always divided into *tiles* so that they can be managed more easily. A tile is the digital equivalent of an individual map in a map series. A tile is uniquely identified by a file name. A collection of themes of vector data covering the same geographic area and serving the common needs of a multitude of users constitutes the *spatial component* of a *geographical database*.

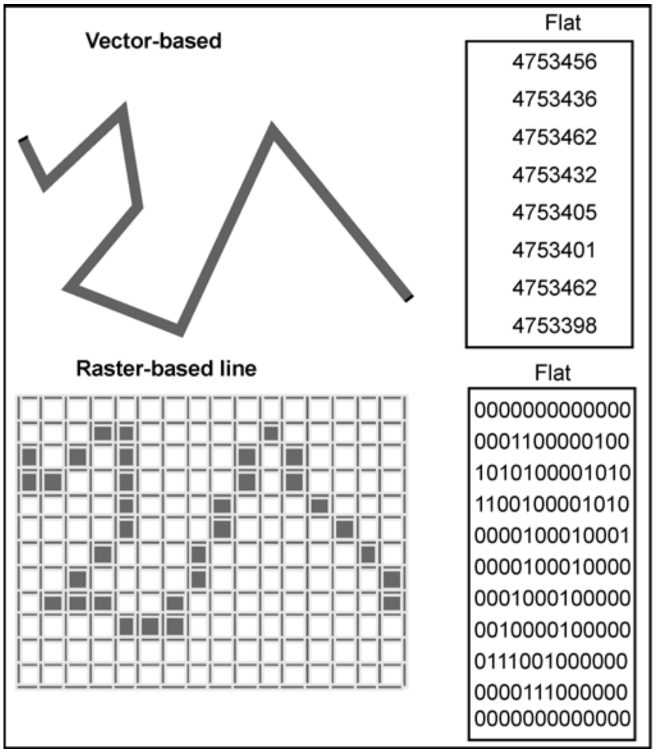


Figure 5.10: Rasters and vectors can be flat files...if they are simple.

The vector method of representing geographic features is based on the concept that these features can be identified as discrete entities or objects. This method is therefore based on the *object view of the real world* (Goodchild, 1992). The object view is the method of information organization in conventional mapping and cartography.

Graphical data captured by imaging devices in remote sensing and digital cartography (such as multi-spectral scanners, digital cameras and image scanners) are made up of a matrix of picture elements (pixels) of very fine resolution. Geographic features in such form of data can be visually recognized but not individually identified in the same way that geographic features are identified in the vector method. They are recognizable by differentiating their spectral or radiometric characteristics from pixels of adjacent features. For example, a lake can be visually recognized on a satellite image because the pixels forming it are darker than those of the surrounding features; but the pixels forming the lake are not identified as a single discrete geographic entity, *i.e.,* they remain individual pixels. Similarly, a highway can be visually recognized on the same satellite image because of its particular shape; but the pixels forming the highway do not constitute a single discrete geographic entity as in the case of vector data.

The method of representing geographic features by pixels is called the *raster method* or *raster data model,* and the data are described as *raster data.* The raster method is also called the *tessellation method.* A raster pixel is usually a square grid cell and a raster pixel represents the generalized characteristics of an area of specific size on or near the surface of the Earth. The actual ground size depicted by a pixel is dependent on the resolution of the data, which may range from smaller than a square meter to several square kilometers.

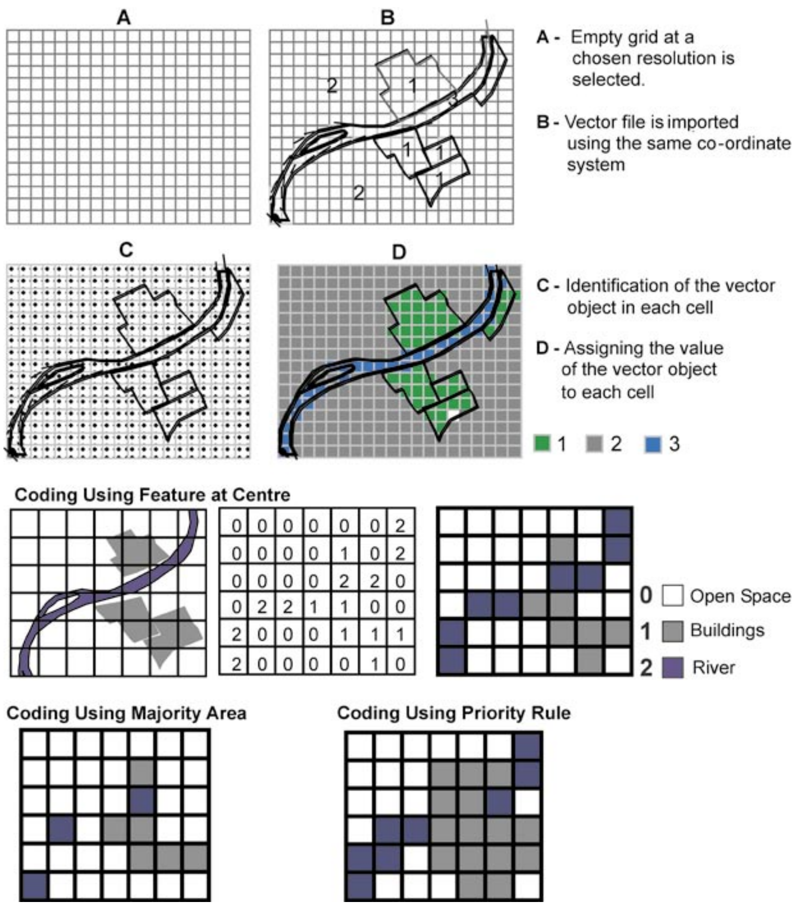


Figure 5.11: Raster data capture-rasterisation.

Raster data are organized by themes, which is also referred to as layers, for example, a raster geographic database may contain the following themes: bed rock geology, vegetation cover, land use, topography, hydrology, rainfall, temperature etc. Raster data covering a large geographic area are organized by *scenes* (for remote sensing images) of by *raster data files* (for images obtained by map scanning).

The raster method is based on the concept that geographic features are represented as surfaces, regions or segments. This method is therefore based on the *field view of the real world.* The field view is the method of information organization in image analysis systems in remote sensing and geographic information systems for resource-and environmental-oriented applications.

Choice between Raster and Vector

Arguments about which was better have been commonplace since the earliest systems were created. Raster databases are appealing because of simplicity of organization, speed of many operations, *e.g.,* overlay, buffers and especially appealing to the remote sensing community who are used to 'pixel' processing. On the other hand, there are many situations in which the raster approach may appear to sacrifice too much detail. Cartographers were appalled by the crude outlines of parcels that resulted in the 'pinking shear' effect of diagonal boundaries represented by grid cell edges. Similarly, surveyors were dismayed by the 'inaccuracy' caused by the cells when portraying linear features and points and situations in which the raster approach sacrificed too much detail. However, computing times for overlaying vector based information can be excessive and early polygon overlay routines were error-prone, expensive, and slow. But today, there are situations in which it is clear that one approach is more functional than the other, *e.g.,* using 'friction' layer to control width of buffer is only feasible in raster. For example, viewshed algorithms to find area visible from a point are feasible with elevation grids (raster DEMs), not with digitized contours.

An important current trend involves linking raster and vector systems, displaying vector data overlying a raster base. Raster data may be from a GIS file (perhaps a remotely sensed image) or from a plain scanned image file. Therefore, the question has evolved from 'Which is best?' to 'Under what conditions is which best and how can we have flexibility to use the most appropriate approaches on a case by case basis?'

Four issues to the discussions of raster versus vector:

- coordinate precision
- speed of analytical processing
- mass storage requirements
- characteristics of phenomena

Box 8: Choice between raster and vector data

	Raster	Vector
Data Collection	Rapid	Slow
Data Volume	Large	Small
Data Structure	Simple	Complex
Geometrical Accuracy	Low	High
Graphic Treatment	Average	Good
Area Analysis	Good	Average
Network Analysis	Poor	Good
Generalization	Simple	Complex

Data Capture

The functionality of GIS relies on the quality of data available. The true value of GIS can only be realized if the proper tools to collect spatial data and integrate them with attribute data are available.

Spatial information is presented in two ways: as vector data in the form of points, lines, and areas (polygons); or as grid data in the form of uniform, systematically organized cells.

Geometric presentations are commonly called digital maps. A digital map would be peculiar because it would comprise only numbers (digits).

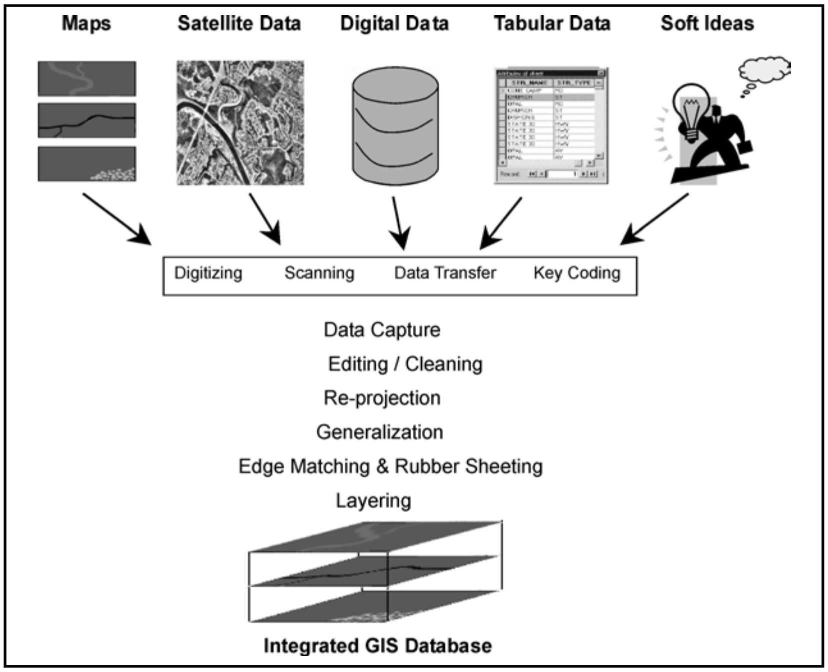


Figure 5.12: GIS data stream.

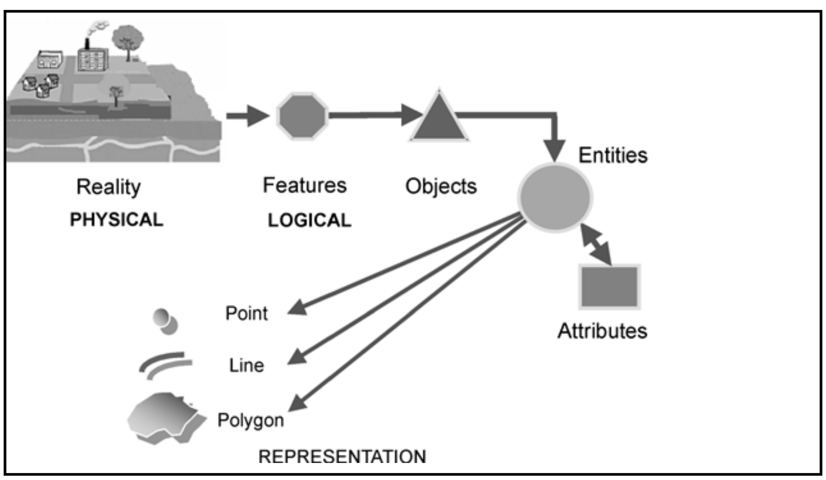


Figure 5.13: Modelling the real world data into GIS.

By their very nature, maps are analogue, whether they are drawn by hand or machine, or whether they appear on paper or displayed on a screen. GIS does not produce digital maps – it produces analogue maps from digital map data. Nonetheless, the term digital map is now so widely used that the distinction is well understood.

Box 9: Possible encoding methods for different data sources.

Data source	Analogue or Digital source	Possible encoding method	Examples
Tabular data	Analogue	• Keyboard • Text scanning	• List of school • Education board publications
Map data	Analogue	• Digitizing • Scanning	• Political maps • Historical maps
Aerial photo	Analogue	• Digitizing • Scanning	• Landuse maps • Water bodies
Tabular data	Digital	• Digital file transfer	• Census data
Satellite image	Digital	• Digital file transfer	• Landuse data

GIS can contain a wide variety of geographic data types originating from many diverse sources. From the perspective of creating geographic databases, it is convenient to classify raster and vector geographic data as primary and secondary (Table 1). Primary data sources are those collected specifically for use in GIS. Typical examples of primary GIS sources include raster IRS, SPOT and IKONOS Earth satellite images, and vector building survey measurements captured using a total survey station. Secondary sources are those that were originally captured for another purpose and need to be converted into a form suitable for use in a GIS project. Typical secondary sources include raster scanned colour aerial photographs of urban areas, and USGS and IGN paper maps that can be scanned and vectorized.

Box 10: General classification of geographic data.

Source	Raster	Vector
Primary	• Digital aerial photographs • Digital remote sensing images	• Survey measurements • GPS measurements
Secondary	✦ Scanned maps ✦ Photographs ✦ DEM generated from maps	✦ Topographic maps ✦ Toponymy databases (Place names)

So, primary geographic data sources are captured specifically for use in GIS by direct measurement. Secondary sources are those reused from earlier studies. Geographic data may be obtained in either digital or analog format. Analog data must always be digitized before being added to a geographic database. Depending on the format and characteristics of the digital data, considerable reformatting and restructuring may be required prior to import.

Here, we describe the data sources, techniques, and workflows involved in GIS data collection. The processes of data collection are also variously referred to as data capture, data automation, data conversion, data transfer, data translation, and digitizing. Although there are subtle differences between these terms, they essentially describe the same thing, *i.e.,* adding geographic data to a database. Data capture refers to direct entry; data transfer is the importing of existing digital data across the Internet. WANs, or LANs; or using CD ROMs, zip disks, or diskettes. Here we focus on the techniques of data collection and its importance to a real-world GIS implementation.

In the early days of GIS, when geographic data were very scarce, data collection was the main project task and it typically consumed the majority of the available resources. Data collection still remains a time consuming, tedious, and expensive process. Usually it accounts for 15 – 50 % of the total cost of a GIS project (Longley, et al., 2001). After an organization has completed basic data collection, their emphasis moves on to data maintenance. Over the multi-year lifetime of a GIS project, data maintenance often turns out to be a far more complex and expensive activity than initial data collection, This is because of the high volume of update transactions in many systems (for example, changes in land parcel ownership, maintenance work orders on a highway transport network etc.) and the need to manage multi-user access to operational databases.

DATA COLLECTION WORKFLOW

Data collection projects involve a series of sequential stages (Figure 5.14). The workflow commences with planning, followed by preparation, digitizing (here taken to mean a range of techniques such as table digitizing, survey entry, scanning, and photogrammetry) or transfer, editing and improvement and, finally, evaluation. Planning is obviously important to any project and data collection is no exception. It includes establishing user requirements, garnering resources (staff, hardware, and software) and developing a project plan. Preparation is especially important in data collection projects. It involves many tasks such as obtaining data, redrafting poor-quality map sources, editing scanned map images, and removing noise (unwanted data such as speckles on a scanned map image). Digitizing and transfer are the stages where the majority of the effort will be expended. It is naive to think that data collection is really just digitizing, when in fact it involves very much more. Editing and improvement follows digitizing / transfer. This covers many techniques designed to validate data, as well as correcting errors and improving quality. Evaluation, as the name suggests, is the process of identifying project successes and failures. Since all large data projects involve multiple stages, this workflow is iterative with earlier phases (especially a first, pilot, phase) helping to improve subsequent parts of the overall project.

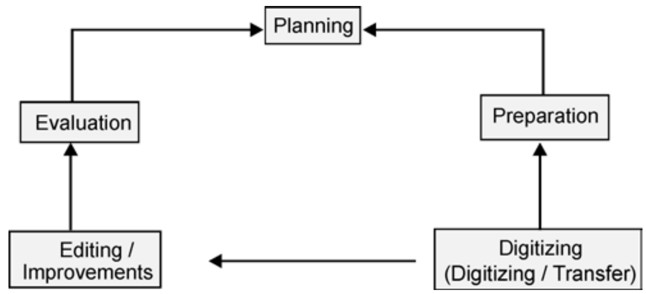


Figure 5.14: Stages in data collection.

Primary Geographic Data Capture

Primary geographic capture involves the direct measurement of objects. It can be in both raster and vector data capture methods.

Raster data capture

The most popular form of primary raster data capture is remote sensing. Broadly speaking, remote sensing is a technique used to derive information about the physical, chemical, and biological properties of objects without direct physical contact. Information is derived from measurements of the amount of electromagnetic radiation reflected, emitted, or scattered from objects. A variety of sensors, operating throughout the electromagnetic spectrum from visible to microwave wavelengths, are commonly employed to obtain measurements (Lillesand and Kiefer, 2004). Passive sensors are reliant on reflected solar radiation or emitted terrestrial radiation; active sensors (such as synthetic aperture radar) generate their own source of electromagnetic radiation. The platforms on which these instruments are mounted are similarly diverse. Although Earth-orbiting satellites and fixed-wing aircraft are by far the most common, helicopters, balloons, etc. is also employed. As used here, the term remote sensing subsumes the fields of satellite remote sensing and aerial photography.

From the GIS perspective, resolution is the key physical characteristic of remote sensing systems. There are three basic aspects to resolution: spatial, spectral, and temporal. All sensors need to trade off spatial, spectral, and temporal properties because of storage, processing, and bandwidth considerations.

Spatial resolution refers to the size of object that can be resolved and the most usual measure is the pixel size. Satellite remote sensing systems typically provide data with pixel sizes in the range 1 meter – 1 km. The cameras used for capturing aerial photographs usually range from 0.1 meter – 5 meters. Image (scene) sizes vary quite widely between sensors – typical ranges include 1000 by 1000 to 3000 by 3000 pixels. The total coverage of remote sensing images is usually in the range 10 by 10 – 200 by 200 km.

Spectral resolution refers to the parts of the electromagnetic spectrum that are measured. Since different objects emit and reflect different types and amounts of radiation, selecting which part of the electromagnetic spectrum to measure is critical for each application area.

The spectral signatures of water, green vegetation, and dry soil are all different. Remote sensing systems may capture data in one part of the spectrum (referred to as a single band) or simultaneously from several parts (multi-band or multi-spectral). The radiation values are usually normalized and resampled to give a range of integers from 0 – 255 for each band, for each pixel, in each image.

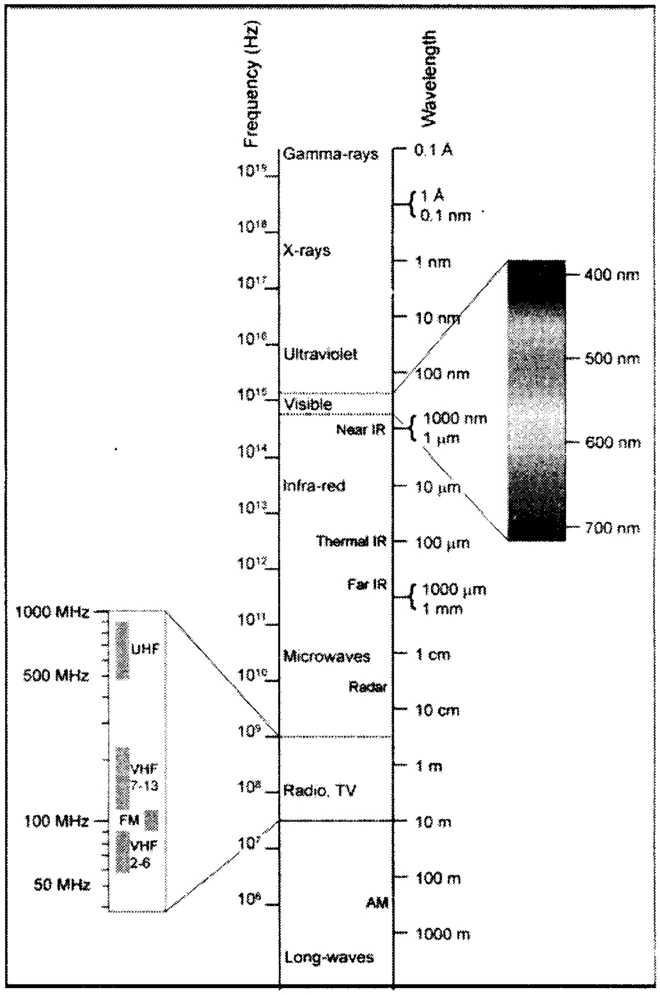


Figure. 5.15: Electromagnetic spectrum.

Temporal resolution, or repeat cycle, describes the frequency with which images are collected for the same area. There are essentially two types of commercial remote sensing satellite: Earth orbiting and geostationary. Earth orbiting satellites collect information about different parts of the Earth surface at regular intervals. To maximize utility, orbits are typically polar, at a fixed altitude and speed, and are Sun synchronous. The Indian Satellite Series (IRS), for example, passes virtually over the poles at an altitude of 904 with a repetitive coverage of 22 days. The satellite carries three sensors – a single band panchromatic sensor measuring in the visible part of the EMR at a resolution of 5.8 meters. Another sensor is

LISS III (Linear Imaging and Self Scanning Sensor) measuring green, red and near infra red radiation at 23.5 meters resolution and WiFS (Wide Field Sensor) measuring red and near infra red radiation at 188 meters resolution.

Aerial photography also has importance, especially in medium to large-scale GIS projects. Although the data products resulting from remote sensing satellites and aerial photography systems are technically very similar (*i.e.,* they are both images) there are some significant differences in the way data are captured and can be interpreted. The most notable difference is that aerial photographs are normally collected using analog optical cameras (although digital cameras are becoming more widely used) and then later rasterized, usually by scanning a film negative.

The quality of the optics of the camera and the mechanics of the scanning process both affect the spatial and spectral characteristics of the resulting images. Most aerial photographs are collected on an ad hoc basis using cameras mounted in airplanes flying at low altitudes (3000 – 10000 meters) and are either panchromatic (black and white) or colour, although multi-spectral cameras/sensors operating in the non-visible parts of the electromagnetic spectrum are also used. Aerial photographs are very suitable for detailed surveying and mapping projects.

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

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

On the other hand, the spatial resolution of commercial satellites is too coarse for many large area projects and the data collection capability of many sensors is restricted by cloud cover. The data volumes from both satellites and aerial cameras can be very large and create storage and processing problems for all but the most modern systems. The cost of data can also be prohibitive for a single project or organization.

VECTOR DATA CAPTURE

Primary vector data capture is a major source of geographic data. The two main branches of vector data capture are ground surveying and GPS.

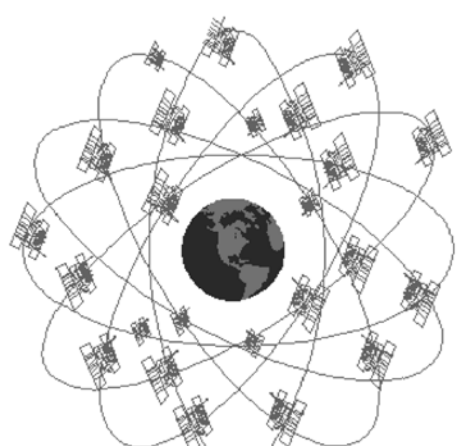
SURVEYING: Ground surveying is based on the principle that the 3D location of any point can be determined by measuring angles and distances from other known points. Surveys begin from a benchmark point. If the coordinate system of this point is known, all subsequent points can be collected in this coordinate system. If it is unknown then the survey will use a local or relative coordinate system.

Since all survey points are obtained from survey measurements their locations are always relative to other points. Any measurement errors need to be apportioned between multiple points in a survey. For example, when surveying a field boundary, if the last and first points are not identical in survey terms (within the tolerance employed in the survey) then errors need to be apportioned between all points that define the boundary. As new measurements are obtained these may change the locations of points. For this reason it is necessary to store both the measurements and the points inside a GIS database until the survey is complete.

Traditionally, surveyors used equipment like transits and theodolites to measure angles, and tapes and chains to measure distances. Today these have been replaced by electro-optical devices called total stations that can measure both angles and distances to an accuracy of 1 millimeter. Total stations automatically log data and the most sophisticated can create vector point, line, and polygon objects in the field, thus providing direct validation.

The basic principles of surveying have changed very little in the past 100 years. Ground survey is a very time-consuming and expensive activity, but it is still the best way to obtain highly accurate point locational data. Surveying is typically used for capturing buildings, land and property boundaries, and other objects that need to be located accurately. It is also used to obtain reference marks for other data capture methods. For example, large-scale aerial photographs and satellite images are frequently georeferenced using points obtained from ground survey.

GPS: The Global Position System (GPS) is a collection of 27 NAVSTAR satellites orbiting the Earth at a height of 12,500 miles, five monitoring stations, and individual receivers. The GPS was originally funded by the US Department of Defence, and for many years military users had access to only the most accurately data. Fortunately this selective availability was removed in May 2000, so that now civilian and military users can fix the x, y, z location of objects relatively easily to an accuracy of better than 10 m with standard equipment.



21 satellites with three operational spares, 6 orbital planes,
55 degree inclinations, 20,200 kilometer, 12 hour orbit.

Figure 5.16: GPS.

The GPS is a network of satellites, monitoring stations, and inexpensive receivers used for primary GIS data capture. In many respects GPS has revolutionized primary data capture, especially since the development of Differential GPS (Box 11), the removal of selective availability, and the creation of low-cost, low-power receivers. Today units costing less than $100 can easily provide locational data at better than 10 m accuracy. One of the drawbacks of GPS, however, is that it is necessary to have three or more satellites in unobstructed view in order to collect measurements. This can especially be a problem in forests and urban areas with tall buildings. GPS is very useful for recording ground control points for other data capture projects, for locating objects that move (for example, combine harvesters, tanks, cars, and shipping containers), and for direct capture of the locations of many types of objects such as utility assets, buildings, geological deposits, and stream sample points.

Box 11: Principles of GPS

GPS works according to a simple principle-the length of time it takes a signal to travel from a satellite to a receiver on the ground. The GPS satellites constantly transmit a coded radio signal that indicates their exact position in space and time. The receiver measures how long it takes the signal to travel from the satellites. By measuring the distance from three or more satellites, the location of the receiver can be obtained by triangulation. If a signal can be obtained from a fourth satellite, then the elevation of the receiver can also be determined.

Although standard GPS receivers can provide locations at accuracies of 5–10 m, it is important to understand that there are several possible sources of error inherent in these locations. Some of the errors are random in nature, while others are systematic and can therefore be corrected. Errors arise from signal degradation due to atmospheric effects, minor variations in the location of the satellites, inaccuracies in the timing clocks, errors in receivers, and variations in the reflection of signals from local objects.

A number of techniques are available to improve the accuracy of GPS measurements. Many GPS receivers perform averaging of measurements to improve apparent accuracy. Others snap measurements to map features. So, for example, in-car navigation systems snap the location of the vehicle to a road centerline.

The accuracy of measurements can also be improved by using Differential GPS. This technique uses two receivers. One is fixed and the other is used to collect measurements. If the location of the fixed (base) receiver is known accurately, comparing the exact location with the location reported by GPS will provide an estimate of error. This error can be used to correct measurements obtained from the roving receiver provided that it is within about 300 km. In some countries, the differential correction information is broadcast freely over airwaves and can be received using a standard radio receiver. Differential GPS can improve accuracy to allow locations to be determined to better than 1 meter.

Strictly speaking, the term GPS refers only to the US Department of Defence System. GLONASS is the Russian version of GPS offering similar coverage and accuracy; Galileo is the European Union's proposed equivalent.

SECONDARY GEOGRAPHIC DATA CAPTURE

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

Box 12: *Data input by a scanner*

There are three different types of scanner generally used for data entry.

Flat-bed scanner – A common PC peripheral, it is small and inaccurate.

Rotating drum scanner – It is expensive and slow but accurate.

Large-format feed scanner – most suitable for capturing data in GIS. It is quicker, cheaper and accurate.

All scanners work on the same principles, where a scanner has a light source, a background (source document) and a lens. During scanning the absence or presence of light is detected as one of the three components moves past the other two.

Precautions for map scanning in GIS:

OUTPUT QUALITY: The output quality of map is very crucial in GIS, it needs to be sharp and clear. Setting up the brightness and contrast levels can enhance the quality of images. In some cases *gamma correction* (a method which looks at histogram of the image and places points strategically along the histogram to isolate data types) or *filtering methods* (selectively removal of noise disturbance).

RESOLUTION: This is the density of the raster image produced by the scanning process. The resolution of scanners is usually measured in dots per inch (dpi) as a linear measurement along the scan line. Commonly, 150 dpi for text, 300 dpi for line maps and higher dpi scanning is done for high quality ortho-photos.

ACCURACY: The accuracy of the scanned image is important if the image needs to be used in GIS. It needs to fit for its intended use in terms of its physical and cartographic quality. That is why cleaning of scanned map is essential before using it in GIS because stains and folding marks in maps can affect the map accuracy.

GEOREFERENCING: The output of a map from scanner needs to be correctly referenced according to the coordinate system used in GIS. Generally, this process is controlled using linear transformation from the row an column number. Distortion across scanned image can create problem if the scanned image is of low quality.

VECTORIZATION: The output from scanned maps are often used to generate vector data. This involves, automatic or user controlled raster to vector conversion. Here the resolution of scanned map is very important because it affects the generalization of features in the map.

Figure 5.17: Using a scanner.

RASTER DATA CAPTURE USING SCANNERS: A scanner is a device that converts hardcopy analog media into digital images by scanning successive lines across a map or document and recording the amount of light reflected from a local data source. The differences in reflected light are normally scaled into bilevel black and white (1 bit per pixel), or multiple grey levels (8, 16, or 32 bits). Colour scanners output data into 8-bit red, green, and blue colour bands. The spatial resolution of scanners varies widely from as little as 100 dpi (4 dots per millimeter) to 1800 dpi (72 dots per millimeter) and beyond. Most GIS scanning is in the range 400 – 1000 dpi (16-40 dots per millimeter). Depending on the type of scanner and the resolution required, it can take from 30 seconds to 30 minutes or more to scan a map. Scanned maps and documents are used extensively in GIS as background maps and data stores.

There are three reasons to scan hardcopy media for use in GIS:

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

An 8 bit (256 grey levels) 400 dpi (16 dots per millimeter) scanner is a good choice for scanning maps for use as a background GIS reference layer. For a colour aerial photograph that is to be used for subsequent photo-interpretation and analysis, a colour (8 bit for each of three bands) 1000 dpi (40 dots per millimeter) scanner is more appropriate. The quality of data output from a scanner is determined by the nature of the original source material, the quality of the scanning device, and the type of preparation prior to scanning (*e.g.,* redrafting key features or removing unwanted marks will improve output quality).

Vector Data Capture

Secondary vector data capture involves digitizing vector objects from maps and other geographic data sources. The most popular methods are manual digitizing, heads-up digitizing and vectorization, photogrammetry, and COGO data entry.

MANUAL DIGITIZING: Manually operated digitizers are much the simplest, cheapest, and most commonly used means of capturing vector objects from hardcopy maps. Digitizers come in several designs, sizes, and shapes. They operate on the principle that it is possible to detect the location of a cursor or puck passed over a table inlaid with a fine mesh of wires. Accuracies typically range from 0.003 inch (0.075 millimeter) to 0.010 inch (0.25 millimeter), Small digitizing tablets up to 12 by 24 inches (30 by 60 centimeters) are used for small tasks, but bigger (typically 50 by 32 inches (120 by 80 centimeters) freestanding tab1e digitizers are preferred for larger tasks. Both types of digitizer usually have cursors with cross hairs mounted in glass and buttons to control capture.

Vertices defining point, line, and po1ygon objects are captured using manual or stream digitizing methods. Manual digitizing involves placing the center point of the cursor cross hairs at the location for each object vertex and then clicking a button on the cursor to record the location of the vertex. Stream mode digitizing partially automates this process by instructing the digitizer control software automatically to collect vertices every time a distance or time threshold is crossed (*e.g.,* every 0.02 inch (0.5 millimeter) or 0.25 second). Stream-mode digitizing is a much faster method, but it typically produces larger files with many redundant coordinates.

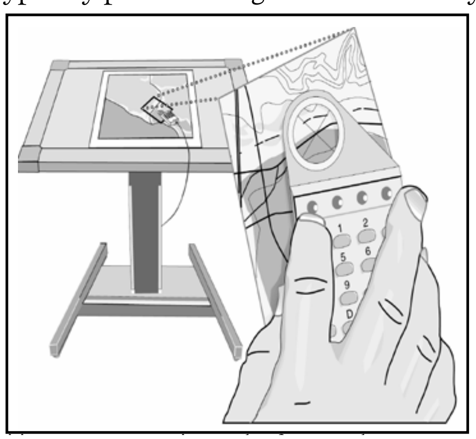


Figure 5.18: Digitizing table.

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

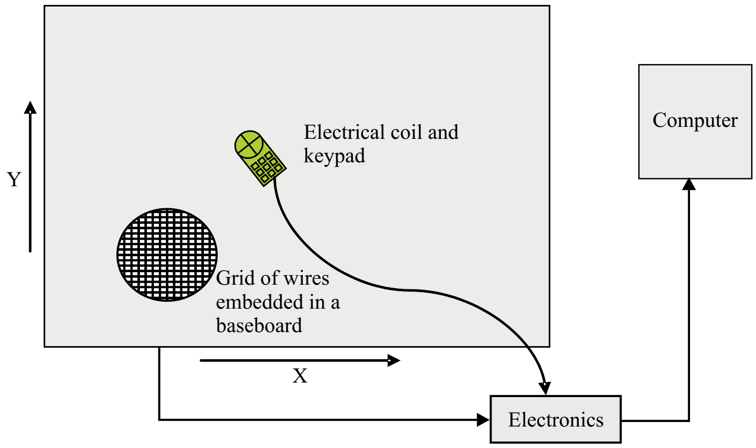


Figure 5.19: The basic components of a digitising tablet.

Terms	Example	Description
Arc		Line feature: a node at each end; vertices at each change of direction.
Node		Endpoint of an arc (also found at intersections between lines).
Vertex		A point on an arc that signals a change of direction.
Pseudo Node		On an (island) arc that connects to itself or where an attribute changes or on a long arc
Dangling Node		Arc endpoint that is not connected.
Label Point		Identifies a point feature or polygon.
Tic		Geographic control point; features can be registered to the same coordinate system.

Figure 5.20: Digitizing terms.

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

A faster and more consistent approach is to use software to perform automated vectorization in either batch or interactive mode. Batch vectorization takes an entire raster file and converts it to vector objects in a single operation. vector objects are created using software algorithms that build simple (spaghetti) line strings from the original pixel values (Figure 10.10). Depending on the size of the raster file, it typically takes 1 – 100 minutes to complete vectorization.

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

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

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

To obtain true georeferenced coordinates from a model it is necessary to georeference photographs using control points. Control points can be defined by ground surveyor nowadays more usually with GPS.

Measurements are captured from overlapping pairs of photographs using stereoplotters. These build a model and allow 3D measurements to be captured, edited, stored, and plotted. Stereoplotters have undergone three major generations of development: analog (optical), analytic, and digital. Mechanical analog devices are seldom used today, whereas analytical (combined mechanical and digital) and digital (entirely computer based) are much more common. It is likely that digital (softcopy) photogrammetry will eventually rep1ace mechanical devices entirely.

The options for extracting vector objects from 3D models are directly analogous to those available for manual digitizing as described above: namely batch, interactive, and manual. The obvious difference, however, is that there is a requirement for capturing z (elevation) values. In the case of manual and interactive methods, this requires a 3D cursor.

Photogrammetric techniques are particularly suitable for highly accurate capture of contours, digital elevation models, and almost any type of object that can be identified on an aerial photograph or image. One type of popular specialist photogrammetry product is the orthophotograph. Orthophotographs result from using a DEM to correct distortions in an aerial photograph derived from varying land elevation. They have become popular because of their relatively low cost of creation (when compared with topographic maps) and ease of interpretation as base maps. They can also be used as accurate data sources for heads-up digitizing.

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

COGO DATA ENTRY: COGO, a contraction of the term coordinate geometry, is a methodology for capturing and representing geographic data. COGO uses survey style bearings and distances to define each part of an object. The COGO system is widely used in North America to represent land records and property parcels (also called lots). Coordinates can be obtained from COGO measurements by geometric transformation (*i.e.,* bearings and distances are converted into X, Y coordinates). Although COGO data obtained as part of a primary data capture activity are used in some projects, it is more often the case that secondary measurements are captured from hard copy maps and documents. Source data may be in the form of legal descriptions, records of survey, tract (housing estate) maps, or similar documents. COGO data are very precise measurements and are often regarded as the only legally acceptable definition of land parcels.

Obtaining Data from External Sources (Data Transfer)

One major decision that needs to be faced at the start of a GIS project is whether to build or buy a database. All the preceding discussion has been concerned with techniques for building databases from primary and secondary sources. This section focuses on how to import or transfer data captured by others. Some of these data are freely available, but many of them are sold as a commodity from a variety of outlets including, increasingly, Internet sites.

There are many sources and types of geographic data. The characteristics and availability of datasets are constantly changing so those seeking an up-to-date list should consult one of the good online sources. The best way to find geographic data is to search the Internet using one of the specialist geographic search engines such as the US NSDI Clearinghouse or the Geography Network. One of the good things about data standards is that there are many to choose from.

An interesting new trend initiated by the Geography Network Project is the idea of providing data online in ready-to-use GIS formats. The Geography Network is global collection of data users and providers connected by the Internet. Information about available data sources can be found by consulting the Geography Network Web site (www. GeographyNetwork.com). Once a useful data source has been located, the actual data can be streamed directly into a browser or desktop GIS. Much of the content on the Geography Network is accessible without charge, but additional commercial content is also provided and maintained by its owners. This information is accessible in the same way as free content, but every time a map is viewed, an online service utilized (for example a retail site suitability or flood risk mapping application), or a dataset downloaded, a charge is recorded by the Geography Network e-commerce system. The Geography Network management organization is responsible for maintaining the e-commerce system and for billing users and paying providers. A critical requirement for providing online geographic data indexing, searching, access and download is good quality metadata.

Geographic Data Formats

One of the biggest problems with data obtained from external sources is that they can be encoded in many different formats. There are so many different geographic data formats because no single format is appropriate for all tasks and applications. It is not possible to design a format that supports. The many different formats have evolved in response to diverse user requirements.

Given the high cost of creating databases many people have asked for tools to move data between systems and to re-use data through open application programming interfaces (APIs). In the former case, the approach has been to develop software that is able to translate data (Figure 10.12), either by a direct read into memory, or via an intermediate file format. In the latter case, software developers have created open interfaces to allow access to data.

Many GIS software systems are now able to read directly Auto CAD DWG and DXF, Microstation DGN, and Shapefile, VPF, and many image formats. Unfortunately, direct read support can only easily be provided for relatively simple product-oriented formats. Complex formats, such as SDTS and UKNTF, were designed for exchange purposes and require more advanced processing before they can be viewed (*e.g.,* multi-pass read and feature assembly from several parts).

More than 25 organizations are involved in the standardization of various aspects of geographic data and geoprocessing. Several of these are country and domain specific. At the global level, ISO (the International Standards Organization) is responsible for coordinating efforts through the work of technical committees TC 211 and 287. In Europe, CEN (Commission European Normalization) is engaged in geographic standardization.

Having obtained a potentially useful source of geographic information the next task is to import it into a GIS database. If the data are already in the native format of the target GIS software system, or the software has a direct read capability for the format in question, then this is a relatively straightforward task. If the data are not compatible with the target GIS software then the alternatives are to ask the data supplier to convert the data to a compatible format, or to use a third-party translation software system, such as the Feature Manipulation Engine from Safe Software lists over 60 Supported geo-graphic data formats) to convert the data. Geographic data translation software must address both syntactic and semantic translation issues. Syntactic translation involves converting specific digital symbols (letters and numbers) between systems. Semantic translation is concerned with converting the meaning inherent in geographic information. While the former is relatively simple to encode and decode, the latter is much more difficult and has seldom met with much success to date.

Box 13: Some examples of geographic data formats

Vector	Raster (Image)
Automated Mapping System (AMS)	Arc Digitized Raster Graphics (ADRG)
ESRI Coverage	Band Interleaved by line (BIL)
Computer Graphics Metafile (CGM)	Band Interleaved by Pixel (BIP)
Digital Feature Analysis Data (DFAD)	Band Sequential (BSQ)
Encapsulated Postscript (EPS)	Windows Bitmap (BMP)
Microstation drawing file format (DGN)	Device-Independent Bitmap (DIB)
Dual Independent Map Encoding (DIME)	Compressed Arc Digitized Raster Graphics (CADRG)
Digital line Graph (DLG)	Controlled Image Base (CIB)
AutoCAD Drawing Exchange Format (DXF)	Digital Terrain Elevation Data (DTED)
AutoCAD Drawing (DWG)	ERMapper
MapBase file (ETAK)	Graphics Interchange Format (GIF)
ESRI Geodatabase	ERDAS IMAGINE (IMG)
Land Use and Land Cover Data (GlRAS)	ERDAS 7.5 (GIS)
Interactive Graphic Design Software (IGDS)	ESRI GRID file (GRID)
Initial Graphics Exchange Standard (IGES)	JPEG File Interchange Format (JFIF)
Map Information Assembly Display System (MIADS)	Multi-resolution Seamless Image Database (MrSID)
MOSS Export File (MOSS)	Tag Image File Format (TIFF; GeoTIFF)
TIGER/line file: Topologically Integrated Geographic Encoding and Referencing (TIGER)	Portable Network Graphics (PNG)
Spatial Data Transfer Standard/Topological Vector Profile (SDTS/TVP)	

Capturing Attribute Data

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

An essential requirement far separate data entry is a common identifier (also called a key) that can be used to relate object geometry' and attributes together following data capture.

Metadata are a special type of non-geometric data that are increasingly being collected. Some metadata are derived automatically by the GIS software system (for example, length and area, extent of data layer, and count of features), but some must be explicitly collected (for example, owner name, quality estimate, and original source). Explicitly collected metadata can be entered in the same way as other attributes as described above.

Managing a Data Capture Project

The management of data capture projects is of critical importance and because there are several unique issues. That said, most of the general principles for any GIS project apply to data capture: the need for a clearly articulated plan, adequate resources, appropriate funding, and sufficient time. In any data capture project there is a fundamental trade-off between quality, speed, and price. Capturing high quality data quickly is possible, but it is very expensive. If price is a key consideration then lower quality data can be captured over a longer period.

A key decision facing managers of data capture projects is whether to pursue a strategy of incremental capture or 'Blitzkrieg' – that is, to capture all data as rapidly as possible. Incremental data capture involves breaking the data capture project into small manageable sub-projects. This allows data capture to be undertaken with lower annual resource and funding levels (although total project resource requirements may be larger).

Whichever approach is preferred, a pilot project carried out on part of the study area and a selection of the data types can prove to be invaluable. A further important decision is whether data capture is to use in-house or external resources. Three factors influencing this decision are: cost – schedule, quality, and long-term ramifications. Specialist external data capture agencies can often perform work faster, cheaper, with higher quality than in-house staff, but because of the need for real cash to pay external agencies this may not be possible. In the short term, project costs, quality, and time are the main considerations.

Data Editing

The process of data encoding is so complex that an error free data input is next to impossible. Data may have errors derived from the original source data or may be during encoding process. There may be errors in co-ordinate data as well as inaccuracies and uncertainness in attribute data. However, good practice in GIS involves continuous management of data quality, and it is normal at this stage in the data stream to make special provision for the identification and correction of errors. It is better to intercept errors before they contaminate the GIS database and go on to infect (propagate) the higher levels of information that are generated. The process is known as data editing or 'cleaning'. Data editing includes – detection and correction of errors; re-projection, transformation and generalization; and edge matching and rubber sheeting.

Detecting and Correcting Errors: Errors in input data may derive from three main sources: errors in the source data; errors introduced during encoding; and errors propagated during data transfer and conversion. Errors in source data may be difficult to identify. For example, there may be subtle errors in a paper map source used for digitizing because of the methods used by particular surveyors, or there may be printing errors in paper based records used as source data. During encoding a range of errors can be introduced. During keyboard encoding it is easy for an operator to make a typing mistake; during digitizing an operator may encode the wrong line; and folds and stains can easily be scanned and mistaken for real geographical features. During data transfer, conversion of data between different formats required by different packages may lead to a loss of data. Errors in attribute data are relatively easy to spot and may be identified using manual comparison with the original data. For example, a forest area can be wrongly identified as agricultural land or if a railway line has been erroneously digitized as a road, then the attribute database may be corrected accordingly. various methods, in addition to manual comparison, exist for the correction of attribute errors.

Errors in spatial data are often more difficult to identify and correct than errors in attribute data. These errors take many forms, depending on the data model being used (vector or raster) and the method of data capture. Chrisman (1997) suggests that certain types of error can help to identify other problems with encoded data. For example, in an area data layer 'dead-end nodes' might indicate missing lines, overshoots or undershoots. The user can look for these features to direct editing rather than having to examine the whole map. Most GIS packages will provide a suite of editing tools for the identification and removal of errors in vector data.

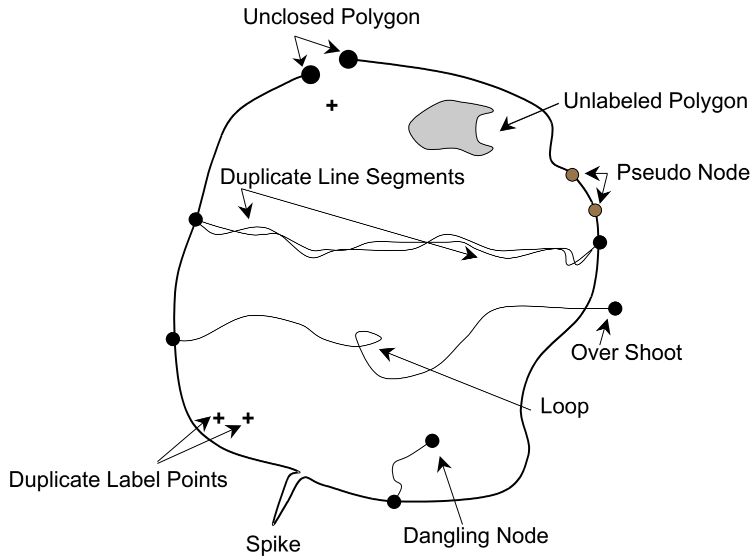


Figure 5.21: Examples of spatial errors.

Box 14: *Common spatial errors*

Error	Description
Missing entities	Missing points, lines or boundary segments.
Duplicate entities	Points, lines or boundary segments that have been digitized twice.
Mislocated entities	Points, lines or boundary segments that have been digitized at wrong place.
Missing labels	Unidentified polygons.
Duplicate labels	Two or more identification labels for same polygon.
Artifacts of digitizing	Undershoot, overshoot, loops, spikes etc.
Noise	Irrelevant data entry during digitizing or scanning.

Corrections can be done interactively by the operator 'on-screen', or automatically by the GIS software. However, visual comparison of the digitized data against the source document, either on paper or on the computer screen, is a good starting point. This will reveal obvious omissions, duplications and erroneous additions. Systematic errors such as overshoots in digitized lines can be corrected automatically by some digitizing software, and it is important for data to be absolutely correct if topology is to be created for a vector data set. Automatic corrections can save many hours of work but need to be used with care as incorrectly specified tolerances may miss some errors or correct 'errors' that never existed in the first place.

Errors will also be present in raster data. In common with vector data, missing entities and noise are particular problems. Data for some areas may be difficult to collect, owing to environmental or cultural obstacles. Similarly, it may be difficult to get clear images of vegetation cover in an area during a rainy season using certain sensors. Noise may be inadvertently added to the data, either when they were first collected or during processing. This noise often shows up as scattered pixels whose attributes do not conform to those of neighbouring pixels. For example, an individual pixel representing water may be seen in a large area of forest. While this may be correct, it could also be the result of noise and needs to be checked. This form of error may be removed by filtering. Filtering involves passing a filter (a small grid of pixels specified by the user-often a 3 × 3 pixel square is used) over the noisy data set and recalculating the value of the central (target) pixel as a function of all the pixel values within the filter. This technique needs to be used with care as genuine features in the data can be lost if too large a filler is used.

RE-PROJECTION, TRANSFORMATION AND GENERALIZATION: Once spatial and attribute data have been encoded and edited, it may be necessary to process the data geometrically in order to provide a common framework of reference. The scale and resolution of the source data are also important and need to be taken into account when combining data from a range of sources into a final integrated database. This section briefly considers the role of re-projection, transformation and generalization in the data stream.

Data derived from maps drawn on different projections will need to be converted to a common projection system before they can be combined or analyzed. If not re-projected, data derived from a source map drawn using one projection will not plot in the same location as data derived from another source map using a different projection system. For example, if a coastline is digitized from a navigation chart drawn in the Mercator projection (cylindrical) and the internal state boundaries of the country are digitized from a map drawn using the Alber's Equal Area (conic) projection, then the state boundaries along the coast will not plot directly on top of the coastline. In this case they will be offset and will need to be re-projected into a common projection system before being combined.

Data derived from different sources may also be referenced using different co-ordinate systems.

The grid systems used may have different origins, different units of measurement or different orientation. If so, it will be necessary to transform the co-ordinates of each of the input data sets onto a common grid system. This is quite easily done and involves linear mathematical transformations.

Some of the other methods commonly used are:

- *Translation and scaling:* One data set may be referenced in 1-metre co-ordinates while another is referenced in 10-metre co-ordinates. If a common grid system of 1-metre coordinates is required, then this is a simply a case of multiplying the co-ordinates in the 10metre data set by a factor of 10.
- *Creating a common origin:* If two data sets use the same co-ordinate resolution but do not share the same origin, then the origin of one of the data sets may be shifted in line with the other simply by adding the difference between the two origins (dx,dy) to its co-ordinates.
- *Rotation:* Map co-ordinates may be rotated using simple trigonometry to fit one or more data sets onto a grid of common orientation.

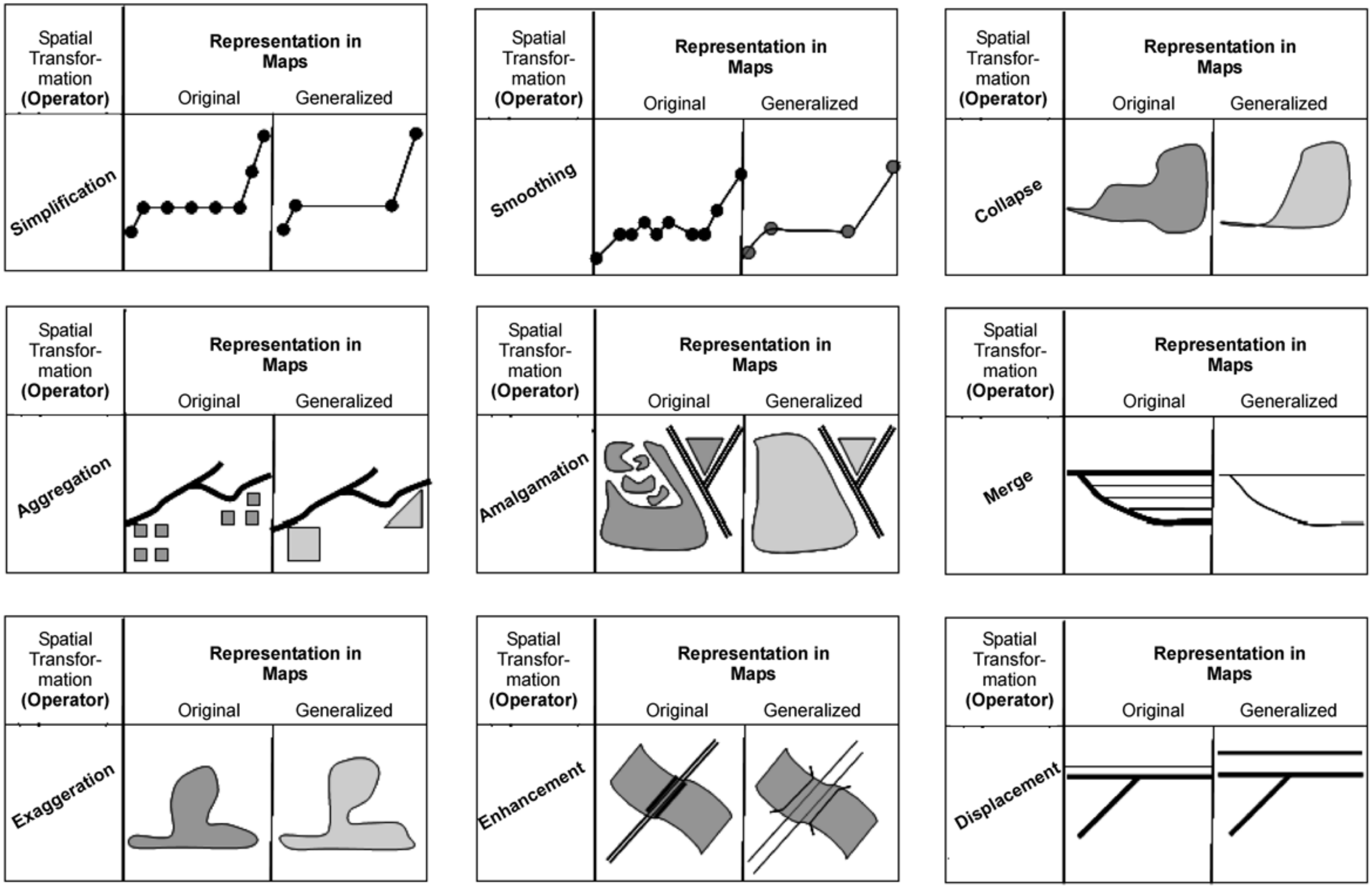


Figure 5.22: Different forms of generalization (Based on McMaster and Shea, 1992).

Data may be derived from maps of different scales. The accuracy of the output from a GIS analysis can only be as good as the worst input data. Thus, if source maps of widely differing scales are to be used together, data derived from larger-scale mapping should be generalized to be comparable with the data derived from smaller-scale maps. This will also save processing time and disk space by avoiding the storage of unnecessary detail. Data derived from large-scale sources can be generalized once they have been input to the GIS. Routines exist in most vector GIS packages for weeding out unnecessary points from digitized lines such that the basic shape of the line is preserved. The simplest techniques for generalization delete points along a line at a fixed interval (for example, every third point).

These techniques have the disadvantage that the shape of features may not be preserved. Most other methods are based on the Douglas-Peucker algorithm (Douglas and Peucker, 1973). This involves the following stages:

i. Joining the start and end nodes of a line with a straight line.
ii. Examining the perpendicular distance from this straight line to individual vertices along the digitized line.
iii. Discarding points within a certain threshold distance of the straight line.
iv. Moving the straight line to join the start node with the point on the digitized line that was the greatest distance away from the straight line.
v. Repeating the process until there are no points left which are closer than the threshold distance.

When it is necessary to generalize raster data the most common method employed is to aggregate or amalgamate cells with the same attribute values. This approach results in a loss of detail which is often very severe. A more sympathetic approach is to use a filtering algorithm. If the main motivation for generalization is to save storage space, then, rather than resorting to one of the two techniques outlined above, it may be better to use an appropriate data compaction technique as this will result in a volume reduction without any loss in detail.

EDGE MATCHING AND RUBBER SHEETING: When a study area extends across two or more map sheets small differences or mismatches between adjacent map sheets may need to be resolved. Normally, each map sheet would be digitized separately and then the adjacent sheets joined after editing, re-projection, transformation and generalization. The joining process is known as edge matching and involves three basic steps.

First, mismatches at sheet boundaries must be resolved. Commonly, lines and polygon boundaries that straddle the edges of adjacent map sheets do not meet up when the maps are joined together. These must be joined together to complete features and ensure topologically correct data. More serious problems can occur when classification methods vary between map sheets. For example, different soil scientist may interpret the pattern and type of soils differently, leading to serious differences on adjacent map sheets. This may require quite radical reclassification and reinterpretation to attempt a smooth join between sheets. This problem may also be seen in maps derived from multiple satellite images. If the satellite images were taken at different times of the day and under different weather and seasonal conditions then the classification of the composite image may produce artificial differences where images meet. These can be seen as clear straight lines at the sheet edges.

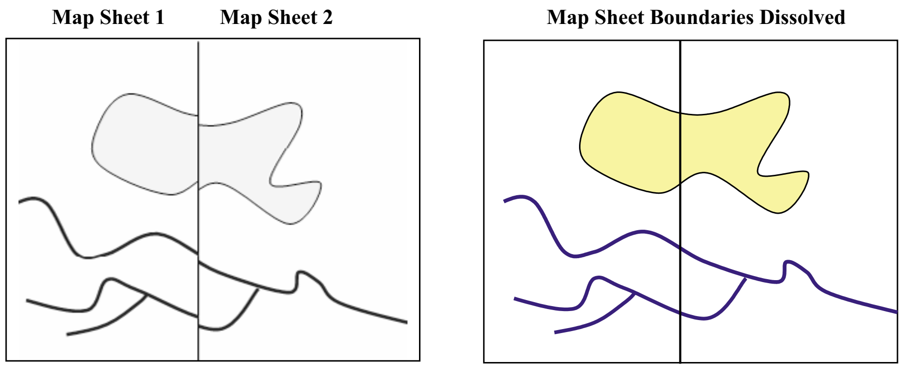


Figure 5.23: Example of edge matching.

Second, for use as a vector data layer, topology must be rebuilt as new lines and polygons have been created from the segments that lie across map sheets. This process can be automated, but problems may occur due to the tolerances used. Too large a tolerance and small edge polygons may be lost, too small a tolerance and lines and polygon boundaries may remain unjoined.

Finally, redundant map sheet boundary lines are deleted or dissolved (Jackson and Woodsford, 1991) note that although some quasi-automatic scanning edge matching is available, in practice the presence of anomalies in the data produced can require considerable human input to the process. Certain data sources may give rise to internal distortions within individual map sheets. This is especially true for data derived from aerial photography as the movement of the aircraft and distortion caused by the camera lens can cause internal inaccuracies in the location of features within the image. These inaccuracies may remain even after transformation and re-projection. These problems can be rectified through a process known as rubber sheeting (or conflation). Rubber sheeting involves stretching the map in various directions as if it were drawn on a rubber sheet. Objects on the map that are accurately placed are 'tacked down' and kept still while others that are in the wrong location or have the wrong shape are stretched to fit with the control points. These control points are fixed features that may be easily identified on the ground and on the image. Their true co-ordinates may be determined from a map covering the same area or from field observations using GPS. Distinctive buildings, road or stream intersections, peaks or coastal headlands may be useful control points. Figure 5.6 illustrates the process of rubber sheeting. This technique may also be used for re-projection where details of the base projection used in the source data are lacking. Difficulties associated with this technique include the lack of suitable control points and the processing time required for large and complex data sets. With too few control points the process of rubber sheeting is insufficiently controlled over much of the map sheet and may lead to unrealistic distortion in some areas.

GEOCODING ADDRESS DATA: Geocoding is the process of converting an address into a point location (McDonnell and Kemp, 1998). Since addresses are an important component of many spatial data sets, geocoding techniques have wide applicability during the encoding and preparation of data for analysis.

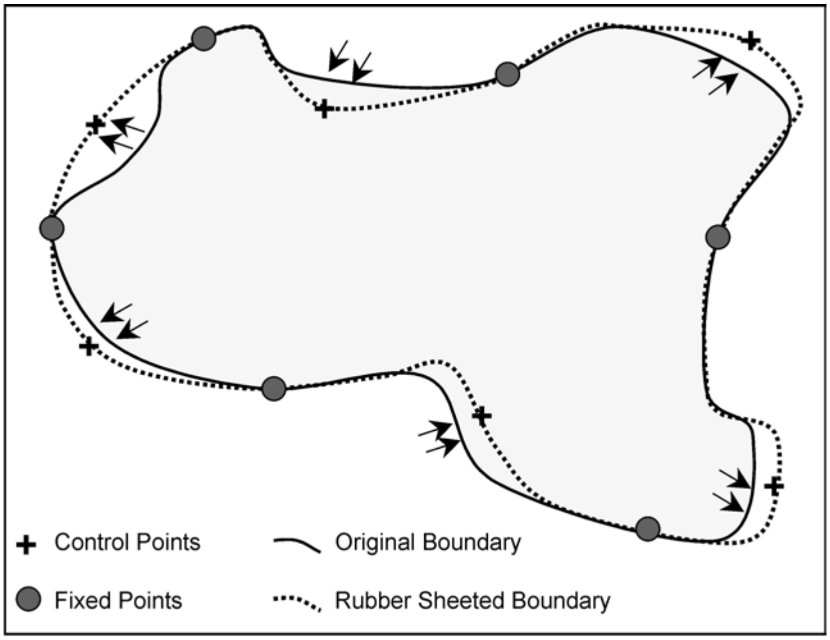


Figure 5.24: Example of rubber sheeting.

During geocoding the address itself, a postcode or another non-geographic descriptor (such as place name, land owner or land parcel reference number) is used to determine the geographical co-ordinates of a location. UK postcodes can be geocoded with an Ordnance Survey grid reference. Several products are available that contain a single data record for each of the 1.6 million postcodes in the UK. In these files, each data record contains the OS Grid Reference and local government ward codes for the first address in each postcode. Many GIS software products can geocode US addresses, using the address, zip code or even place names. Address matching is the process of geocoding street addresses to a street network. Locations are determined based on address ranges stored for each street segment. Geocoding can be affected by the quality of data. Address data are frequently inconsistent: place names may be spelt incorrectly, addresses may be written in different formats and different abbreviations exist for words that appear frequently in addresses, the use of standards for address data is particularly relevant to geocoding.

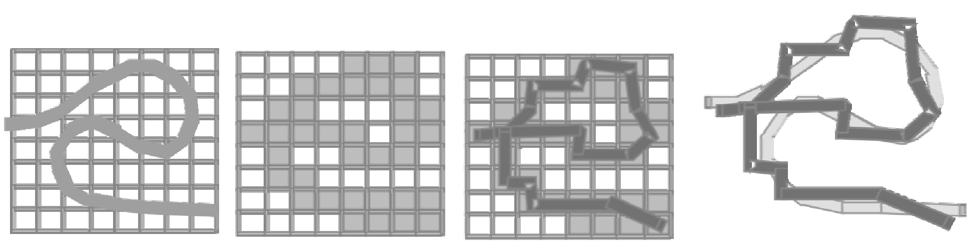


Figure 5.25: Vector to raster exchange errors.

DATA CONVERSION

While manipulating and analyzing data, the same format should be used for all data. When different layers are to be used simultaneously, they should all be in vector or all in raster format. Usually the conversion is from vector to raster, because the biggest part of the analysis is done in the raster domain. vector data are transformed to raster data by overlaying a grid with a user-defined cell size. Sometimes the data in the raster format are converted into vector format. This is the case especially if one wants to achieve data reduction because the data storage needed for raster data is much larger than for vector data.

Remote-sensing images are digital datasets recorded by satellite operating agencies and stored in their own image database. They usually have to be converted into the format of the spatial (raster) database before they can be downloaded.

GEOGRAPHIC DATA – LINKAGES AND MATCHING

Linkages: A GIS typically links different sets. Suppose we want to know the mortality rate to malnutrition among children under 10 years of age in any state. If we have one file that contains the number of children in this age group, and another that contains the mortality rate from malnutrition, we must first combine or link the two data files. Once this is done, we can divide one figure by the other to obtain the desired answer.

Exact Matching: Exact matching means when we have information in one computer file about many geographic features (*e.g.,* towns) and additional information in another file about the same set of features. The operation to bring them together is easily achieved by using a key common to both files -- in this case, the town name. Thus, the record in each file with the same town name is extracted, and the two are joined and stored in another file.

Hierarchical Matching: Some types of information, however, are collected in more detail and less frequently than other types of information. For example, land use data covering a large area are collected quite frequently. On the other hand, land transformation data are collected in small areas but at less frequent intervals. If the smaller areas nest (*i.e.,* fit exactly) within the larger ones, then the way to make the data match of the same area is to use hierarchical matching -- add the data for the small areas together until the grouped areas match the bigger ones and then match them exactly.

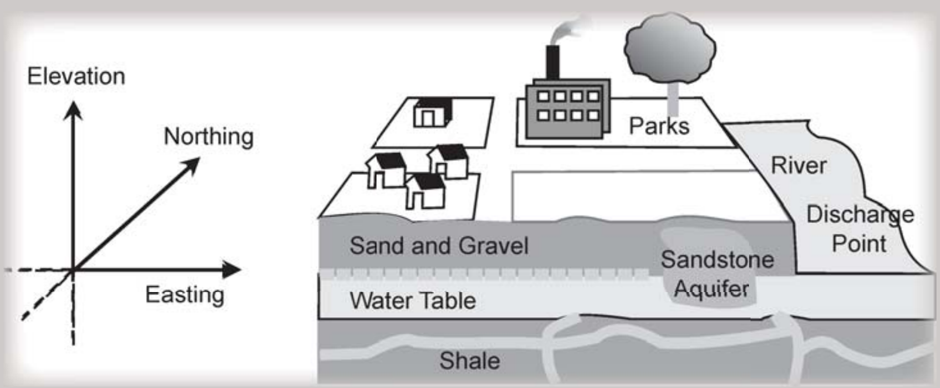
Fuzzy Matching: On many occasions, the boundaries of the smaller areas do not match those of the larger ones. This occurs often while dealing with environmental data. For example, crop boundaries, usually defined by field edges, rarely match the boundaries between the soil types. If we want to determine the most productive soil for a particular crop, we need to overlay the two sets and compute crop productivity for each and every soil type. This is like laying one map over another and noting the combinations of soil and productivity.

A GIS can carry out all these operations because it uses geography, as a common key between the data sets. Information is linked only if it relates to the same geographical area.

This page intentionally left blank

Chapter 6

GIS and the Real World Model



In many ways GIS presents a simplified view of the real world. Since the processes involved are seldom straightforward because realities are irregular and constantly changing, perception of the real world depends on the observer. For example, a surveyor might see a road as two edges to be surveyed, the roadwork authority might regard it as an asphalt surface to be maintained, and the driver will see it as a highway. Moreover, the real world may be described in terms of countless phenomena, from basic subatomic particles up to the dimensions of oceans and continents. The complexity of the real world, as well as the broad spectrum of its interpretations, suggests that GIS system designs will vary according to the capabilities and preferences of their creators. This human factor can introduce an element of constraint, as data compiled for a particular application may be less useful elsewhere.

The systematic structuring of the data determines its ultimate utility and consequently the success of the relevant GIS application. This aspect is also characteristic of the data available in traditional maps and registers. The real world can be described only in terms of models that delineate the concepts and procedures needed to translate real world observations into data that are meaningful in GIS. The process of interpreting reality by using both a real-world and a data model is called data modelling. The principles involved are illustrated in Figures 6.1 and 6.2.

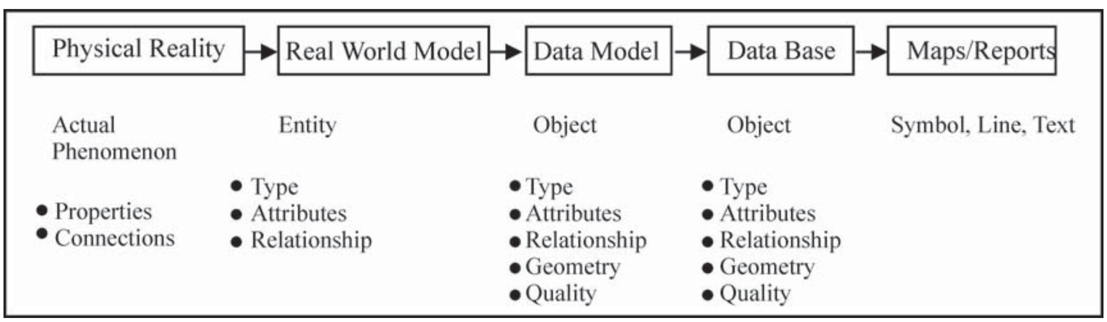
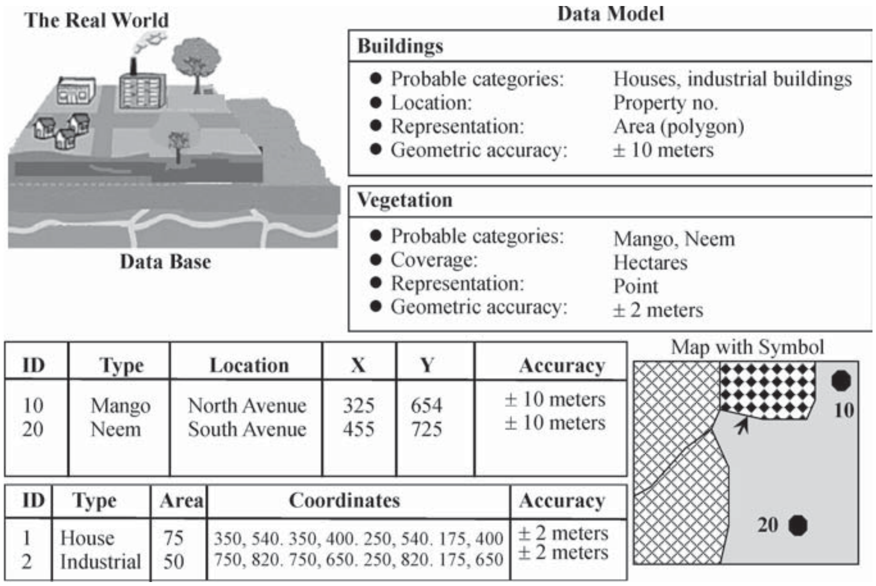


Figure 6.1: GIS makes simplified models to represent real world models. The data model is transferred to a database that can handle digital data, from which the data can be presented.

Real World Model

The arrangement of the real-world model determines which data need to be acquired. The basic carrier of information is the entity, which is defined as a real-world phenomenon that is not divisible into phenomena of the same kind. An entity consists of:

- Type classification
- Attributes
- Relationships



ID	Type	Location	X	Y	Accuracy
10	Mango	North Avenue	325	654	± 10 meters
20	Neem	South Avenue	455	725	± 10 meters

ID	Type	Area	Coordinates	Accuracy
1	House	75	350, 540. 350, 400. 250, 540. 175, 400	± 2 meters
2	Industrial	50	750, 820. 750, 650. 250, 820. 175, 650	± 2 meters

Figure 6.2: The transformation of the real world into GIS is achieved by means of simplification and models in the form of maps and reports.

GEOGRAPHICAL DATA

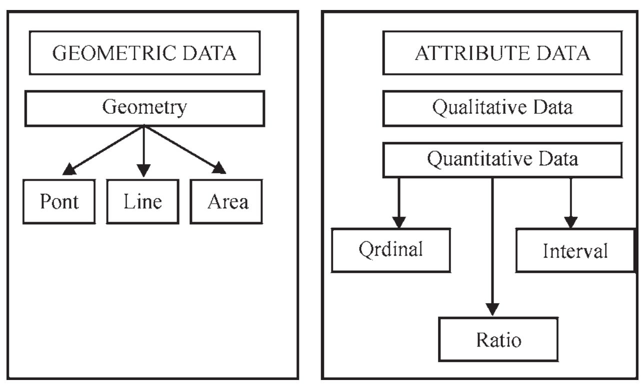


Figure 6.3: Geographical data can be divided into geometric and attribute data.

Entity Types

The concept of entity types assumes that uniform phenomena can be classified as such. During the classification process, each entity type must be uniquely defined to preclude ambiguity. For example, 'house' must be defined in such a way that 'detached house at No.10, Marris Road Civil Lines' is classified under 'house' and not under 'industrial building'.

Some user organizations may need to classify entity types into categories as well as according to type. For example, national highways, state highways, urban roads, and village roads might come under the 'roadways' category; alternatively, all entities within a specific geographical area might belong to a unique category of that area. In geographical data an entity type is also known as the nominal scale or qualitative data (Figure 6.4).

ENTITY ATTRIBUTES: Each entity type may incorporate one or more attributes that describe the fundamental characteristics of the phenomena involved. For example, entities classified as 'buildings' may have a 'material' attribute, with legitimate entries 'frame' and 'masonry' and a 'number of stories' attribute with legitimate values of 1 to 10, and so on.

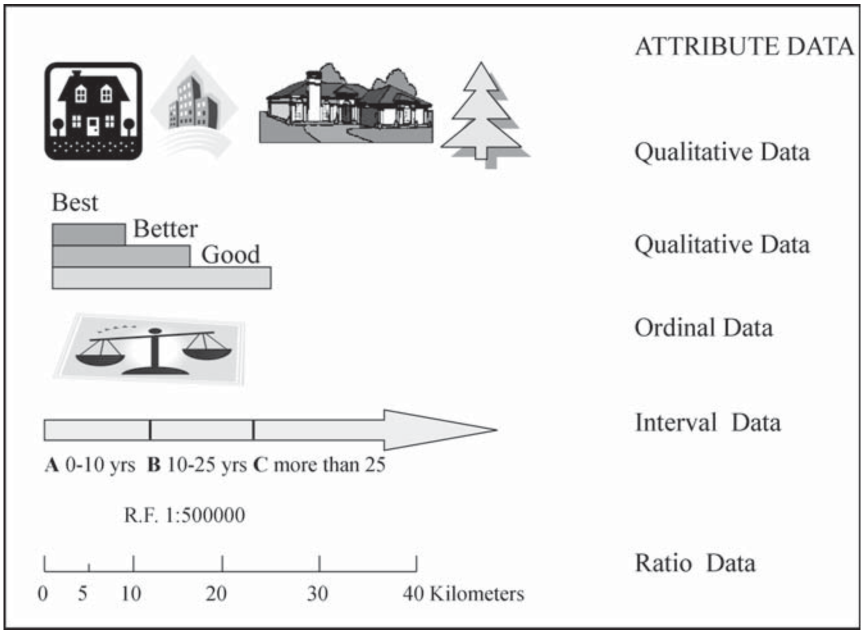


Figure 6.4: Attribute data consists of qualitative or quantitative data. Qualitative data specify the types of object, while quantitative data can be categorized into ratio data, data measured in relation to a zero starting point; interval data, data arranged into classes; and ordinal data, which specify quality by using text.

In principle an entity may have any number of attributes. For example, a lake may be described in terms of its name, depth, water quality, or fish population as well as its chemical composition, biological activity, water colour, potability, or ownership. Attributes may also describe quantitative data, which may be ranked in three levels of accuracy: ratio, interval, and ordinal. The most accurate are ratio or proportional attributes, such as length and area, which are measured with respect to an origin or starting point and on a continuous scale. Interval data, such as age and income category, comprise numerical data in groups and are thus less accurate. The least accurate are ordinal data of rank, such as 'good', 'better' and 'best' which describe qualitative data in text form. These could also be characterized as quality data.

Relation	Example
Pertains/belongs	A depth figure pertains to a specific shoal, or a pipe belongs to a larger network of contiguous pipes.
Comprises	A state comprises districts, which in turn comprise townships.
Located in/on	A particular building is located on a land parcel.
Borders on	Two properties have a common border.

Figure 6.5: Examples of relations often exist between entities.

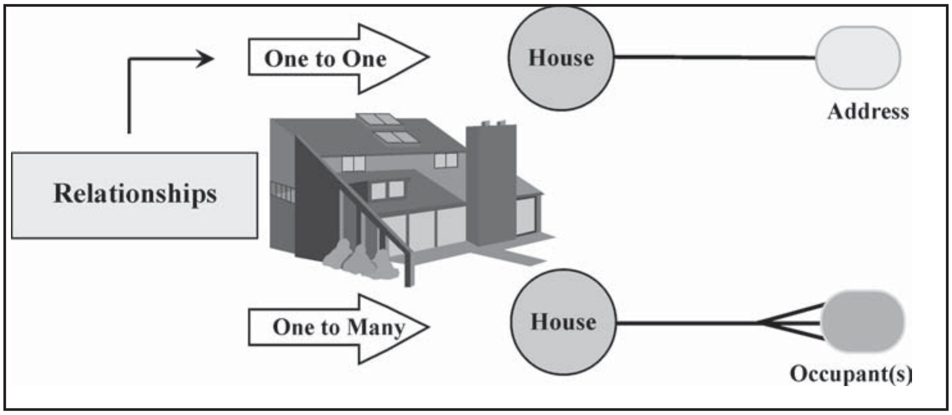


Figure 6.6: A single entity can be described by several objects (i.e., there are many relationships between entities).

Entity Relations

Relations often exist between entities. Typically, these include (Figure 6.7):

Objects are characterized by:

- Type
- Attributes
- Relations
- Geometry
- Quality

Real-world models and entities cannot be realized directly in databases, partly because a single entity may comprise several objects. For instance, the entity 'Marris Road' may be represented as a compilation of all the roadway sections between intersections, with each of the sections carrying object information. Multiple representations produced by such divisions may promote the efficient use of GIS data. This means that information-carrying units and their magnitudes must be selected before the information is entered in a database. For example, the criteria for dividing a roadway in sections must be selected before the roadway can be described.

Objects: Objects in a GIS data model are described in terms of identity type, geometric elements, attributes, relations, and qualities. Identities, which may be designated by numbers, are unique: no two objects have the same identity. Type codes are based on object classifications, which can usually be transferred from entity classifications. An object may be classified under one type code only.

Data models may be designed to include:

- Physical objects, such as roads, water mains, and properties
- Classified objects, such as types of vegetation, climatic zones, or age groups

- Events, such as accidents or water leaks
- Continuously changing objects, such as temperature limits
- Artificial objects, such as elevation contours and population density
- Artificial objects for a selected representation and database (raster)

Geographical Representation of Objects

Graphical information on objects may be entered in terms of:

- Points (no dimensions)
- Lines (one dimension)
- Areas (two dimensions)

POINTS: A point is the simplest graphical representation of an object. Points have no dimensions but may be indicated on maps or displayed on screens by using symbols. The corner of a property boundary is a typical point, as is the representative coordinate of a building. It is, of course, the scale of viewing that determines whether an object is defined as a point or an area. In a large-scale representation a building may be shown as an area, whereas it may only be a point (symbol) if the scale is reduced.

LINES: Lines connect at least two points and are used to represent objects that may be defined in one dimension. Property boundaries are typical lines, as are electric power lines and telecommunications cables. Road and rivers, on the other hand, may be either lines or areas, depending on the scale.

AREAS: Areas are used to represent objects defined in two dimensions. A lake, an area of grassland, or a city may typically be represented by an area. Again, physical size in relation to the scale determines whether an object is represented by an area or by a point. An area is delineated by at least three connecting lines, each of which comprises points. In databases, areas are represented by polygons (*i.e.,* plane figures enclosed by at least three straight lines intersecting at a like number of points). Therefore, the term polygon is often used instead of area.

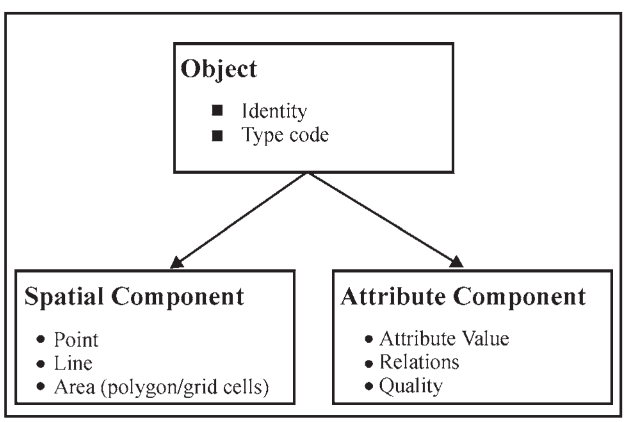


Figure 6.7: In a data model, objects are categorized as object classifications, geometric elements (point, line, area), attributes, relations between the entities and quality definitions of these descriptive elements.

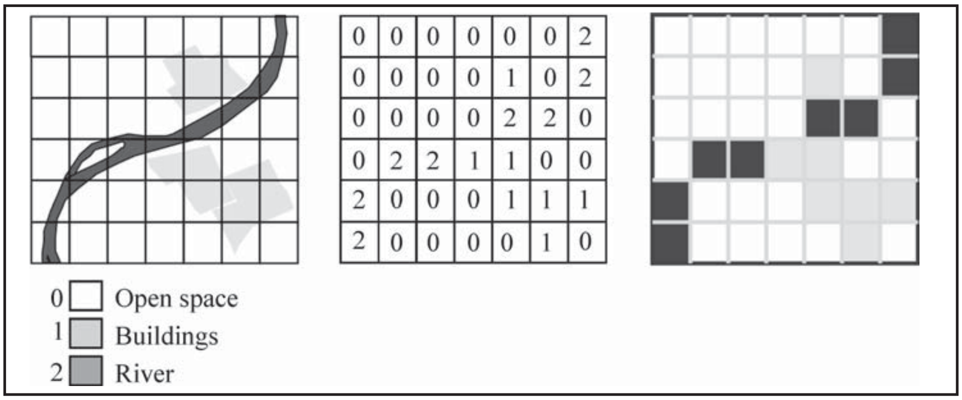


Figure 6.8: Land use/ Land cover in the form of a raster map. The land use is registered in a raster system with cells. Each category is assigned its own symbol on the map.

Physical reality is often described by dividing it into regular squares or rectangles so that all objects are described in terms of areas (Figure 6.8). This entire data structure is called a grid. Population density is well suited to grid representation; each square or rectangle is known as a cell and represents a uniform density or value. The result is a generalization of physical reality. All cells of a grid in a data model or a database are of uniform size and shape but have no physical limits in the form of geometric lines.

In the traditional layer based data model heights are treated as attributes to the objects, not as a part of the geometry. But the real world is three dimensional.

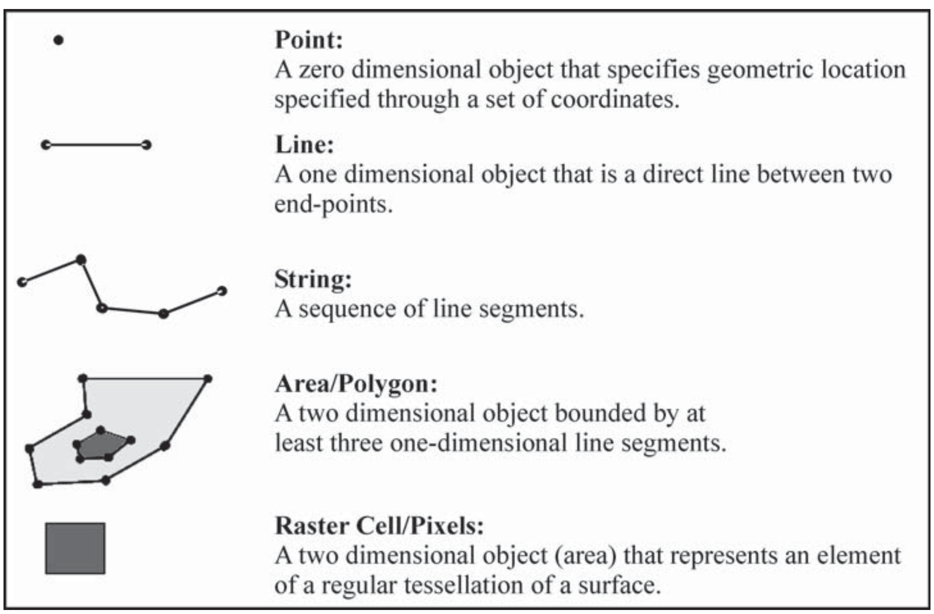


Figure 6.9: Point, line, area objects in GIS.

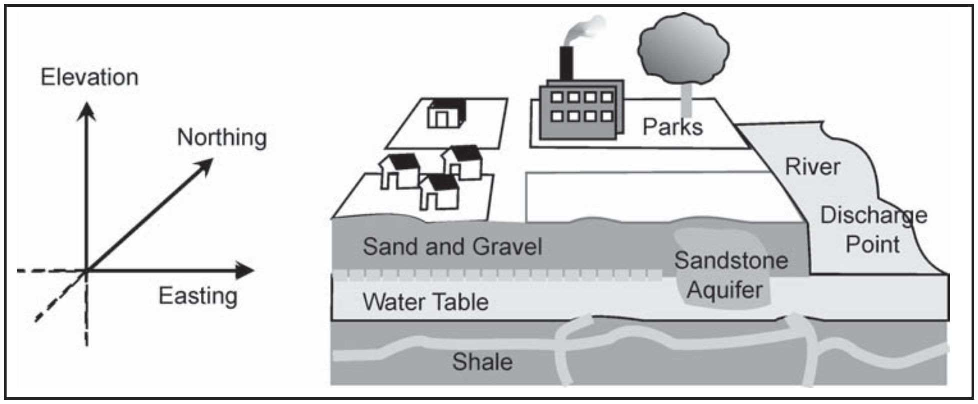


Figure 6.10: The world is three-dimensional with phenomena having a location and surface area in both elevation and ground plane.

Object Attributes

Object attributes are the same as the entity attributes of the real-world model. Attributes describe an object's features and may thus be regarded as a computer's 'knowledge' of the object. In practice, object attributes are stored in tables (Figure 6.11), with objects on lines and attributes in columns. Theoretically, attribute values connected to grid data can be presented in the same way. Each grid cell corresponds to an object.

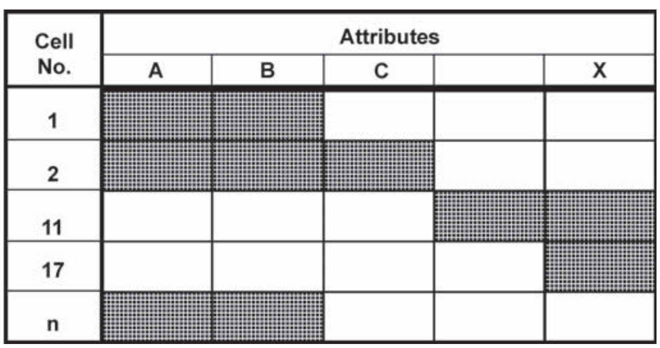


Figure 6.11: In principle, the difference between vector data and raster data is not that great. Raster data could well be arranged in tabular form with each cell number representing a line and each attribute (layer or raster) a column.

Object Relations

Object relations are the same as entity relations in the real-world model. Differentiation is made between:

1. Relations that may be calculated from:
   a. The coordinates of an object: for example, which lines intersect or which areas overlap
   b. Object structure (relation), such as the beginning and end points of a line, the lines that form a polygon, or the locations of polygons on either side of a line

2. Relations that must be entered as attributes, such as the division of a townships in to different wards or the levels of crossing roads that do not intersect.

Quality

The true value of any description of reality depends on the quality of all the data it contains:

- Graphics
- Attributes
- Relations

Graphical data accurate to ± 0.1 meter obviously describe reality more faithfully than data accurate to ± 100 meter. Similarly, recently updated data are preferable to five-year-old data (which bring in temporal factors).

In the initial data modelling stage, the assessment of the data quality should include:

- Graphical accuracy (such as ± 1.0 meter accuracy)
- Updating (when and how data should be updated)
- Resolution/detailing (*e.g.,* whether roads should be represented by lines or by both road edges)
- Extent of geographical coverage, attributes included, and so on
- Logical consistency between geometry and attributes
- Representation: discrete versus continuous
- Relevance (*e.g.,* where input may be surrogate for original data that are unobtainable)

Information on the quality of data is important to users of the database.

From Database to GIS to Map

Once a data model is specified, the task of realizing it in a computer is technical and the task of entering data is simple and straightforward, albeit time consuming. A database need seldom be made to suit a data model, as many databases compatible with GIS applications are now on the market. The problem at hand is more one of selecting a suitable database with regard to:

- Acquisition and control
- Structure
- Storage
- Updating and changing
- Managing and exporting/importing
- Processing
- Retrieval and presentation
- Analyses and combinations

Needless to say, a well-prepared data model is vital in determining the ultimate success of the GIS application involved. Users view reality using GIS products in the form of maps with symbols, tables, and text reports.

Shortcomings of the Traditional GIS Data Model

Entities and Fields: In the real world, one specific area or field may have many different characteristics; one area will in reality represent a number of entities or object types, such as coniferous forest, protected area, property no. 118/1/B, and so on. We experience on a daily basis that it is the area as an entity that carries the information. However, in our real-world model we split phenomena into entities (entity: a real-world phenomenon that is not divisible into phenomena of the same kind) and allow the entities to be bearers of information. This model will allow an entity to represent only one phenomenon (*e.g.,* only Aligarh fort or only protected area). To adapt the model to reality, overlapping phenomena (entities/objects) are separated into different layers. Reality is thus adapted to fit into a layer system, which is also traditionally used in map presentation. In the real world, areas are not divided into any form of horizontal two-dimensional physical layer not even geological layers (strata) are presented in this way.

We can say that geometry – where coordinates define points, lines, and areas – is in many ways an artificial concept and an unnatural way to describe reality. Coordinates are not tangible and are never used in our everyday description of reality. Instead, we define a phenomenon's location in relation to other phenomena with which the recipient of the information is familiar. We can therefore establish that our model of reality is not perfect. During the 1990s, new models have been developed, known as object-oriented models, which to a certain extent can allow for the fact that the entity bearing the information can represent many phenomena. Object-oriented database systems are currently little used in commercial GIS but would appear to have many advantages over traditional database systems.

Uncertainty: To regard the real world as consisting of geometric constructs (points, lines, areas) means viewing objects as discrete data model representations. That is, all objects have clearly defined physical limits. These limitations are most obvious on maps, where lines imply sharp demarcations with no smooth, continuous transitions.

A discrete data model does not always suit reality. Difficulties arise in depicting phenomena that lack clear physical demarcation, such as soil types, population densities, or prevailing temperatures. There can also be uncertainty in the attribute values to be retained. In the traditional discrete model, entities or objects are defined as being either within specific classes or outside them and thus operate only with areas that are homogeneous with respect to limitation and classification. In reality, phenomena will often vary even within small, limited areas. For example, coniferous forest often contains deciduous trees, population density is variable, and terrain surface changes continuously. Once again, we have established that our real-world model is not perfect and that it is closely linked to traditional mapping concepts. Some of these problems can be partly solved by using the fuzzy set theory, which allows an object to belong only partially to a class. The fuzzy theory has as yet been little used in commercial GIS software; thus the significance of this type of deficiency in the data model is left to the person interpreting the results (maps and reports) of the GIS process.

CONCEPTUAL GENERALIZATION: When points, lines, and polygons are selected as the geometric representation of objects, this very often results in a generalization of the real world; a town can be represented by a point rather than a polygon, and a road will frequently be represented by a center line and not two road verges. The need to divide objects into classes also results in a generalization. For example, an area of forest that is mainly coniferous, with some deciduous, will often be generalized and classified as coniferous, not as a combination. Thus conceptual generalization is also a method for handling uncertain elements.

It will always be necessary to make choices about such generalizations in relation to the real world when making data models. This may be seen as a problem, but generalization is also a technique that makes it possible to obtain an overview of our complex reality. It can also be difficult to create data models that have a uniform and clear definition of the objects' classes. For example, does a pedestrian area that is accessible to emergency vehicles classify as a road?

Role of Maps in Data Modelling

Maps are, in general, good sources for describing objects and their attributes. However, maps always represent particular models of the real world, and GIS should represent the real world, not the maps that depict it. For instance, ferry routes are often shown by dotted lines on maps, whereas in transport planning data models should form integral parts of a contiguous road network. As a rule of thumb, therefore, always look at a map as a data source, not as a data model.

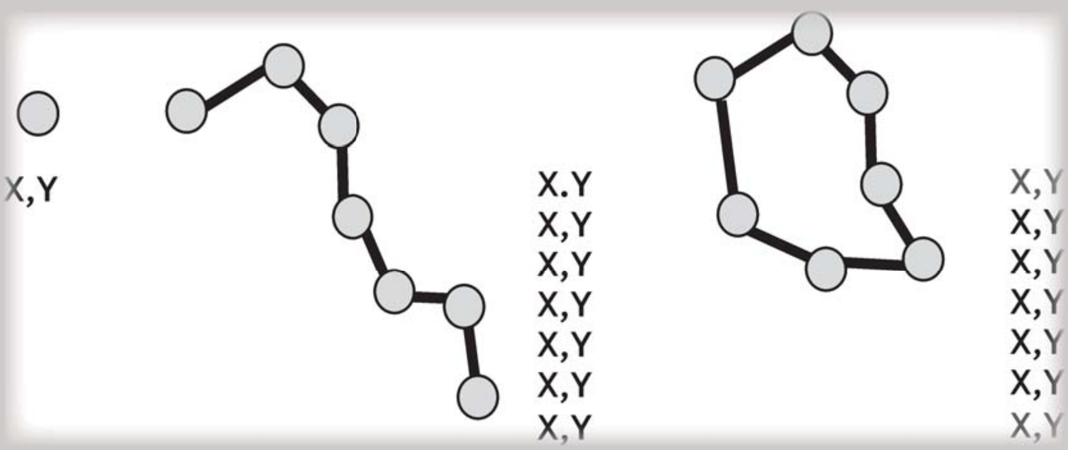
Extension of the Reality Concept

The traditional model for transformation from the real world to GIS, as described above, has its obvious faults. In addition, it only describes flat and unchanging reality. Models for describing objects in three-dimensional space and terrain have not yet been discussed, nor has the fourth dimension – time – and its inroads into a geographical data model. The same applies to models for dealing with objects (traffic) moving along defined networks.

Here it is also most practical to use the same basic concept: a geometry consisting of points, lines, and polygons, and attributes that describe the objects or phenomena. Elevation values can be linked to points, lines, and polygons and thereby give the objects a position in space. The surface of the terrain can be described with the help of sloping areas or with the help of horizontal surfaces with an elevation value linked as an attribute. Elevation values can also be linked to objects such as towers, wells, and buildings as attributes. The time factor can be accommodated by storing all historical data for the objects, such as changes in the geometry or attribute values. The movement of objects (traffic) along a road network can be simulated by assigning attribute values to elements in the network. These should be values that are relevant to the transfer speeds, and the sum of attribute values for different routes will be the measurement of passage in time or distance. Undoubtedly the traditional data model concept has definite drawbacks when describing these new real-world elements. We must accept that the real world is too complex to be described in full at present, although researchers are continuously engaged in developing improved models.

Chapter 7

Basic Data Models in GIS



GIS depicts the real world through models involving geometry, attributes, relations, and data quality. In this chapter, the realization of models is described, with the emphasis on geometric spatial information, attributes, and relations.

Spatial information is presented in two ways: as vector data in the form of points, lines, and areas (polygons); or as grid data in the form of uniform, systematically organized cells. Geometric presentations are commonly called digital maps. Strictly speaking, a digital map would be peculiar because it would comprise only numbers (digits). By their very nature, maps are analog, whether they are drawn by hand or machine, or whether they appear on paper or displayed on a screen. Technically speaking, GIS does not produce digital maps-it produces analog maps from digital map data. Nonetheless, the term digital map is now so widely used that the distinction is well understood.

Vector Data Model

The basis of the vector model is the assumption that the real world can be divided into clearly defined elements where each element consists of an identifiable object with its own geometry of points, lines, or areas (Figure 7.1). In principle, every point on a map and every point in the terrain it represents is uniquely located using two or three numbers in a coordinate system, such as in the northing, easting, and elevation Cartesian coordinate system. On maps, coordinate systems are commonly displayed in grids with location numbers along the map edges. On the ground, coordinate systems are imaginary, yet marked out by survey control stations. Data usually may be transformed from one coordinate system to another.

With few exceptions, digital representations of spatial information in a vector model are based on individual points and their coordinates. The exceptions include cases where lines or parts of lines (*e.g.,* those representing roads or property boundaries) may be described by mathematical functions, such as those for circles or parabolas. In these cases, GIS data include equation parameters: for example, the radii of the circles used to describe parts of lines. Together with the coordinate data, instructions are entered as to which points in a line are unconnected and which are connected. These instructions can subsequently be used to create lines and polygons and to trigger 'pen up' and 'pen down' functions in drawing.

Coordinate systems are usually structured so that surveys along an axis register objects in a scale of 1: 1; that is, 1 meter along the axis corresponds to 1 meter along the ground. In principle, the degree of accuracy of measurements along an axis is decided by the type of measuring method applied, while the required degree of precision will naturally influence the amount of work required to gather the data.

Mathematically, a vector is a straight line, having both magnitude and direction. Therefore, a straight line between two data coordinate points on a digital map is a vector-hence the concept of vector data used in GIS and the designation of vector-based systems.

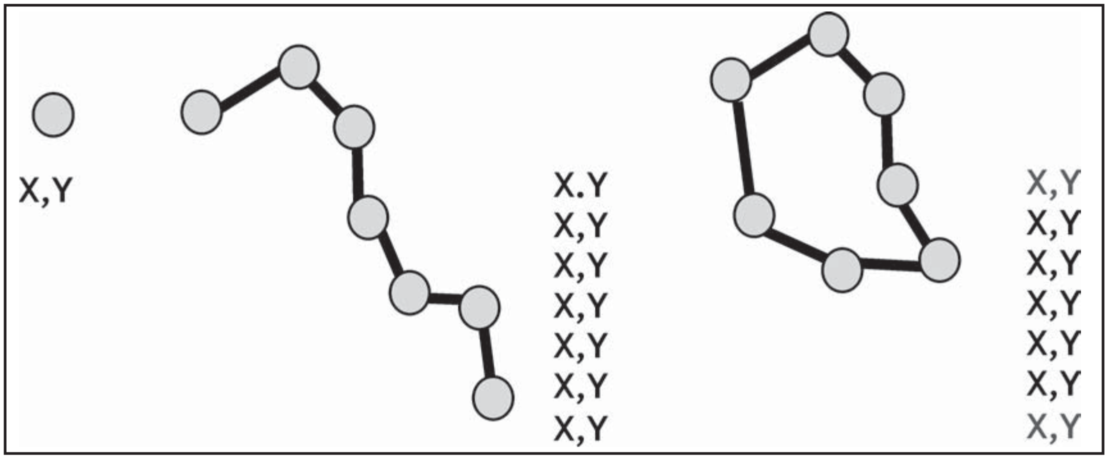


Figure 7.1: The vector data model, objects are described as points, lines or areas (polygons). These three geometric phenomena are described individually by a single point in a coordinate system and with connected lines (lines and area features).

In a vector model, points, lines, and areas (polygons) are the homogeneous and discrete units that carry information. As discussed above, these three types of object may be represented graphically using coordinate data. However, the objects may also carry attributes that can be digitized, and all digital information can be stored.

Coding Digital Data for Map Production

In any map, data are traditionally coded. Roads, contour lines, property boundaries, and other data indicated by lines are usually shown in lines of various widths and colours. Symbols designate the locations of mosques, airports, and other buildings and facilities. In other words, coordinates and coding information identify all objects shown on a map.

Similarly, the digital data used to produce maps are also coded, usually by the assignment of numerical codes used throughout the production process – from the initial data to computer manipulation and on to the drawing of the final map. Each numerical code series contains specific codes assigned to the objects in the group.

In thematic coding, which may be compared to the overlay separation of conventional map production (Figure 7.2), data are divided into single-topic groups, such as all property boundaries. Information on symbol types, line widths, colours, and so on, may be appended to each thematic code, and various combinations of themes may be drawn. Data may be presented jointly in this way only if all objects are registered, using a common coordinate system.

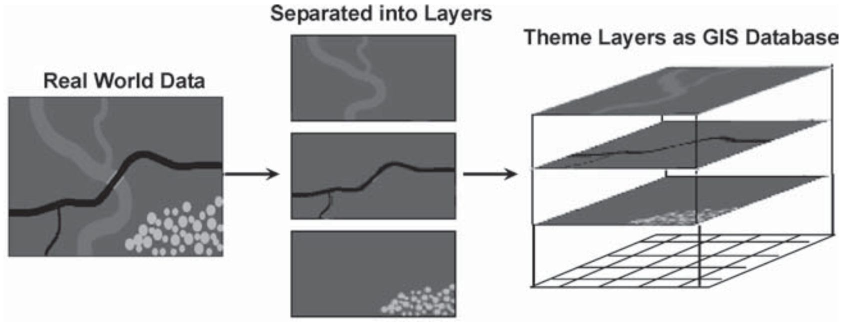
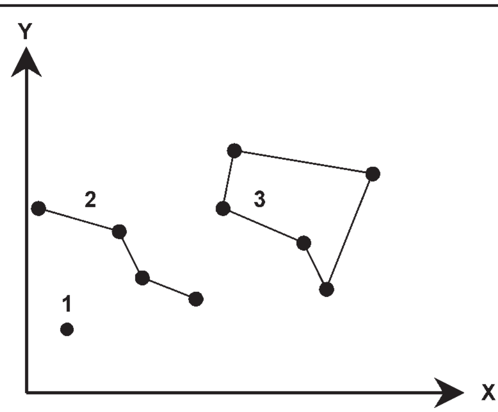


Figure 7.2: Theme codes in the digital map data can be send to separate data into thematic layers.

Coding Digital Data for GIS

Point objects may easily be realized in a database because a given number of attributes and coordinates is associated with each point (Figure 7.3). Line and polygon objects are more difficult to realize in a database because of the variation in the number of points composing them. A line or a polygon may comprise two points or 1000 or more points, depending on the extent of the line and the complexity of the area, which is delineated by a boundary line that begins and ends at the same point. Object spatial information and object attributes are often stored in different databases to ease the manipulation of lines and areas, but in some systems they are stored together. As pivotal attributes are often available in existing computer memory files, dividing the databases conserves memory by precluding duplicate storage of the same data.



Geometry	ID	Attributes		Coordinates
Point	1	A	B	(Single Point) 2,3
Line	2	C	D	(String) 1,7 3,6 5,5 7,4
Area	3	E	F	(Closed Polygon) 8,7 11,6 12,5 14,9 9,15

Figure 7.3: Each object is assigned attributes and coordinates. The number of coordinates for lines and polygons depends on the length or circumference of the object.

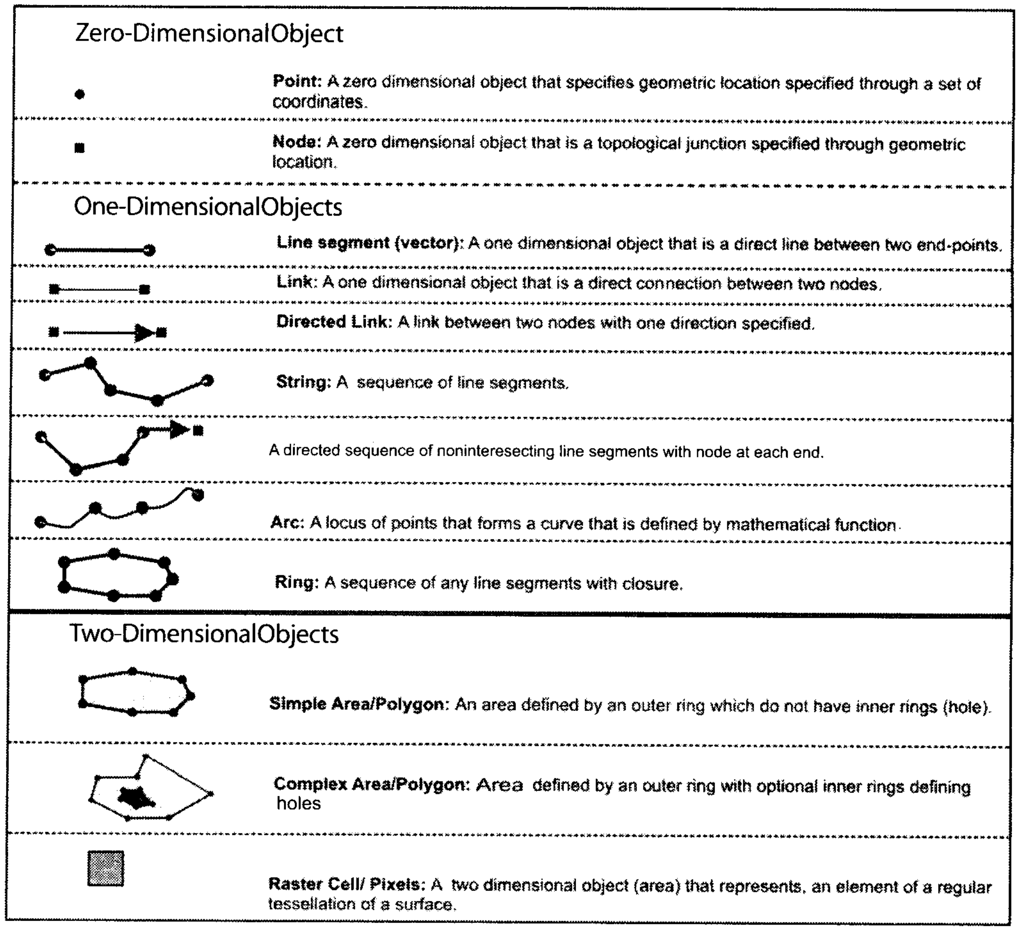


Figure 7.4: Some geometric objects.

Here we will take a look at the key features of some of the many data structures which have been developed to store vector data. In doing so, we will discover that ideas from mathematics have been very important in the development of these data structures.

Storing Points and Lines

Figure 7.5 shows a simple vector layer containing examples of all three vector data types: points (*e.g.*, the houses), lines (*e.g.*, the roads and rivers) and areas (*e.g.*, the forests, agricultural fields). The locational information for these features is stored using geographical coordinates but how is this information actually stored on the computer?

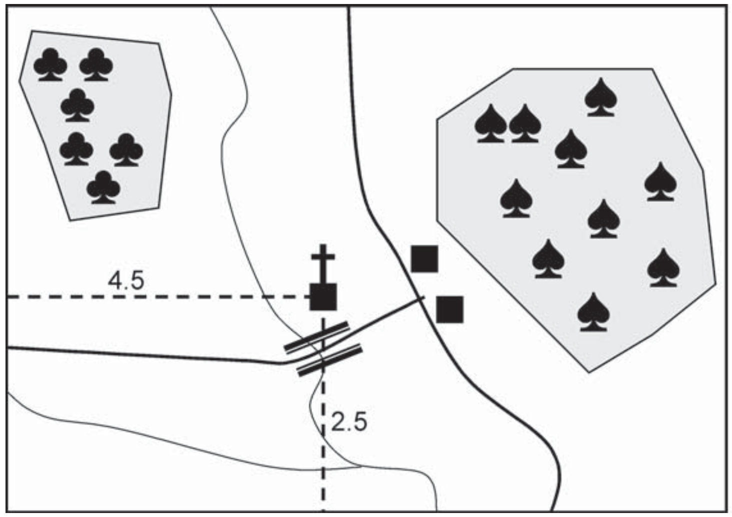


Figure 7.5: Imaginary map.

Table 7.1: Coordinate data for points.

Building 1	4.5	2.5
Building 2	5.8	2.9
Building 3	6.0	2.2

Let us start with points, since these are the simplest features. Consider the Departmental Store near the bridge. We need to be able to store the position of this building and to do this we can measure its distance from the left hand corner of the map both horizontally and vertically as shown. This will give us a pair of figures. If we do the same for all three buildings, we can produce a small table (Table 7.1). This table is clearly very suitable for storage on the computer – it could be held in Excel spreadsheet for example or as a table in a dBase database.

Using this information we can draw a new version of our map by raking a piece of graph paper and plotting the positions of the three points. What is more, we can measure the distance between any two points, so by storing the location of the points as a pair of coordinates, we can do three things:

1. Store the locational data from the map on a computer.
2. Use this information to reproduce the map.
3. Make simple calculations from the data.

In practice, measuring the coordinates from the origin of the map is not very useful, especially if the area we are interested in covers more than one map sheet. Almost all maps have a grid on them which allows us to read off the position of points in some more useful coordinate system. Many, for example use the UTM (Universal Transverse Mercator) system which gives a standard way of representing positions for any point in the region. The principle is still the same however – the vector representation of the point is as a pair of coordinates. We also need to be able to store the attributes of each point, and table 7.2 indicates one possible way of doing this – using extra columns in the table.

Table 7.2: Attribute data for points.

	X Coordinate	Y Coordinate	Feature Code	Building Material	Name
Building 1	4.5	2.5	Temple	Stone	Krishna's
Building 2	5.8	2.9	House	Brick	Shaan's
Building 3	6.0	2.2	House	Stone	Sameer's

The first new column is what is known as a feature code. This indicates what category a feature falls into – in the figure each building has been identified as being either a departmental store or a house. This information could then be used when redrawing our map to determine what sort of symbol to use to represent each feature–a small black square for a house, a special symbol for a departmental store for example. The other columns are used to record information which is specific to each building, such as its name (if appropriate) and building material.

Now let us see how we can extend this idea to enable us to store the same data for lines and areas. Consider the road coming down from the top of the map between the two forest areas. Large parts of it are relatively straight, and so we can approximate its course by a series of short straight lines. To store the location of the lines, we simply need the coordinates of the points between each straight section, and so this gives us the following as our representation of the road.

Table 7.3: Coordinate data for part of a line.

4.5	10.0
4.5	5.7
5.5	2.5
6.5	0.3
6.8	0.0

Using this set of numbers we can redraw our road in the same way as we did with the points-in this case we would plot the position of each point on a sheet of graph paper, and join successive points with a straight line. We can calculate the length of any of the straight sections simply by calculating the distance between the points at each end – adding all these up gives us an estimate of the length of the line.

With a curved line, such as a river, this representation by a series of straight sections will only be an approximation of course as shown in Figure 7.6.

The only way to improve the match between the real line and our series of straight segments is to add extra points along the line. The more points we add, the closer the match, but of course each extra point means more data to be stored and hence larger files on the computer.

Again, the basic table of X Y coordinates for a line is a relatively simple file which could be stored in a spreadsheet or a database package. However, we also want to store attributes for this line, and this is when we begin to run into difficulties.

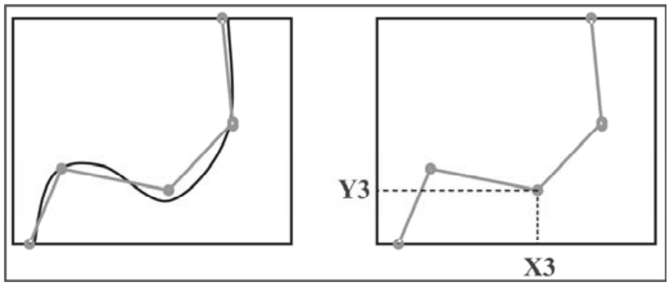


Figure 7.6: Approximating a curved line with a series of straight segments.

Left: original line in black, with positions of digitized points connected by straight lines in grey.

Right: digitized representation of the original line, showing X and Y coordinates for third point.

Imagine that we wish to store the following information about our roads.

Table 7.4: Attribute data for line's.

Name	Surface Quality	Peak Traffic Flow
Khair Road	Fair	600
Anupshahar Road	Good	1000

This is a very simple table. The problem comes when we try and combine both sets of information into one table. For each road in our table, its course is represented by a large number of X Y coordinates rather than the single set we had for a point feature. We could try and simply add the X and Y columns on as we did with the points (Table 7.5).

This means that each feature is now represented by a different number of rows, depending on how many X Y coordinate pairs we have. To keep each feature in a single row, we could add more columns to each row as shown in table 7.6.

However, now we have different numbers of columns for each feature. We could add an extra column to each row indicating the number of X Y coordinates used to represent that road. However whichever way we do it, we still have a rather messy table to deal with, compared with the neat simplicity of the table for point data. As an added difficulty, many of the database systems commonly used today will not allow tables where each row has a different number of columns, or where a single feature is represented by more than one row – this is certainly true of the relational database systems often used in GIS.

Table 7.5: Adding locational information to attribute table for lines. Rows containing data for Khair Road are shaded light grey.

Name	Surface Quality	Peak Traffic Flow	X Coordinate	Y Coordinate
Khair Road	Fair	600	4.5	10.0
			4.5	5.7
			5.5	2.5
			6.5	0.3
			6.8	0.0
Anupshahar Road	Good	1000	0.0	1.5
			3.6	1.5
			5.5	2.5

Table 7.6: Alternative method of adding locational information to attribute table for lines.

Name	Surface Quality	Peak Traffic Flow	X_1	Y_1	X_2	Y_2	X_3	Y_3	X_4	Y_4	X_5	Y_5
Khair Road	Fair	600	4.5	10.0	4.5	5.7	5.5	2.5	6.5	0.3	6.8	0.0
Anupshahar Road	Good	1000	0.0	1.5	3.6	1.5	5.5	2.5				

Because of this, many vector GIS systems solve the problem by storing the locational and attribute data separately. The attribute data is stored in a standard database package, but the locational data is stored using specially written software which can handle its more complicated structure. Such systems are often referred to as geo-relational, because the attributes are often held in a relational database, with the geographical or locational data being handled separately. The best known of these systems is probably ARC/INFO, in which the attribute data is handled by a standard database package – INFO – and the locational data is handled by specially written software – ARC.

Having separated the location and the attribute data, such systems then have to make sure that they can link back together again when necessary for example if a user selects a particular feature by querying the database, then in order to draw that feature on the screen, it will be necessary to retrieve its locational data. This is done by making sure that each feature has some sort of unique identifier which is stored with both the locational and attribute data.

The separate storage of attribute and spatial information data requires that all objects in the attribute tables be associated with the corresponding spatial information. This association is achieved by inserting spatially stable and relevant attribute data or codes from the attribute table into the spatial information, or vice versa. In other words, identical objects have the same identities in both databases.

The identity (ID) codes used to label and connect spatial information and attribute table data are most often numerical, but may be alphanumerical. If the data are ordered in a manuscript map, each object may be assigned a serial number used in both the spatial information and the attribute databases. ID codes allow differentiation between objects, whereas theme codes allow for differentiation between different groups of objects. In theory, identity codes and thematic codes are both attribute data. However, they are very closely tied to geometry and are therefore often treated as such, as described above.

Spatially defined objects without attributes need no identifiers, but they are required for all objects that are listed in attribute tables, and manipulated spatially. Identifiers are normally entered together with the relevant data, but they may also be entered later, using an interactive human – machine process such as keying in identifiers for objects pointed out on the screen.

Table 7.7: Locational data for buildings.

Building – ID	X Coordinate	Y Coordinate
1	4.5	2.2
2	5.8	2.9
3	6.0	2.2

Table 7.8: Attribute data for buildings.

Building – ID	Feature Code	Building Material	Name
1	Temple	Stone	Krishna's
2	House	Brick	Shaan's
3	House	Stone	Sameer's

Let us first see how this works with the point data. In the original table (Table 7.1) we had a column which simply identified each point as Building 1, Building 2 etc. Instead we will now have a column labelled Building – ID which will contain the Identification number of each building. This number has nothing to do with any characteristic of the building in the real world, but is simply a number assigned to each building on the map. We can then split our original table into two tables (see Tables 7.7 and 7.8), one each for the locational and attribute data.

In the case of the road data, we might use the road identification number as our unique ID but this will not be a good idea if we wish to distinguish between different parts of the Khair road for example, so again we will simply use a number starting at 1. Our attribute table will now be as shown in Table 7.9.

Table 7.9: Modified attribute table for roads.

Road – ID	Name	Surface Quality	Peak Traffic Flow
1	Khair Road	Fair	600
2	Anupshahar Road	Good	1000

STORING AREA BOUNDARIES

Now that we have covered some of the basics of storing points and lines in a GIS, let us return to the third major type of spatial feature – the area. Figure 7.7 shows a simple area of forest, and one way to store this area is by storing the line which defines its boundary, as shown in Figure 7.7B.

We already know that we can store lines as a sequence of X, Y coordinate values – the only difference in this case is that the end of the line joins the start to make a closed boundary. As with the other lines, we have considered we must approximate the boundary in order to store it. The points would be stored in order; with the coordinates of the last being the same as the first on some systems the coordinates of the last point are simply assumed to be the same as the first. Figure 7.7B is what mathematicians call a polygon a closed shape described by a series of straight lines – and in the GIS literature the term polygon is often used to refer to areas.

As with points and lines, we will probably wish to store attributes for our areas. With points and lines we simply added a label to the point and line data stored in the database, but it makes less sense to add a label to the boundary of an area – we naturally think of attributes as being associated with the interior of an area rather than its boundary. Therefore, it is very common to store a centroid for each area, which is a point that is located inside the polygon as shown in Figure 7.7. The centroid can be defined by hand when the area is digitized, but many systems will automatically define one if this is not done. The centroid is commonly used to give a position for labels when drawing maps of area features and for this reason centroids are normally positioned near the centre of the area (as their name implies).

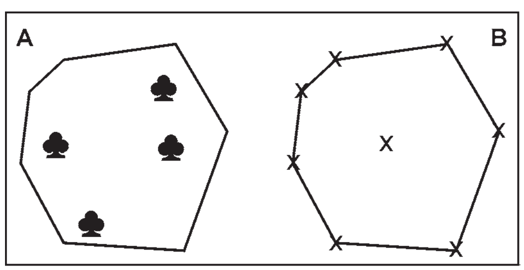


Figure 7.7: Storage of area feature.

The use of centroids means that to store a single area in our GIS, we actually need to store two things – the line defining the boundary and the point defining the centroid. In fact, things become more complicated still because so far we have only dealt with the simplest type of area. Figure 7.8 shows our original forest area as part of a land use map. Rather than a single area we now have a series of areas which neighbour each other, completely covering the area of the map. This type of map is very common – other examples are soil maps, geology maps and maps showing administrative areas such as states, or districts.

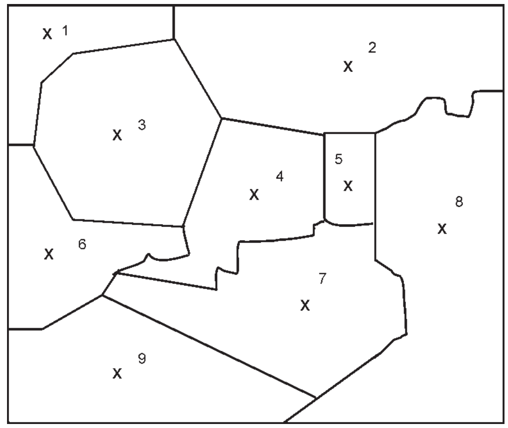


Figure 7.8: Land use map – an example of multiple areas.

Each area has a centroid, with an identifier associated with it, and this identifier is used as the link to a table containing the attributes for the areas (Table 7.10).

We can still use the simple method of storing the areas shown on Figure 7.7 but we will run into a number of problems. If we consider polygon 2, we can see that this shares part of its boundary with our original forest between points A and B on the map. However, although we have already stored this part of the line in storing the woodland boundary, we have to store it again, otherwise there will be a gap in the boundary of polygon 2. If we look at the whole map we will see that the majority of the boundary lines lie between two areas, and will be stored twice in this way-the result is that we will store nearly twice as much data as necessary.

This is not the only problem. When we store the boundary, we choose a series of points along the line, and connect these by straight lines. When the same line is digitized a second time, slightly different points will be chosen, with the result shown in Figure 7.9.

Table 7.10: Attribute table for land use map.

Polygon ID	Land Use
1	Vacant Land
2	Vacant Land
3	Forest
4	Urban
5	Agriculture
6	Vacant Land
7	Vacant Land
8	Forest
9	Agriculture

This shows the part of the boundary of polygon 2 between points A and B in black and the same part of the boundary of polygon 3 in grey. Because the two lines do not coincide, there are small areas of overlap, and small gaps between the two areas. These mismatches areas are called sliver polygons, because they are usually very small and thin.

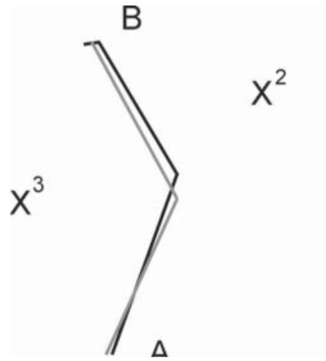


Figure 7.9: Sliver polygons as a result of digitizing the same line twice.

There is a third problem with this method of storing area boundaries which arises if we wish to use our data for analysis rather than simply map drawing. We may wish to produce a new GIS layer which simply shows urban and non – urban areas. To do this we have to merge together all the polygons in which the land use is not urban – *i.e.,* to dissolve the boundaries between them resulting in a new layer looking like the one in figure 7.1.

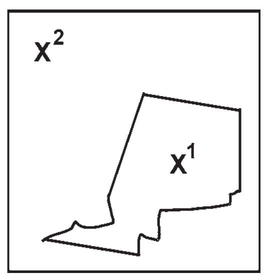
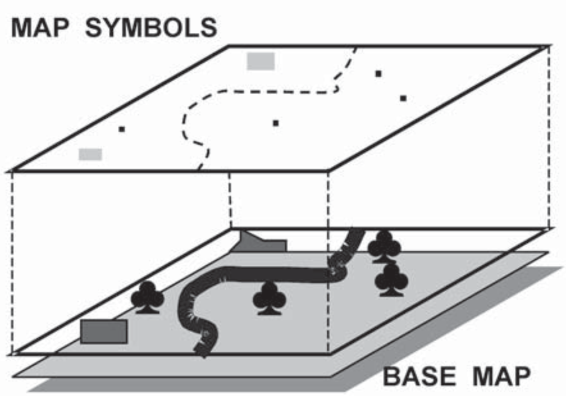


Figure 7.10: Map of urban and non-urban areas created using a polygon dissolve operation.

This operation is called a polygon dissolve, and is quite common in GIS analysis. However, it is difficult to do with the simple method of storing the area boundaries. If we consider our original forest, we do not know which part of the boundary has an urban area on the other side (so that we will need to keep that part of the boundary) and which has non-urban land use on the other side (so that we can drop that part of the line). In technical terms, we do not have any information about the contiguity of our polygons – which ones are neighbours to each other – and to store this information we need a different method of storing our area boundaries. However, it should be said that the method of storing areas described here has one great advantage, which is its simplicity. In addition, it is perfectly adequate for many of the operations we need to have in a GIS – we can store attributes for our areas, measure their aerial extent and perimeter, and produce maps based on them. This is why this method of storing areas is very common in mapping packages.

Some systems tie a polygon's ID code to a characteristic point in the polygon, known as the label point. Label points may be computed or identified interactively on screen, and codes may be entered manually for the relevant polygons. The attribute values of the polygon are then linked to this label point. Today, systems are available which treat polygons as independent objects. Plotting may be controlled by appending drawing instructions to the thematic code, to the individual identifiers, or to other object attribute values. In a finished map, tabular data appear on a foreground map against the background of a base map derived from the remaining map data. Look-up tables are usually used to translate tabular data to map symbols.



Look-up Table

Thematic Code	Map Symbols
10	Dots
20	Dotted Lines
30	Shaded Area

Table

ID	Thematic Code	Map Coordinates
Tree	10	X_1, Y_1
River	20	X_2, Y_2
Builtup Land	30	X_3, Y_3

Figure 7.11: Drawing instructions are designated in look-up tables. Thematic code values or attribute values are input values in the tables, while output values can be symbol types, colours, line thickness etc.

Spaghetti Model

Digital map data comprise lines of contiguous numerals pertaining to spatially referenced points. Spaghetti data are a collection of points and line segments with no real connection (Figure 7.12). What appears as a long, continuous line on the map or in the terrain may consist of several line segments which are to be found in odd places in the data file. There are no specific points that designate where lines might cross, nor are there any details of logical relationships between objects. Polygons are represented by their circumscribing boundaries, as a string of coordinates so that common boundaries between adjacent polygons are registered twice (often with slightly differing coordinates). The lines of data are unlinked and together are a confusion of crossings.

Unlinked (spaghetti) data usually include data derived either from the manual digitizing of maps or from digital photogrammetric registration. Consequently, spaghetti data are often viewed as raw digital data. These data are amenable to graphic presentation-the delineation of borders, for example – even though they may not form completely closed polygons. Otherwise, their usefulness in GIS applications is severely limited.

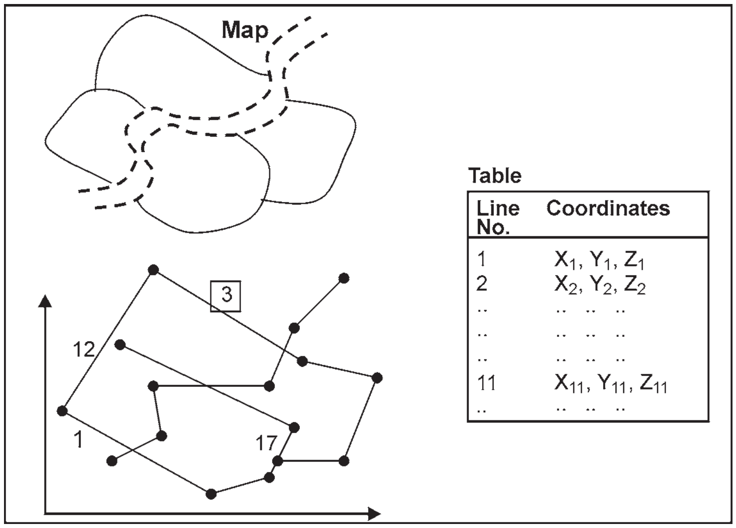


Figure 7.12: Spaghetti data is a term used to describe digital map data with crossing lines, loose ends, double digitization of common boundaries between adjacent polygons, etc. These data lie in a pile, just like spaghetti. Several line segments are found at odd places in the data file.

One drawback is that both data storage and data searches are sequential. Hence search times are often unduly long for such routine operations as finding commonality between two polygons, determining line intersection points, or identifying points within a given geographical area. Other operations vital in GIS, such as overlaying and network analysis, are intractable. Furthermore, unlinked data require an inordinate amount of storage memory because all polygons are stored as independent coordinate sequences, which mean that all lines common to two neighbouring polygons are stored twice.

Storing Area Boundaries: The Topological Approach

To overcome the limitations of the simple method of storing polygons, GIS systems draw on ideas first developed in a branch of mathematics called ***topology***, topology can be broadly explained as the way in which area data is stored in GIS systems.

If we look at the land use map in Figure 7.13, we can see that each area boundary is made up of a series of line sections, which meet at junctions such as the points marked A and B. If we identify each of those junctions, and store the location of the line section joining them, we will have stored all the boundary information, but without duplication.

The junctions are called nodes, and the sections of lines between them have various names in the GIS literature – arcs, chains, segments, edges and links. Since this method of storing area boundaries is often called link – and – node, we will use the term link.

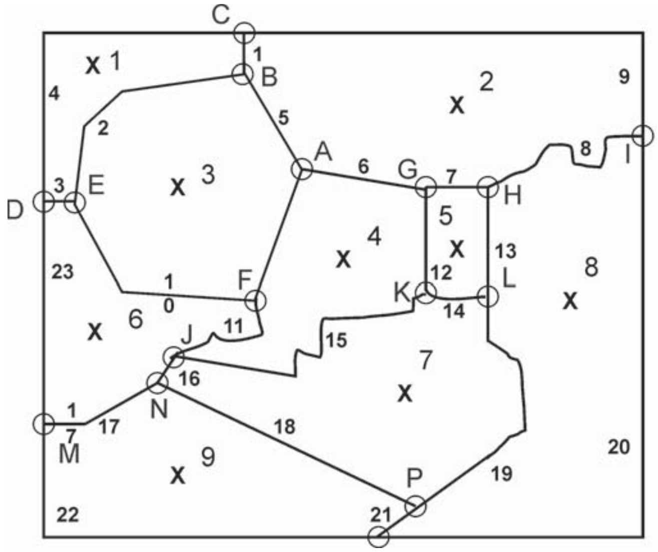


Figure 7.13: Land use map as stored in a topological GIS.

In figure 7.13, each link has been given a number (in italic numerals) and each node identified by a letter. For each link, we can store the location of the end points (the nodes) and of a series of positions along the line recording its location. However, we also need to know which link belongs to which areas, remembering of course that most links form part of the boundary of two polygons. This is done by storing for each link the identifier of the polygons on either side. For this purpose, the links must have a direction so that we know which side of the line a polygon is on-so for link 5 for example we might store (Table 7.11).

Table 7.11: Information stored for a link in a link and node data structure.

Link ID	Start Node	End Node	Left Polygon	Right Polygon	X_1	Y_1	X_2	Y_2	X_n	Y_n
5	A	B	3	2						

The direction of the link is not important – we could equally well store this link as running from B to A with polygon 2 on the left and 3 on the right. The same information is stored for every link on the map. Links such as number 4 which form part of the border of the map only have a polygon on one side – here it is customary to assign a number to the area ‘outside’ the map, so that link 4 might be recorded as having area 1 on one side and area 0 on the other .

Since we know which areas lie on either side of each link, we can use this information to construct the boundary of any area. For example, to construct the boundary of polygon 3 (our original forest area), we go through the complete set of links identifying all those with 3 to the left or right. We then attempt to join them together into a closed boundary by matching the node identifiers. For example, if we start with link 5, this ends at node B we then need a link which joins it at node B. Depending on the direction it has been digitized, this may start or end with node B, so we must check for both. We would continue this process until we arrived back at node A. In Figure 7.13, this process would fail because one of the links bordering polygon 3 has not been given an identifier and so will not be stored in the database. This illustrates one of the strengths of the link and node structure it can be used to check for errors in the data (which is how this structure originated).

The structure also solves our other problems. First, each link is only stored once, thus saving the problem of duplicating data. Second, we now have information about the relationship between areas, which can be useful for analysis. For example, if we wish to expand the urban area in polygon 4, it is simple to identify which land parcels border it, and which might therefore be considered as potential development sites.

The same link and node structure can also be used for line data, where it is the connections between lines at the nodes which is the important element the left/right area identifiers are generally ignored. However, knowing about connections between lines means that we can trace routes through networks.

The key to the link and node method of storing area boundaries is in the information which describes the relationships between objects on the map – the left/right identifiers and the to/from node identifiers for each link. Because of the origin of this structure in the mathematical subject of topology, this information is often described, somewhat loosely, as topological data. The X, Y coordinate pairs giving the location of points along the line are collectively known as the geometrical data. The two are shown together in the table for link 5 above, but in some systems they are actually stored separately. This means that in many GIS systems, the storage of area data is rather complex since each area requires one centroid and one or more links describing the boundary. In turn, each link will require one entry for the topological data plus one for the geometrical data.

So What Actually is Topology?

The study of relationships such as contiguity (whether objects are neighbours or not) is part of the mathematical subject of topology, which is concerned with those characteristics of objects which do not change when the object is deformed. For example, imagine the land use map shown in Figure 7.13 printed on a sheet of thin rubber. If the rubber were stretched, then some properties of the areas would change, such as their size and shape. However, no amount of stretching could make polygon 3 border polygon 7 – this would involve cutting the sheet or folding it over on itself. Hence, the connection (or lack of it) between areas is a topological property of the map. Containment is another example of a topological property, since no amount of stretching will move centroid 3 outside its polygon. One of the earliest people to study such properties was the Swiss mathematician Leonhard Euler and one of the classic problems he studied, the Konigsberg bridge problem, has a direct relevance to the use of topological ideas in GIS. In the town of Konigsberg, there was an island in the Pregel River, which was connected to the banks of the river by seven bridges.

The local people believed that it was impossible to plan a route which started and ended in the same place but crossed every bridge only once. However, nobody could prove whether this was in fact correct, Euler realized that the problem had nothing to do with the distances or directions involved, but depended solely on the connections between places. He reformulated the problem, by representing each of the land masses as points, and the connections between them as lines (Figure 7.14).

This representation is called a graph by mathematicians. The key to the problem is the number of lines which meet at any given vertex of the graph – if this is an even number we can reach that vertex and leave by a different line. If it is an odd number, then eventually the pairs of entry/exit lines will be used up and there will be only one unused line joined to that vertex – *i.e.*, we can visit the vertex but can't leave again without using a line for the second time. Therefore, it is only possible to make a round trip walk if all vertices have an even numbers of lines, or if there are just two vertices at which an odd number of lines meet (which will have to be the start and end points of the route). In the case of the Konigsberg bridges neither condition is satisfied, proving that the round trip cannot be made and that the locals were right.

Another mathematician, Henri Poincaré, realized that graph theory could be applied to maps in general, and his ideas were used by staff at the US Bureau of the Census to help in processing the data for the 1980 census. A key part of processing the census data was in handling the map of the street network of the United States of America. This indicated where each address was located and which block it fell in. In order to automate the processing of census data, it was necessary to have an accurate database which indicated which block each address was in. Compiling such a database was an enormous task in which errors were bound to be made, and so some way was needed for checking the accuracy of the results.

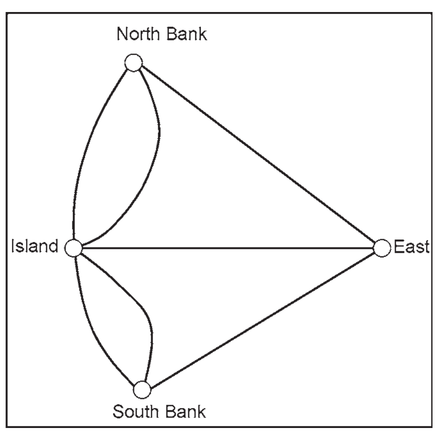


Figure 7.14: Sketch of the Konigsberg bridge problem.

The map in Figure 7.15 is a fictitious example of part of the street network in an American city. It can be seen that each block is surrounded by sections of street which meet at junctions, if we treat each street intersection as a vertex, we can regard the street network as a mathematical graph. What is more, if we consider the part of the graph which surrounds an individual block (*e.g.,* Block 5 in the above figure) it is clear that this will form a single connected circuit.

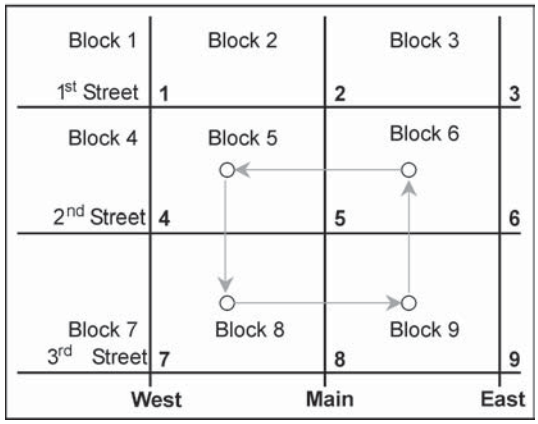


Figure 7.15: Fictitious city blocks illustrating Poincarés dual graph model of maps.

However, we can also use a graph to represent the relationship between the blocks. First, we represent each block by a single point as shown with blocks 5, 6, 8 and 9 in Figure 2.20. The points are then joined by lines if they lie on either side of a street section and we have a graph similar to the one constructed by Euler for the Konigsberg bridge problem. If we take the portion of this graph which surrounds a single street intersection (like the example shown in grey in Figure 7.15 which surrounds node 5), then this should form a single connected circuit as with the first graph.

We can therefore construct two graphs – one based on the streets surrounding a block, one on the blocks surrounding a street intersection, and it was this model of maps as paired graphs which came from the work of Poincaré. Mathematically, the two graphs are exactly the same, since both will consist of a single closed loop. If we can automatically create these graphs from our street map, and check that they do form closed loops we will have a way of checking the data. This is exactly what the staff at the Bureau of the Census managed to do when they developed a data structure called DIME.

How it Works? The DIME Example

So far we have considered these graphs by drawing them on the original map. However, computers cannot 'see' maps, so we must devise a method of storing the data in a format which allows the computer to construct and test the graph around any block or junction. The system which was devised was called DIME – (Dual Independent Map Encoding) and was based upon a data structure in which one record was created for each street segment the part of a street between two junctions. For the map in Figure 2.20, a portion of the DIME – file might be as shown in Table 7.12.

Table 7.12: DIME data structure for fictious city map.

Segment	From	To	Block Left	Block Right	Street Name
1	1	2	2	5	1st
2	2	3	3	6	1st
3	4	5	5	8	2nd
4	5	6	6	9	2nd
5	7	8	8	11	3rd
6	8	9	9	12	3rd
7	4	1	4	5	West
8	7	4	7	8	West
9	5	2	5	6	Main
10	8	5	8	9	Main

If we look at block 5 in Figure 7.15, it is very simple for us to see that it is surrounded by four connected street segments. In order to check this on the computer using the data structure we first find those records in the DIME file which have 5 on either their left or right. Since the DIME file is simply a table, this could be done using a standard database query, or by writing a program to read each record and select those in which the block left or block right value was 5. In this case, we will find segments 1, 3, 7 and 9. We then need to check that these segments form a closed loop, and this is most easily done if they all go in the same direction around the block. If they all have block 5 on their right hand side, this means they will form a clockwise loop. At the moment, only segments 1 and 7 have 5 as their right hand block–however to change the direction of the other two all we need to do is switch the left and right blocks and then to and from nodes to produce the records shown in Table 7.13.

Table 7.13: Records from DIME file related to block 5. All records are modified so that block 5 is on right hand side of the street.

Segment	From	To	Block Left	Block Right	Street Name
1	1	2	2	5	1st
3	5	4	8	5	2nd
7	4	1	4	5	West
9	2	5	6	5	Main

We can now start at any segment and try and trace round the loop. We start at segment 1, noting that our start point is node 1. The end of this segment is node 2, so we look for a segment which starts at node 2. We find segment 9, which ends in node 5. This leads us to segment 3 which ends in node 4, and this leads us to segment 7, which ends in node 1, our starting point. If for any reason we can't complete this loop, we know there is an error in the data, such as a segment completely missed out, or one in which the block numbers or node numbers were wrong. For example, if segment 2 had 5 as its right block instead of 6 this would create three errors:

- Block 3 would not be correct because of the 'missing' segment.
- Block 6 would not be correct because of the 'missing' segment.
- Block 5 would close but would have a segment unused.

This checking process can, also be carried out using the graph based around the street junctions. If we consider junction 5, then we can identify the segments which meet at this point because they will have 5 as either their start or end node (Table 7.14) .

We then adjust these so that they all end at node 5 as shown in Table 7.15. Now, if we start at a segment, then the block on the right of that segment must be the left hand block of one other segment, which in turn will share its right hand block with one other segment, until we work our way around to the starting point.

Table 7.14: Records from DIME file relating to junction 5.

Segment	From	To	Block Left	Block Right	Street Name
3	4	5	5	8	2nd
4	5	6	6	9	2nd
9	5	2	5	6	Main
10	8	5	8	9	Main

In both the block and junction checking, we are using the left/right and from/to information to trace around the topological graph. Since we know from mathematical theory that there must be one closed graph in each case, if we fail to find this, we know we have an error in the data.

Table 7.15: Records from DIME file relating to junction 5 modified so that the street ends at junction 5.

Segment	From	To	Block Left	Block Right	Street Name
3	4	5	5	8	2nd
4	6	5	9	6	2nd
9	2	5	6	5	Main
10	8	5	8	9	Main

Notice that the original DIME file did not contain any geographical coordinates. The geographical referencing was done via addresses, since the original file had extra fields which have not been shown so far, which indicated the range of addresses on either side of the segment. This allowed any address to be matched with its block number (and census tract number), and also allowed summary statistics to be produced for any block by aggregating data for all the addresses (Table 7.16).

Table 7.16: Storage of address information in the DIME data structure.

Segment	From	To	Block Left	Block Right	Street Name	Left Address Low	Left Address High	Right Address Low	Right Address High
1	1	2	2	5	1st	12	24	13	25

We have now seen how ideas from topology lead to the DIME data structure. Since this was intended simply for handling data for streets, the segments in the DIME file were all straight – if a street curved, it was simply broken up into segments. In order to develop this into a more general purpose data structure, it was necessary to allow the lines between the nodes to take on any shape, as described by a set of XY coordinates, in this way we reach the link and node data structure.

Topology Model – Connections and Relationships between Object

Topology deals with geometric properties which remain invariable under certain transformations, such as stretching or bending. The topology model is one in which the connections and relationships between objects are described independent of their coordinates; their topology remains fixed as geometry is stretched and bent. Hence, the topology model overcomes the major weakness of the spaghetti model, which lacks the relationships requisite to many GIS manipulations and presentations.

The topology model is based on mathematical graph theory and employs nodes and links. A node can be a point where two lines intersect, an endpoint on a line, or a given point on a line. For example, in a road network the intersection of two roads, the end of a cul-de-sac, or a tunnel edit may generate a node. A link is a segment of a line between two nodes. Links connect to each other only at nodes. A closed polygon consisting of alternating nodes and links forms an area. Single points can be looked upon as a degenerate node and as a link with zero length (Laurini and Thompson, 1992). Theme codes should be taken into consideration when creating nodes to ensure that they are created only between relevant themes (*e.g.,* at the junction between a national highway and a state highway, not between roads and property boundaries).

Unique identities are assigned to all links, nodes, and polygons, and attribute data describing connections are associated with all identities. Topology can therefore be described in three tables (Figure 7.16):

i. The polygon topology table lists the links comprising all polygons, each of which is identified by a number.
ii. The node topology table lists the links that meet at each node.
iii. The link topology table lists the nodes on which each link terminates and the

polygons on the right and left of each link, with right and left defined in the direction from a designated start node to a finish node. The system creates these tables automatically.

A table with point coordinates to the links ties these features to the real world and permits computations of distances, areas, intersections, and other numerical parameters. The geometry of the objects is stored in its own subordinate table (see Figure 7.16). Numerous spatial analyses may then be performed, including:

- Overlaying
- Network analyses
- Contiguity analyses
- Connectivity analyses

Topological attribute data may be used directly in contiguity analyses and other manipulations with no intervening, time – consuming geometric operations.

Once the topology has been created, a map can be plotted with solid colours. This is not possible with spaghetti data. Thematic layers of topological data can also be used to steer the plotting sequence. The sequence influences what becomes visible on the map. For example, a green area superimposed on a white house will render the house invisible on the map (unless the house creates a window in the area). Topology requires that all lines should be connected, all polygons closed, and all loose ends removed. Even gaps as small as 0.001 millimetre may be excessive, so errors should be removed either prior to or during the compilation of topological tables.

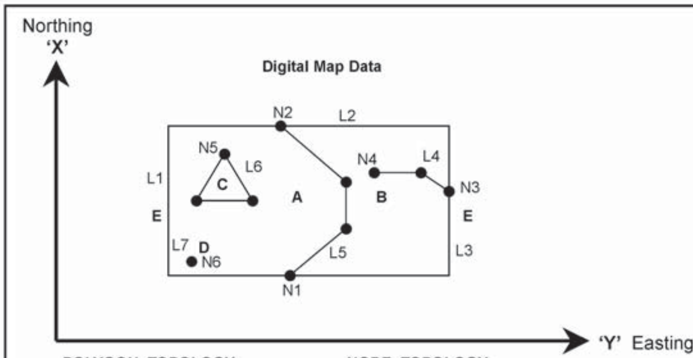
A function known as snap can also be used in digitalization. Using the snap function with a defined tolerance of, say, 1 millimetre, a search can be carried out around the end of a line or around an existing point which is assumed to have the same coordinates as the last point registered. When this point is found, the two points will be snapped together to form a common node, thereby closing the polygon. The same procedure can be carried out automatically on existing data. A node can also be created in existing data by calculating the point of intersection between lines. Meaningless loose ends can be removed by testing with a given minimum length.

Topological information permits automatic verification of data consistency to detect such errors as the incomplete closing of polygons during the encoding process. The graph theory contains formulas for the calculation of such data errors. There has to be a fixed relationship between the number of nodes, lines, and polygons in one data set. A run-through of the data in positive and negative directions will produce the same result.

The topological model has a few drawbacks. The computational time required to identify all nodes may be relatively long. Uncertainties and errors may easily arise in connection with the closing of polygons and formation of nodes in complex networks (such as in road interchanges). Operators must solve such problems. When new data are entered and existing data updated, new nodes must be computed and the topology tables brought up to date.

Topological data may require a longer plotting time than spaghetti data because of the separation of lines into nodes and links. However, the overall advantages of the topology model over the spaghetti model make it the prime choice in most GISs. Today, efficient software and faster computers enable topology to be established on-the-fly; thus the disadvantages of topological data as compared to spaghetti data have become less important.

Usually, map data are not stored in a contiguous unit, but rather, divided into lesser units that are stored according to a selected structure. This structure may be completely invisible to the user, but its effects, such as rapid screen presentation of a magnified portion of a map, are readily observable.



POLYGON TOPOLOGY

Polygon	Links
A	L1, L5
B	L2, L3, L5
C	L6
D	L7
E	L1, L2, L3

NODE TOPOLOGY

Node	Links
N1	L1, L3, L5
N2	L1, L2, L5
N3	L2, L3, L4
N4	L4
N5	L6
N6	L7

LINK COORDINATES

Link	Coordinates			
L1	4, 10	4, 4	11, 4	11, 9
L2	11, 9	11, 16	8, 16	
L3	8, 16	4, 16	4, 10	
L4	8, 16	9, 15	9, 13	
L5	11, 9	8, 11	6, 11	4, 10
L6	10, 7	7, 8	7, 5	10, 7
L7	5, 5			

LINK TOPOLOGY

Link	Start Node	End Node	Left Polygon	Right Polygon
L1	N1	N2	E	A
L2	N2	N3	E	B
L3	N3	N1	E	B
L4	N3	N4	B	B
L5	N2	N1	B	A
L6	N5	N5	A	C
L7	N6	N6	A	A

Figure 7.16: Topology model have geometric objects. Digital map data are represented by nodes and links. The objects attributes and relationships are described by storing nodes and links in tables, i.e. Polygon Table; Node Topology Table; Link Topology Table and an additional table showing Objects Geographical Coordinates.

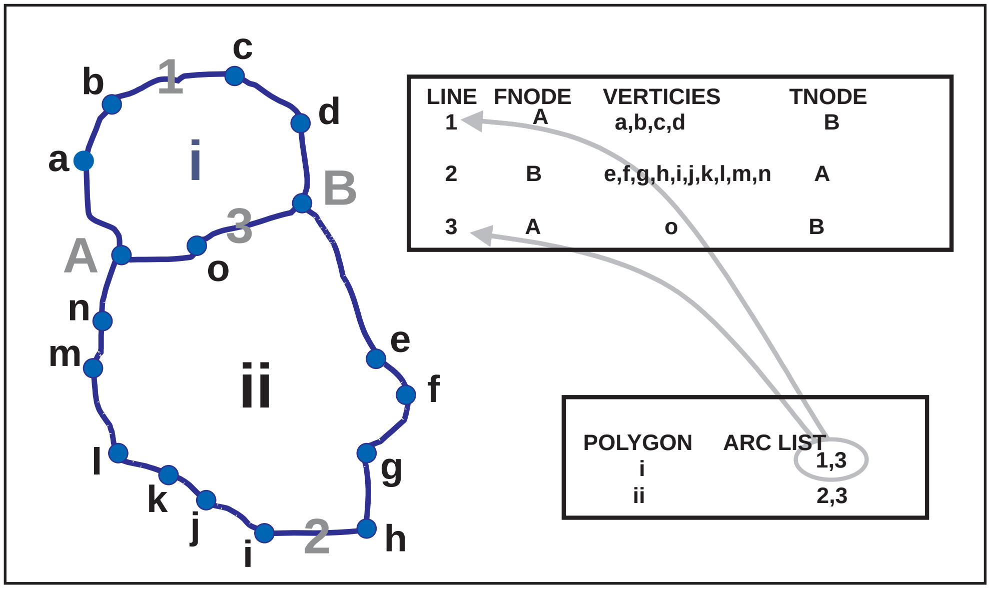


Figure 7.17: Representing polygon in a topology.

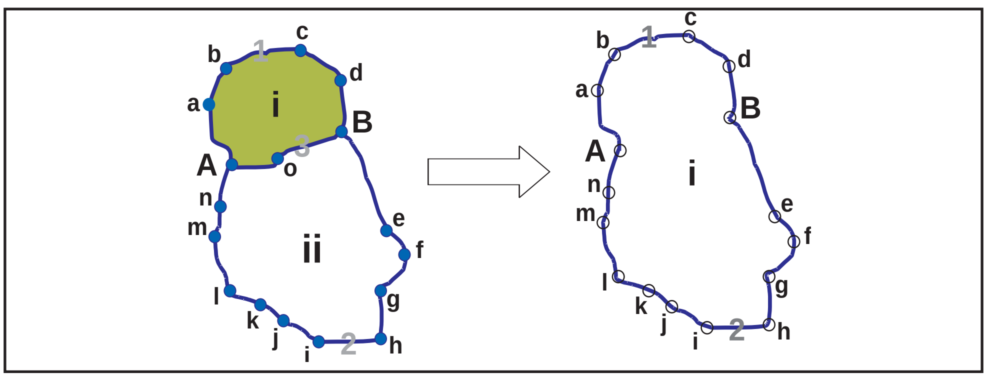


Figure 7.18: Impact of editing in topology.

Data Compression

The amount of memory needed can be reduced by using data compression techniques. Most of these automatic techniques are based on removing points from continuous lines (contour lines, etc.). Good data compression techniques, therefore, are those that preserve the highest possible degree of geometric accuracy. The most basic technique involves the elimination of repetitive characters: for example, the first character of all coordinates along a particular axis. The repetitive character needs to be entered only once; subsequently, it may be added to each set of coordinates. This particular technique has no effect on the geometry. Savings in characters stored are illustrated in Table 7.17.

Table 7.17: Simple data compression. The volume of data to be stored is reduced to a single entry, assigning the value common to all coordinate values.

Original Data		Compacted Data	
Northing	Easting	10000	80000
10,234	80565	234	565
10245	80598	245	598
10167	80324	167	324
--	--	--	--
--	--	--	--
--	--	--	--

There are other automatic methods of removing points. One simple means is to keep only every nth point on a line. The lower the value of n, the greater the number of points

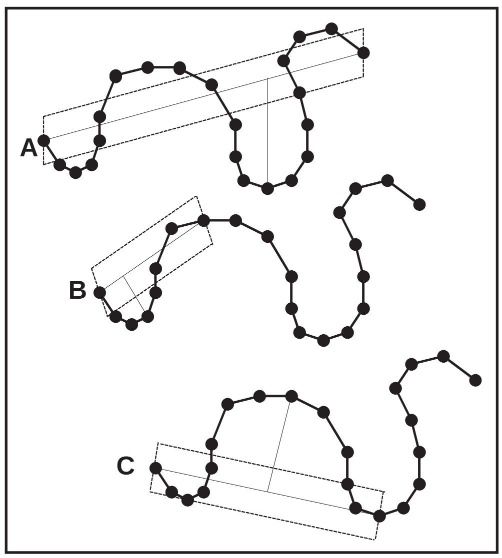


Figure 7.19: Douglas-Peucker Method, this helps in saving storage space.

that will be removed. This method does not take into account geometric accuracy; however, this can be compensated for by testing the curvature of the line. One method is to draw a straight line between the first and last points on a curved stretch of line and to calculate the orthogonal distance from each point on the curved line below the straight one. Points that are closer than a given distance from the straight line will be removed. The endpoint of the straight line is then moved to the point with the greatest distance and the same procedure for removing points is repeated. This continues until all the relevant points are removed. This method is known as the Douglas-Peucker algorithm (Figure7.19).

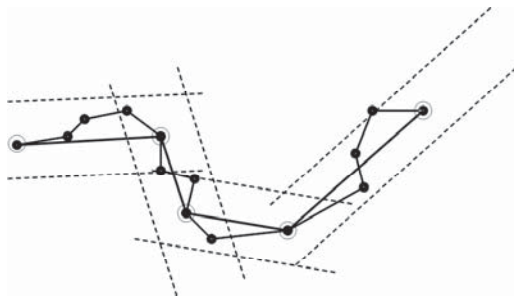


Figure 7.20: Example of reducing lines by the use of corridor. The numbers of points needed to describe a line are reduced by moving forward a corridor of a given width until it touches the digitized line. All the points on the line in the corridor, apart from the first and last, are deleted. This process is repeated until the entire line is trimmed.

Points of little or no value in describing a line may be eliminated by moving a corridor step by step along a line and deleting points that are closer to the neighbouring point than a given value or where the vectors create an angle that is smaller than the given value (Figure 7.20). Contours and other lines can also be replaced with mathematical functions, such as straight lines, parabolas, and polynomials. A spline function comprises segments of polynomials joined smoothly at a finite number of points so as to approximate a line. A spline function can involve several polynomials to build a complex shape. It has been reported that a spline function representing nautical chart data has reduced data volume by 95 %.

The amount of memory required to store a given amount of data often depends on the format in which data are entered. Some formats contain more administrative routines than others, some have vacant space. Thus, the gross volumes stored are frequently related to format.

Storing Vector Data

The information content of the data is designated not in the format but ancillary to it, for example, in a heading. Typical specifications for information content might include field assignments, such as the point number in the first field, the thematic code in the second, easting in the third, northing in the fourth, and elevation in the fifth. The meanings of the numeric codes used must also be given. The spaghetti data are stored in a simple file structure and in the order in which the data have been registered.

Users of conventional maps know the frustrations of extracting information from maps produced by various agencies using differing map sheet series, varying scales and coordinate systems, and frequently, different symbols for the same themes. Moreover, the cartographic version of Murphy's Law dictates that the necessary information is all too often located in the comers where four adjoining map sheets meet.

Database storage of cartographic data can overcome these problems because it involves standardization of data through common reference systems and uniform formats. Cartographic data from various sources can, with few limitations, be combined. The results are then independent of map sheet series and scales.

Standardized storage makes the presentation of data compiled from dissimilar sources much easier. For example, uniform storage formats permit the combination of telecommunications administration network data with property survey data, or of geological information from 1:50,000 scale maps with vegetation data from 1:20,000 scale maps.

Digital map data are stored in databases, the computerized equivalent of conventional file drawers and cabinets. Although data entries in a database can be updated far more rapidly than data printed on map sheets on file, the information is found more quickly from map sheets than by searching in a database. This is because a single map sheet contains an enormous amount of information, usually equivalent to 100,000 or more sets of coordinates. A sequential computer search of 100,000 items in a database is slow even for the most powerful computers in comparison with a quick visual scan of a map sheet. Therefore, 'smart' programs known as database management systems (DBMSs) have been compiled to maintain, access, and manipulate databases. The various DBMSs differ primarily in the ways in which the data are organized. Their selection and use are vital in GIS applications because they determine the speed and flexibility with which data may be accessed.

It is usual to split topological data into different thematic layers to simplify storage and to improve access to data. This division is done so that no overlap occurs between polygons within each thematic layer. For example, property boundaries are stored in one layer while other data overlapping the property, such as roads, buildings, and vegetation boundaries, are stored in another. The disadvantage of this system is that common lines between objects (*e.g.,* roads and properties) that are stored in different layers have to be removed several times. This problem can be avoided by using object-based storage.

The Choice Between Spaghetti and Topology Models

When digitizing lines such as those on land-use maps, the borders of surfaces are digitized both as spaghetti data and as separate objects. When creating topology, this model is converted to a layer model. The discussion of spaghetti and topology is very much based on the assumption that a class of area entities is always a tiling of the plane in which every point lies in exactly one polygon. However, the problems related to spaghetti and topology have changed somewhat during recent years with the advent of new GIS software which treats polygons as independent objects that may overlap and need not fill the plane, and with systems permitting shapes. Many of th e traditional arguments for area coverage/layer model and use of topology are based on the assumption of needing to avoid computation. New and more powerful computers eliminate the need for reduction in calculation time. Today, topology can easily be built on-the-fly.

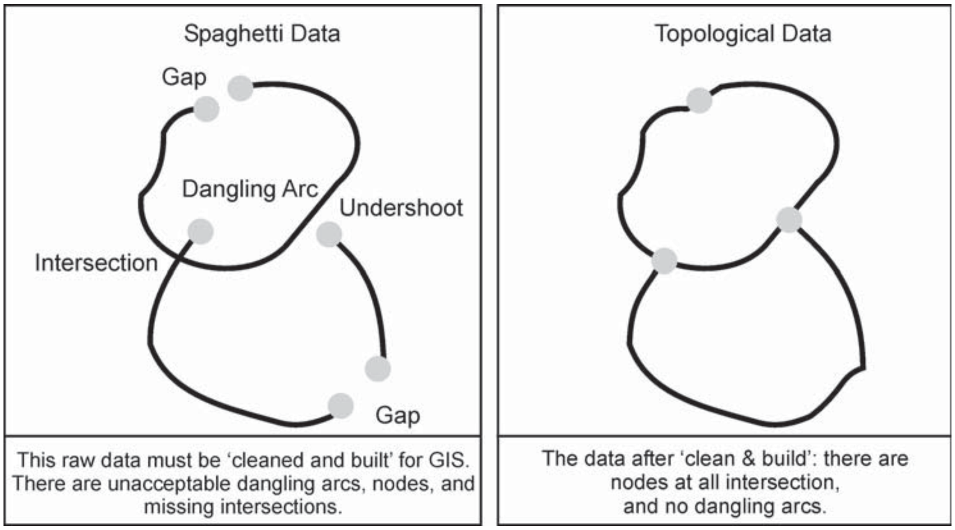


Figure 7.21: Spaghetti data versus topological data.

RASTER DATA MODELS

Raster data are applied in at least four ways:

i. Models describing the real world
ii. Digital image scans of existing maps
iii. Compiling digital satellite and image data
iv. Automatic drawing driven by raster output units

In the first example, raster data are associated with selected data models of the real world: in the second and third, with compilation methods, and in the fourth, with presentation methods.

RASTER MODEL

The raster model represents reality through selected surfaces arranged in a regular pattern. Reality is thus generalized in terms of uniform, regular cells, which are usually rectangular or square but may be triangular or hexagonal. The raster model is in many ways a mathematical model, as represented by the regular cell pattern (Figure 7.22). Because squares or rectangles are often used and a pictorial view of them resembles a classic grid of squares, it is sometimes called the grid model. Geometric resolution of the model depends on the size of the cells. Common sizes are 10 × 10 meters, 100 × 100 meters, 1 × 1 kilometer, and 10 × 10 kilometers. Many countries have set up national digital elevation models based on 100 × 100 – meters cells. Within each cell, the terrain is generalized to be a flat surface of constant elevation.

The rectangular raster cells, usually of uniform size throughout a model, affect final drawings in two ways. First, lines that are continuous and smooth in a vector model will become jagged, with the jag size corresponding to cell size. Second, resolution is constant: regions with few variations are as detailed as those with major variations, and vice versa.

The cells of a model are given in a sequence determined by a hierarchy of rows and columns in a matrix, with numbering usually starting from the upper left comer (Figure 7.23). The geometric location of a cell, and hence of the object it represents, is stated in terms of its row and column numbers. This identification corresponds to the directional coordinates of the vector model. The cells are often called pixels (picture elements), a term borrowed from the video screen technology used in television and computer displays. A pixel is the smallest element of an image that can be processed and displayed individually. The raster techniques used in GIS are siblings of the rasters long used to facilitate the manipulation and display of information and consequently are suited to computerized techniques.

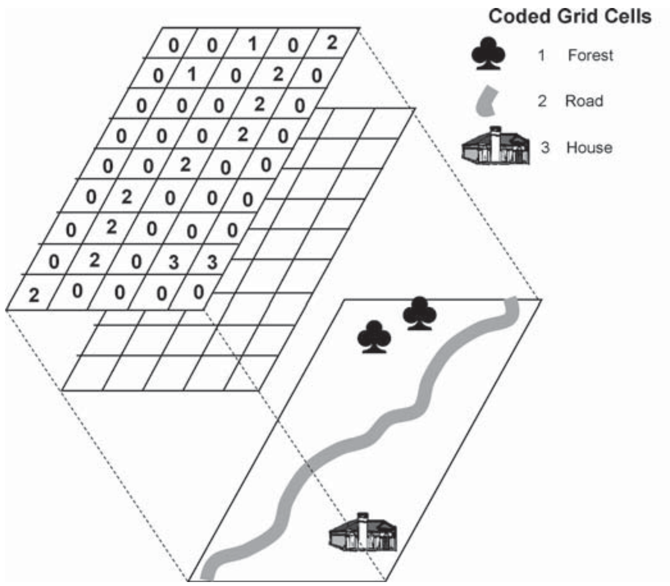


Figure 7.22: Raster data can be visualized as a grid lying over the terrain. Each grid cell has a code stored in the database describing the terrain within that particular cell.

Realizing the Raster Model

Raster models are created by assigning real-world values to pixels (Figure 7.23). The assigned values comprise the attributes of the objects that the cells represent-and because the cells themselves are in a raster, only the assigned values are stored. Values, usually alphanumeric, should be assigned to all the pixels in a raster. Otherwise, there is little purpose in drawing empty rows and columns in a raster.

Consider a grid of cells superimposed on the ground or on a map. Assigning the values/ codes of the underlying objects/features to the cells creates the model. The approach is comprehensive because everything covered by the raster is included in the model. Draping a ground surface in this way regards the ground or map as a plane surface.

Some GISs can manipulate both numerical values and text values (such as types of vegetation). Hence cell values may represent numerous phenomena, including:

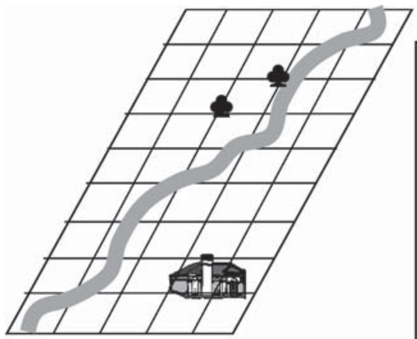
- Physical variables, such as precipitation and topography, respectively, with amounts and elevations assigned to the cells
- Administrative regions, with codes for urban districts, statistical units, and so on
- Land use, with cell values from a classification system
- References to tables of information pertaining to the area(s) the cells cover, such as references to attribute tables
- Distances from a given object
- Emitted and/or reflected energy as a function of wavelength – satellite data.

A single cell may be assigned only one value, so dissimilar objects and their values must be assigned to different raster layers, each of which deals with one thematic topic (Figure 7.24). Hence in raster models as in vector models, there are thematic layers for topography, water supply systems, land use, and soil type. However, because of the differences in the way attribute information is manipulated, raster models usually have more layers than those in vector models. In a vector model, attributes are assigned directly to objects. For instance, a pH value might be assigned directly to the object 'lake.' In a raster model, the equivalent assignment requires one thematic layer for the lake, in which cells are assigned to the lake in question, and a second thematic layer for the cells carrying the pH values. Raster databases may, therefore, contain hundreds of thematic layers.

In practice, a single cell may cover parts of two or more objects or values. Normally, the value assigned is that of the object taking up the greater part of the cell's area, or of the object at the middle of the cell, or that of an average computed for the whole of the cells.

Cell locations, defined in terms of rows and columns, may be transformed to rectangular ground coordinates, for example, by assigning ground coordinates to the center of the upper left cell of a raster (cell 0, 0). If the raster is to be oriented north-south, the columns are aligned along the northing axis and the rows along the easting axis. The coordinates of all cell corners and centers can then be computed using the known cell shapes and sizes.

Object relations, which in the vector model are described by topology, are only partly inherent in the raster structure. When the row and column numbers of a cell are known, the locations of neighbouring cells can easily be calculated. In the same way, cells contained in a given polygon may be located simply by searching with a stipulated value. It is much more difficult, however, to identify all the cells located on the border between two polygons. Polygon areas are determined merely by adding up constituent cells. Some operations, though, are more cumbersome. An example of this is computation of a polygon's perimeter length, which requires a search for, and identification of, all the cells along the polygon's border.



Cell Code

	0	1	2	3	4
0	0	0	1	0	2
1	0	1	0	2	0
2	0	0	0	2	0
3	0	0	0	2	0
4	0	0	2	0	0
5	0	2	0	0	0
6	0	2	0	0	0
7	0	2	0	3	3
8	2	0	0	0	0

Cell Table

Cell No.	Cell Value
00	0
01	0
02	1
03	0
04	2
10	0
11	1
–	–
–	–
–	–
–	–
70	0
71	2
72	0
73	3
74	4
–	–
–	–

Code List

0	Unmapped
1	Forest
2	Road
3	House

Figure 7.23: A line number and column number define the cell's position in the raster data. The data are then stored in a table giving the number and attribute value of each cell.

Overviews of phenomena in a given area are obtained from a raster model quickly and easily by searching all the thematic layers for cells with the same row and column numbers. Raster data are normally stored as a matrix, as described above. However, they can also be stored in tabular form, where each individual cell in a raster forms a line in the table.

Storing Raster Data Structures

Using a raster GIS we could store a set of spatial data in the form of a grid of pixels. Each pixel will hold a value which relates to some feature of interest at that point in space. These values are normally one of three possible types.

I. Binary – A value which indicates the presence or absence of a feature of interest. For example, in a layer representing roads, we might use 1 for pixels which contained part of road, and 0 for pixels which did not.

II. Enumeration – A value from some classification. For example, a layer representing soils might contain codes representing the different soil types–1 for alluvial, 2 for red soil etc. Since the values are not directly related to the soil type, there would have to be a key of some sort indicating the meaning of each value.

III. Numerical – An integer or floating point number recording the value of a geographical phenomenon. In the soil example, we might have measurements of soil moisture content. A common example of this kind of raster layer is when the values represent the height of the land surface, in which case the layer is often referred to as a Digital Elevation Model (DEM).

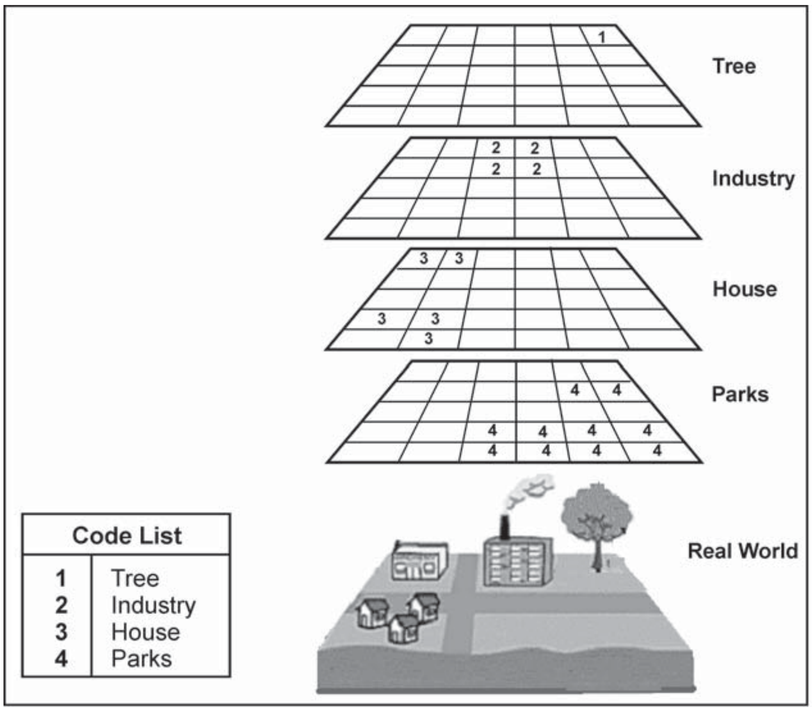


Figure 7.24: Only one attribute value is assigned to each cell. Objects that have several attributes are therefore represented with a number of raster layers, one for each attribute.

The raster data model has the great virtue of simplicity but it can produce very large files. The precision with which, 1 raster layer can represent spatial data is related to the size of the pixel – we cannot represent anything which is smaller than a pixel. This means that pixel sizes need to be small, but the result is very large raster grids. What is more, for many applications, even smaller pixel sizes arc desirable but halving the pixel size would increase the number of pixels by a factor of 4.

Raster Data Structure: The Array

The simplest method of storing a raster layer in the memory of the computer is using a data structure called an array. All programming languages have a structure called an array, which can be used to store and access lists of items. We consider alternative methods for searching through a list of entries in a telephone book to find one that matched with a particular name. The full list of names to be searched could be stored and accessed using an array as follows:

1. Array NAMES [1…... 64]
2. Read names from file into names array
3. i = 1
4. FOUND = false
5. repeat until FOUND = = true or i > 64
6. if NAMES [I] = = 'Sameer' then FOUND = true
7. i = i + 1

This is the brute force algorithm for searching the list. The first line sets up an array with space for 64 names and the actual names are read from a file into this array. At this point, the first few elements of the array might contain the following:

NAMES [1]	Sameer
NAMES [2]	Shaan
NAMES [3]	Krishna

Lines 5 to 8 go through this array, one item at a time, comparing the value with the name we are looking for – Sameer – until either this is found, or the entire list has been searched, Each element in the array is identified by a number and this number is used to retrieve the correct element from memory. The array is available in programming language, because there are so many cases where it is necessary to deal with collections of related pieces of information, It is also possible to have arrays which have both rows and columns, and these are what can be used to store raster data.

The array is also an extremely efficient data storage mechanism but to understand why, it is necessary to understand something of the way a computer operates. Everyone is familiar with the idea of the storage of data on secondary media, such as floppy disks, zip disks and CD-ROMs. However, in order to be used by the computer, the data must first be transferred from the disk into the computer's memory. The only part of the computer that can actually do anything with the data is the Central Processing Unit or CPU. CPUs vary greatly in the way they are made and what they can do, but almost all have one thing in common – they can only deal with a few pieces of information at a time. This may seem surprising. How can a computer perform complex tasks, if it can only deal with a few things at a time? To see how it is able to do this, imagine being asked to work out the following sum on a calculator – 300 + ((25 × 320)/100). The steps involved would probably be:

i. Multiply 25 by 320 – this gives 8000.
ii. Divide 8000 by 100 – this gives 80.
iii. Add 80 and 300 which gives 380 – the answer.

Here we notice that in each step just three numbers are involved, – two numbers which are input to the calculation, and one answer. It is always possible to break down problems in this way, into a series of steps which involve very few pieces of information, and this is exactly how computers are programmed to solve problems. The input data values are stored in the computer memory. From here the first two are passed to the CPU and operated on – 25 and 320 are multiplied to give 8000. This answer can be held in the CPU while the next input is fetched from memory. The 8000 is then divided by 100 to give 80. After the final step, the answer to the problem is passed back to the memory to be stored and allow the CPU to pass on to the next problem.

So why do we need memory? Why not simply pass the information directly between the secondary storage and the CPU? The answer is that this would be very slow, for three reasons. First, the transfer of information to and from disks along cables is inherently slower than transfer to and from memory which all takes place on printed circuit boards. Second, disks are mechanical devices which rotate, and this places an inherent limit on the speed with which information can be accessed. In contrast, memory works purely electronically. Third, finding an individual piece of information on a disk is relatively slow. Disks normally hold a large number of files, and once the correct file has been found, it is necessary to read through it to find the correct piece of data. In contrast, computer memory is designed to make it very easy to find individual data values.

Remember that each data value is held in one or more bytes of storage. If we have a file containing a set of numbers, we can visualize these as being held in a series of boxes, one number per box. In a disk file the individual boxes are not normally distinguished – when the file is read, every box is retrieved starting with the first. In contrast, memory is organized so that every box has what is called an address, which is basically a number which uniquely identifies it. So our list of names might look like this in memory (Figure 7.18).

Table 7.18: Storage of an array in computer memory.

Address	1295	1296	1297	1298
Contents	Sameer	Shaan	Krishna	

The circuitry in the computer is designed so that the information can be retrieved from any of the boxes equally quickly by passing the address to the CPU. It is rather as if the postal, service worked by having a direct connection between every individual house and the post office.

What a computer program has to do therefore is work out the addresses of the boxes which contain the data which it needs. In the case of the array this is extremely simple. When an array is set up in a program, the program takes a note of the address of the first element – in this case the name ‘Sameer’ which is stored in box 1295. The addresses of any of the other element – can then be worked out from the index number which is normally given in brackets after the name of the array. So when the program refers to NAMES [2] what the computer actually does is as follows:

1. The index value of this element is 2.
2. The first element in this array has an index value of 1.
3. This element is therefore (2–1) = 1 box on from the start.

4. The address of the first element is 1295.
5. The address of this element is therefore 1296.

This may seem long winded, but the computer only has to do two calculations – find how far along the array this element is (step 3) and use this to work out the actual address (step 5) – one subtraction and one addition. The calculation in step 3 produces what is sometimes called the offset – how far the element is from the start of the array. In many programming languages, the first array element is labelled 0 so that it is not necessary to perform the calculation in step 3 – the offset is simply the number of the element which can be added directly to the start address.

3	A	A	A	A
2	A	B	B	B
1	A	A	B	B
0	A	A	A	B
	0	1	2	3

Figure 7.25: An example of simple raster array.

To see how this relates to GIS, let us consider the simple raster layer shown in figure 7.26. Note that for clarity the pixel values are shown as letters which will help to distinguish them from the numerical memory addresses in the explanations which follow. In practice, most GIS systems only allow the storage of numerical data in pixels.

Instead of a list of names, we now have a set of rows and columns. When we want to identify a particular element in the array we will need to give both a row and column number – for instance IMAGE [3,3] to refer to the element in the top right hand corner. So does this mean we need a special form of memory which can handle 2D data and two sets of memory addresses – one for rows and one for columns? The answer is no in both cases – we still store our array in a sequence of memory locations in exactly the same way as for our list of names, as shown in table 7.19.

In order to do this we have to decide what sequence we will use to read the values from the rows and columns into memory. In this case, we have started in the bottom left hand corner, and proceeded from left to right along each row in turn until we reach the top. This will make the explanation of some of the other ways of storing raster data a little simpler, but in practice many GIS and Image Processing systems start at the top left and work their way down. There is no single agreed convention however and most GIS and Image Processing systems contain commands to ‘flip’ raster images which have been read in from systems which use a different convention.

So when a program refers to a particular pixel, such as IMAGE [2, 3] how does the computer know which memory location to go to? The size of the array will have been stated at the start of the program. In our pseudo-code notation for example, a four by four array would be declared as follows:

Array IMAGE [0 . . 3, 0. . 3]

Notice that the rows and columns are both numbered from 0 to 3. It may seem more natural to number the rows and columns from 1, but in fact starting at zero makes some operations a little easier. As before, the program knows the address of the first item in the array – the pixel in the lower left hand corner. It also knows how many pixels are in each column. So if we count along two complete rows, and then three pixels along from the start of this row, we will be able to work out the address of the array element we want.

$$\text{address} = (\text{nrow*rowsize}) + \text{ncolumn}$$

$$\text{address} = (2*4) + 3 = 11$$

Table 7.19: Storage of array in computer memory.

Address	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Value	A	A	A	B	A	A	B	B	A	B	B	B	A	A	A	A

We may like to check for ourself that IMAGE [2, 3] is' the 11th array element starting from the lower left hand corner.

Note that this calculation is not explicitly performed by the person writing a program in a language such as FORTRAN or C, who declares and uses arrays simply by putting the row and column positions in brackets after the name of the array. When the program is translated into an actual executable program by the compiler, one of the things which is done is to translate these statements into the sequence of operations which will calculate the address of the item and transfer it from memory to the CPU.

One important feature of this address calculation is that no matter how large the array held in memory, the retrieval of an item from it will take exactly the same amount of time. This is indicated by saying that the operation takes 0(1) time – the speed is the same, no matter how large the problem. Arrays can therefore be very efficient in terms of processing time. However, they are very inefficient in terms of storage, since every single pixel takes one element of storage. In order to assess the storage efficiency of various methods of handling raster arrays, it is easiest to think in terms of the number of rows or columns rather than the total number of pixels. For any given geographical area, this is determined by the resolution of the raster layer – halve the resolution and the number of rows and columns both double. However, the total number of pixels goes up by a factor of 4. This means that the array, which stores every pixel, has $O(n^2)$ storage efficiency, which is not very efficient at all.

Compression of Raster Data

If the cell values of a raster model are entered in fixed matrices with rows and columns identical to those of the registered data, only the cell values need to be stored; row and column numbers need not. Even when only the cell values are stored, the volumes of data can easily become unwieldy. Typical operations may involve 200 thematic layers, each containing 5000 cells. The total number of cell values stored is thus 200 × 5000 = 1 million. A Landsat satellite raster image contains about 7 million pixels, a Landsat TM image about 35 million pixels.

Various devices may be employed to reduce data volume and, consequently, storage memory requirements. Cells of the same value are often neighbours because they pertain to the same soil type, the same population density of an area, or other similar parameters. Thus cells of the same value in a row may be compacted by stating the value and their total. This type of compacting, called run – length encoding. Further compacting may be achieved by applying the same process recursively to subsequent lines.

Saving Space: The Run Length Encoding and Quad-Trees

The main disadvantage of the array is the size of the files when data is stored in this way. In the early days of GIS development, this was a serious problem. Even with modern computers with enormous amounts of disk space and memory, it still makes sense to reduce data sizes for a number of reasons. First, the transfer of data from disk to memory is considerably slower than the speed with which the same information can be processed once it is held in memory – therefore smaller files means quicker execution times. Second, the smaller the file size, the more images can be held in memory at one time. GIS analysis often involves viewing or using several layers – it is much slower if every time a new one is selected a file has to be moved out of memory to make way for it.

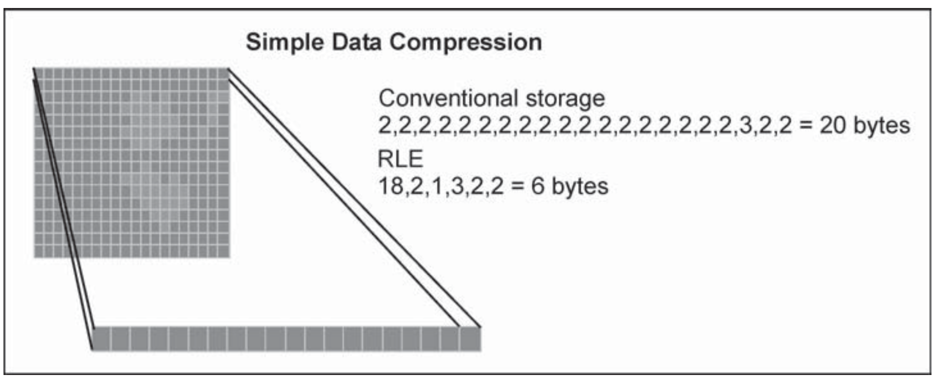


Figure 7.26: Much of the raster consists of areas which contain pixels of same value adjacent to each other.

The simplest strategy for reducing file sizes is to use the smallest possible amount of storage for each pixel. The storage of floating point numbers there is a need to store both a mantissa and an exponent, and as many digits as possible. For this reason floating point numbers are held in memory using at least 32 bits and often more. However, the same is not true of integers. An integer is held by converting the number from base 10 to base 2 and then storing the base 2 number. A single byte, with 8 bits, allows for a maximum integer of 255 as shown in table 7.20. If one of the bits is used to indicate the difference between positive and negative numbers then the range is from 128 to 127. Either of these is sufficient to hold the data in many raster layers, which often use small integer numbers – for instance Boolean layers only use the values 0 and 1 to indicate false and true respectively. Indeed these could be held using a single bit, but this is not normally an option which is available. However, the use of single byte integers is commonly available, and where appropriate will reduce the file size, and hence memory usage by a factor of 4 compared with using 32 bit words.

A second strategy for dealing with large files, is to hold only part of the layer in memory at anyone time. In order to assess the efficiency of this approach, we have to consider two issues – how much memory will be needed, and how many times will we have to transfer data between memory and disk storage. Assume we have a layer of size n (*i.e.,* with n^2 pixels in total). To process the whole array, we will have to transfer all n^2 pixels between the disk and memory, whether we copy them one at a time, or all at once. However, there is an extra overhead of time every time we ask for a transfer, because the system first has to find the location of the file on the disk, then find the particular part of the file we are requesting. Therefore, we need to try and minimize the number of times we go back and get extra data from the disk.

If we hold the whole array in memory then this uses $O(n^2)$ storage, but only requires 1 read and write operation between the disk and the memory. At the other extreme, we could read each pixel as we need it and write it back to disk afterwards. This now uses 1 unit of storage but $O(n^2)$ read/write operations. The first option is very quick, but uses a lot of memory the second uses almost no memory but would be painfully slow. A compromise is to read one row at a time into memory, process it and write it out to disk – this uses $O(n)$ storage, and also $O(n)$ read/write operations. The difference between these approaches can be quite marked. Wise (1995) describes an example where this was a real issue. The work was on the problem of capturing raster data from scanned thematic maps, such as soil or geology maps. As part of this a program was written which processed a scanned image, replacing pixels which represented things like text labels, lines etc. with the value for the likely soil or geology category at that point. The program was written for what was then the latest version of the IDRISI GIS, which worked under MS-DOS, and could therefore only access 640 Kb of memory. Even with nothing else stored in the memory, the largest image size which could be held in memory would only have been just over 800 columns by 800 rows – in contrast, by processing a row at a time images of up to 640000 columns could be handled, with no limit on the number of rows.

Table 7.20: Examples of storage of integers in bytes.

Binary	Decimal
00000000	0
00000001	1
11111111	255

These strategies may help, but there are also other things we can do in order to reduce the size of the image which needs to be stored on disk or held in memory. Each of the three main types of values stored in raster GIS layers – binary, enumerated and numerical. In the case of the first two, because the features we are representing occupy regions of the map, the raster layers contain large number of pixels with the same value next to one another. We can exploit this characteristic to save storage space and the simplest way to do this is to use what is called run length encoding. Consider the simple raster layer we used earlier, which is repeated as shown in Figure 7.27.

3	A	A	A	A
2	A	B	B	B
1	A	A	B	B
0	A	A	A	B
	0	1	2	3

Figure 7.27: An example of simple raster array.

Table 7.21: Storage of a run length encoded layer in computer memory.

Address	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Value	3	A	1	B	2	A	2	B	1	A	3	B	4	A		

When we stored this as a full array, the first 3 pixels all contained the same value – A. What we have is a sequence or run of pixels, and instead of storing each one we can store the information about the run – how long it is and what value the pixels have. Applying this to the whole layer produces the result shown in table 7.21. Even with this small example we have reduced the number of bytes of storage used for the layer. But will we always save space in this way? The answer unfortunately is no. Imagine a layer in which every pixel was different from its neighbours, such as a DEM. Every pixel would take 2 bytes of storage instead of 1 – 1 to record a run length of 1, and one for the value itself – so the file size would double.

The final raster data structure we will consider is called the quadtree, and it extends the idea of run length encoding to 2D. If we look at Figure 7.27, we can see that there is a block of 4 pixels in the lower left hand corner which all have the value A. Instead of storing four small pixels, it would be far more efficient to store 1 pixel, which was four times the size of the 'normal' pixel. This is the basis of the quadtree method in which the pixel size is allowed to vary across the image, so that uniform areas are stored using a few large pixels, but small pixels are used in areas of variation. To illustrate how this works, let us apply it to the layer shown in Figure 7.27. At the first stage, the layer is divided into four quadrants, as shown in figure 7.28. Each quadrant is numbered from 0 to 3 in the sequence shown, which is known as Morton order after its inventor. The reason for this particular sequence will become clear later. If we examine each quadrant, we call see that quadrant 0 does not need to be subdivided any further – the values in all the pixels are the same, and so our new pixel 0 can be used to store this data. The three other quadrants are not uniform and so must be subdivided again. Notice that each of the new pixels is labelled by adding a second digit, also in Morton order, to the digit from the first level of subdivision.

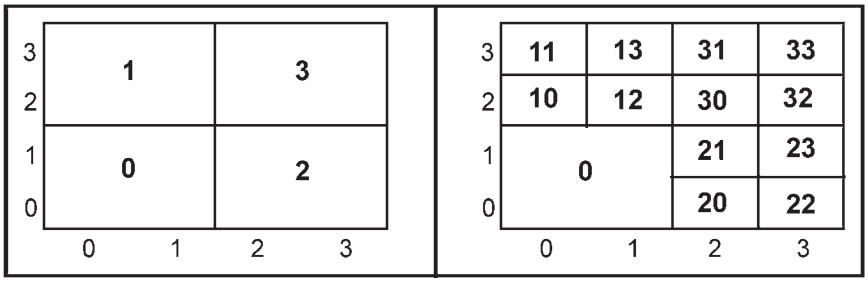


Figure 7.28: Quadtree subdivision of layer shown in figure 7.28.

Table 7.22: Storage of quadtree in memory.

Address	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Quadtree Address	0	1	2	3	10	11	12	13	20	21	22	23	30	31	32	33
Value	A	4	8	12	A	A	B	A	A	B	B	B	B	A	B	A

Because this is only a 4 by 4 image, the process stops at this point. But how do we store this information in memory, especially now that the pixels are no longer the same size? One method is shown in table 7.22.

The first four memory locations (with addresses 0 to 3) are used to store the results of the quadrants from the first subdivision of the image. The first quadrant, labelled 0, was uniform, and so we can store the pixel value – A. The second quadrant, labelled 1, was not uniform, and so we are going to need four bytes to store whatever we found when we subdivided this quadrant. The next available, location is at address 4, so we store this address in location 1. Since this address is pointing to the location of another piece of information, it is known as a pointer. We have to do the same thing for quadrants 2 and 3, storing pointers to addresses 8 and 12 respectively. The four address locations starting at 4 are used to store the results of subdividing quadrant 1 to produce 10, 11, 12 and 13 – since these were all uniform, we can simply store the pixel values, and in fact this is true for all the remaining pixels.

In this case, we have not saved any space at all compared with the original array method, because there are not enough large uniform areas on the layer. In fact, as with run length encoding, we could end up storing more information. If quadrant 0 had not been uniform, we would have needed an extra four bytes of storage to store the individual pixel values making up this quarter of the image. However, in real world examples the savings in space can be considerable.

In this example, the image was conveniently the right size to make it possible to divide it into four equal quadrants. When this is not the case (*i.e.,* most of the time) the image is expanded until its sides are both powers of 2, filling the extra pixels with a special value indicating 'no data'. This increases the amount of storage of course, but since the extra pixels are all the same, they can generally be represented using fairly large pixels, and the additional data are more than offset by the savings due to the quadtree itself.

So why is it called a quadtree? The quad part is obvious, from the subdivision into quadrants. The tree comes from a common way of representing such data structures in Computer Science as shown in figure 7.29.

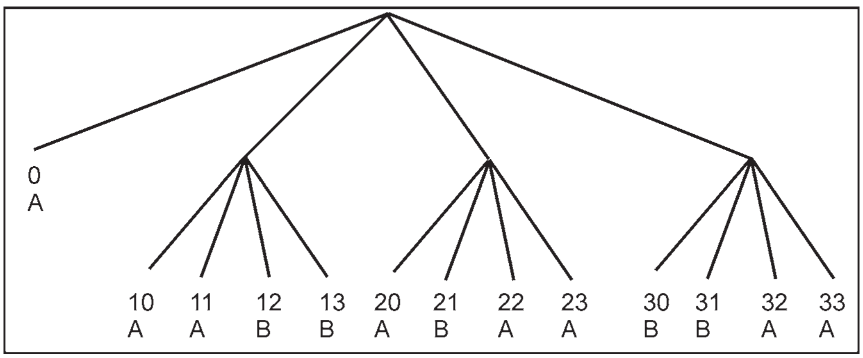


Figure 7.29: Graphical representation of quadtree.

The first level of subdivision is represented as four branches from the original image. Where quadrants are subdivided, then four further branches are drawn, giving a tree-like structure. The ends of the lines are all called nodes – those where the process ends (as in the case of 0 and 10, 11,12 and 13 etc. in the diagram) are called leaf nodes, while those which represent points of further subdivision are called branch nodes. The first node of the tree is called the root node, even though trees are usually drawn with their origin at the top, as in Figure 6.9.

In table 7.22, the Morton addresses of the pixels have been shown, for clarity. In fact, this information is not stored in reality because there is a simple method for calculating it from the row and column number of a pixel.

The advantages of the quad-tree model are

- Rapid data manipulation because homogeneous areas are not divided into the smallest cells used
- Rapid search because larger homogeneous areas are located higher up in the point structure
- Compact storage because homogeneous squares are stored as units
- Efficient storage structure for certain operations, including searching for neighboring squares or for a square containing a specific point

The disadvantages of the quad-tree model are

- Establishing the structure requires considerable processing time.
- Protracted processing may prolong alterations and updating.
- Data entered must be relatively homogeneous.
- Complex data may require more storage capacity than ordinary raster storage.

Automatic Conversion between Vector and Raster Models

GIS applications sometimes require data in a form differing from that which is available. As a result, many GIS now have facilities for automatic conversion between vector and raster models. Raster data are converted to vector data through vectorization. The reverse process, which is just as common, is rasterization. In vectorization, areas containing the same cell values are converted to polygons with attribute values equivalent to the pre – conversion cell values (Figure 7.30). In the reverse process of converting polygons to cells, each cell falling within a polygon is assigned a value equal to the polygon attribute value (Figure 7.31).

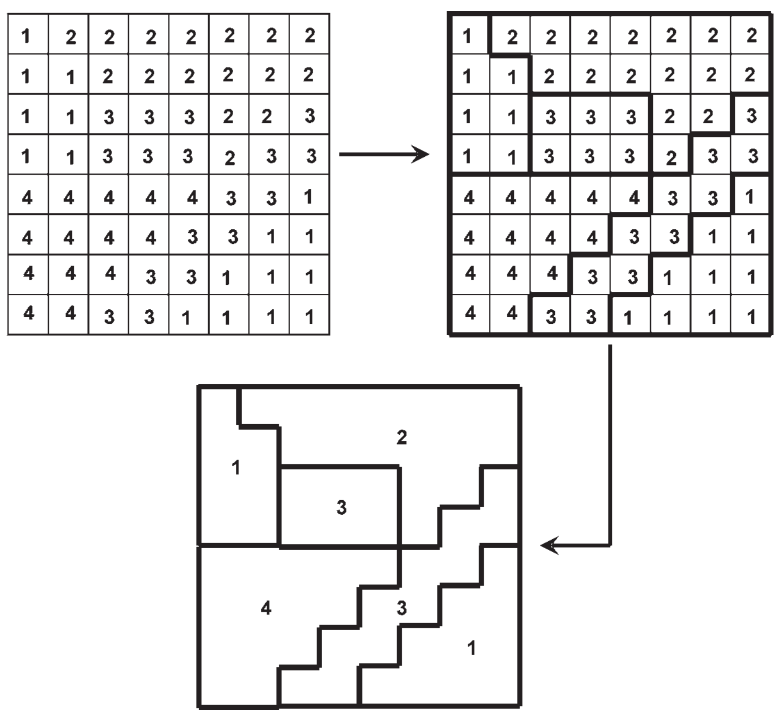


Figure 7.30: Conversion of raster data to vector data; first, each raster cell is assigned an attribute value; secondly, boundaries are set up between different attribute classes and finally, polygons are created by storing X and Y coordinates.

Various routines are available for converting raster data to vector data, and vice versa. The former is the more complex and time consuming of the two processes and different conversion programs can yield differing results from the same set of raster data. Normally, some information/data are lost in conversions. Consequently, converted data are less accurate than original data. These conversion processes apply specifically to data, not to the conversion of raster data from scanned maps into vector form.

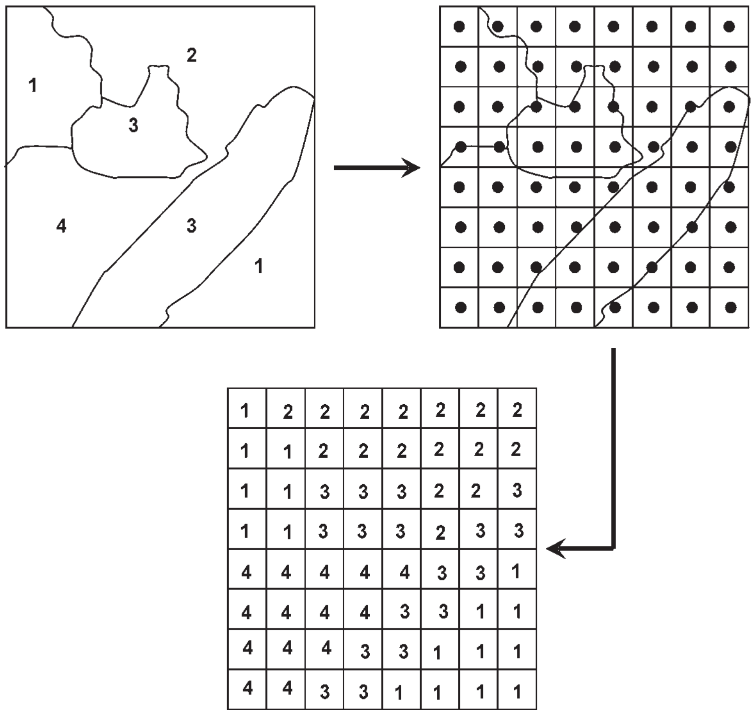


Figure 7.31: Conversion of vector data to raster data; first, polygons are coded; secondly, a grid with the right cell size overlays the polygons; here, the polygons that contain the centre of the individual cells are Identified; finally each cell is assigned the attribute code of the polygon to which it belongs.

Vector versus Raster Models

One of the basic decisions in GIS design involves the choice between vector and raster models, each of which has advantages and disadvantages. In the vector model, the observation units are end points and/or variable line or polygon magnitudes, whereas the raster model presupposes fixed observation areas in a grid. Otherwise, the models are identical.

Vector and raster data have varying ability to represent reality. It is not always easy to recognize vector data's discrete objects out in the terrain. This applies especially to phenomena with diffuse borders, such as vegetation and population density. However, many real phenomena are related to locations. Measurements are often made at points, infrastructures are often related to lines, and administrative units are frequently described in terms of defined areas of various shapes and sizes.

Raster GIS emphasizes properties: here, the basic units of observation are regular cells in a raster. Not all phenomena are related directly to such grid patterns. At the time of writing, satellite data, digital ortophoto, and digital elevation data account for the bulk of data available in raster form. In many countries, national elevation data models have been established based on a fixed grid (*e.g.,* 100 × 100 meters). Other types of data usually have to be reworked to a greater or lesser degree to suit rasters. The accessibility of raster data may thus be a major problem and perhaps the greatest drawback of a raster GIS in comparison with a vector GIS. However, as we have seen, there are methods for converting data from vector to raster.

A vector model, on the other hand, often requires the time-consuming and costly compilation of digital map data, while maps are integral parts of the data compiled for a raster model. Maps may be drawn for all cells as soon as they are assigned values.

Despite oversimplification from a functional viewpoint, vector data may be considered best suited for documentation, while raster data are more adept at showing the geographical variation of phenomena. Another simplification might be that vector data are preferable for line presentations, while raster data are superior for area presentations.

To date, the vector model has been dominant in commercial GIS implementations. The raster model, on the other hand, has been used more frequently in natural resource planning and management and also in teaching because it is more easily explained and used. Many newer GIS can manipulate both vector and raster models. With dual capability, a GIS can exploit the respective advantages of both: vector data might be converted to raster data to perform overlaying or other operations more easily performed using rasters, and then converted back to vector data.

Attribute Data and Computer Registers

With the advantages of easy updating, rapid search, and the flexible superimposition of data, the computerized filing of information has become commonplace in administrative work. Frequently, inaccessible, massive quantities of traditional records and files have been replaced by workstations from which very large amounts of information are rapidly accessible. Physical separation by rooms, buildings, national borders, or intervening distances is no longer a barrier to ready availability of information.

Table 7.23: Geometric content is often limited to identifications geometry/coordinates and topology, while attribute content often comprises location (address), various descriptions of the object and timing.

Georeferencing	Attributes
ID	ID
Geometry/	Location
Coordinates	Description
Topology	Time period

In the days when all records were kept on paper, each agency, organization, or user structured its own manual files. The result was a proliferation of parallel files, often containing nearly identical material. Computerization permits a simplification and coordination of registration efforts and can eliminate duplication and rationalize the overall filing process. In the public sector, central registers have been established as a common resource for numerous users.

Some of these registers are important in GIS applications. Others are of less interest. In many countries, though, work is under way to make public registers available to GIS users. Upon entry, register data are selected (structured) so that registers contain uniform and limited data. As for digital map data, register data are stored using formats. There is no general pattern for register content, but usually the items registered will have identities, locations, descriptive details, time and date notations, and sometimes references to other registers.

Coding and Entering Attribute Data

Attribute data may be coded for several reasons in order to:

- Establish an ID code between geometry and attributes
- Conserve computer memory
- Ease input work
- Ease verification of data entered
- Simplify subsequent searches for data in databases

The coding of geographical data is not new. Systems have been established in many fields for coding telephone lines, water pipes, manholes, streets, properties, buildings, the names of towns, and so on. Indeed, codes have been used for many reasons, not least as file access keys or to conserve the space used on file cards.

Coding of attribute data often includes data structuring. Codes are often assigned according to a hierarchical classification system devised to ease such data operations as searching and sorting. Examples include the official codes widely used for addresses, names of towns, highways, and so on. The type of data may be specified for each field, such as integer (land-use code), decimal (area), and text (name). Code tables may be compiled and used with the main table to produce more meaningful printouts from the system.

Attribute data may be entered relatively easily in most GISs, either manually via a keyboard or by importing data from an existing register. ID codes are usually entered together with the attribute data. They may also be registered or edited into compilations of attribute data which initially have no codes.

Storing Attribute Data

Attribute data are usually most easily and expediently stored in tabular form. Each line in a table represents an object, each column an attribute. Attribute data are therefore often called tabular data and are normally stored in a relational database.

Data on different types of objects are usually stored in separate tables, each dedicated to a single object type. In each table, line formats and lengths are identical throughout. The number of columns may be extended by combining several tables, either by using a common access key or by entering new attributes manually.

In principle, table design is independent of whether the geometrical data to which attributes refer are in the form of vector data or raster data. However, table content must be relevant to the objects, so each object or line must have a stable identity or access key. Data available in existing computerized registers are not always in convenient tabular form. As a result, conversions and round about methods must often be used to access such data for GIS uses.

Linking Digital Map and Register Information

Common identifiers in map data and attribute data permit moving from map data to attribute data, and vice versa. Attribute data which basically lack georeferencing may be linked to geography. As illustrated in Table 7.24, this is possible if the attribute data that lack

Table 7.24: Data elements in one data set can be used as an access key to another data set, thereby acting as a link between other data.

Digital Map Database

Point no.	Theme code	Building	X Coordinate	Y Coordinate
2971	2.1	Nishat	10211300	527330

Property Register

Building	Plot no.	Owner	Year of construction	Use
Nishat	11/1	Shaad	1998	Residential

Land Register

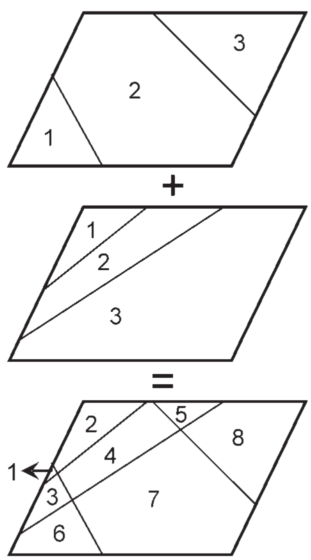
Plot no.	Land owner	Occupation	Area	Address
11/1	Fawwad	Business	10000 sq. m	Shamshad Market

Personal Register

Land owner	Age	Education	Sex	Address
Fawwad	35 Yrs.	Graduate	Male	859/6 Azad Enclave

georeferencing have access keys in common with attribute data that have other access keys in common with map data. The connection is then from attribute data to other attribute data to map data.

This illustrates one of the distinctive capabilities of GIS. Data that initially contain no geographical information or referencing may be given geographical dimensions and may therefore be used to enhance and present data in new ways, in maps or on screen. It is not always necessary to link geometry with attributes. In some instances, the geometry can be stored directly together, with the attribute data linked to each register object. This can occur in the case of, for example, a building register with coordinates representing each building or a register of measurement points for use in registering pollution leve1s. We have also mentioned that identification codes and theme codes are, in principle, attribute data though stored together with geometry.



ID	Attribute
•	Cultivated
•	Grazing
•	Forest

+

ID	Attribute
•	Development
•	Recreation
•	Preservation

=

ID		Attribute		Attribute
1	1	Cultivated	1	Development
2	2	Grazing	1	Development
3	1	Cultivated	2	Recreation
4	2	Grazing	2	Recreation
5	3	Forest	2	Recreation
6	1	Cultivated	3	Preservation
7	2	Grazing	3	Preservation
8	3	Forest	3	Preservation

Figure 7.32: Attribute data can be made comparable by superimposing geometry from dissimilar geographical units to get the integrated data.

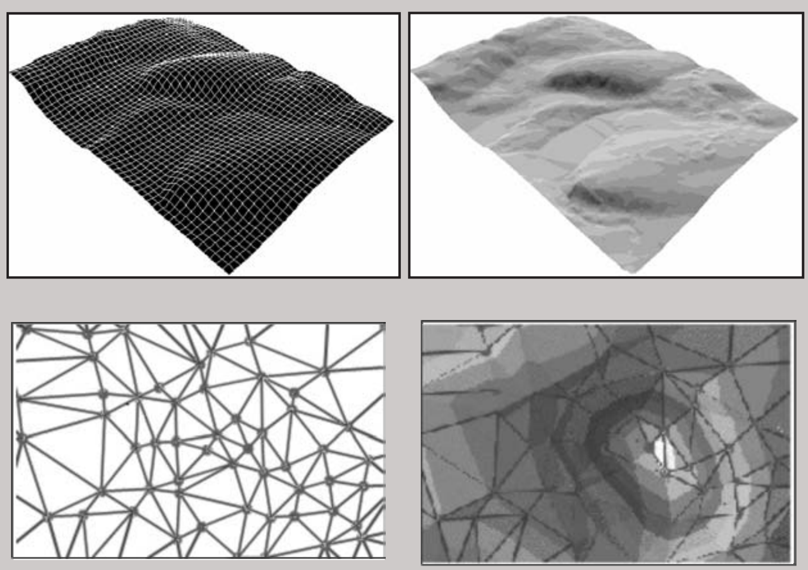
Map data may be used not just to link maps and attributes, but also geographically to link dissimilar attributes. Superimposing dissimilar data, such as geological data and vegetation data, is often hampered by a lack of commonality between the observations made in the field. That is, the observation areas listed in the respective attribute tables cannot be listed together because they refer to different sets of locations. In GIS this problem can be solved by using cartographic integration, in which overlay techniques are used to combine geometry from two dissimilar thematic maps into a single synthesized map. The synthesized map contains numerous new objects and areas, all of which are related to the two original thematic maps. Hence, the objects in the synthesized map comprise the least common units between the original maps and are therefore called ***integrated terrain units*** (ITUs; Figure 7.32). An attribute table is associated with the ITUs. In it, the ITUs are listed on the lines and the elements of the original thematic maps are in the columns. This table contains all the relevant attributes and therefore may be used in further analyses of the data.

Cartographic integration is straightforward when the areas of the original map data contain more or less homogeneous data, such as property data, land use, vegetation, and geology. Complexities arise when properties are not evenly distributed over an area. Consider, for example, a typical city with an unevenly distributed population that averages 500 persons/km^2. An ITU might locate in an uninhabited area of the township and hence misrepresent the facts of its population. In all such cases, rules must be contrived to designate how attributes shall be divided among ITUs. The ITU is also called a ***basic spatial unit*** (BSU) and defined as a fundamental area unit which has homogeneous properties in the context of a particular subject.

This page intentionally left blank

CHAPTER 8

ADVANCED DATA MODELS



Surface Representation

The models discussed in the previous Chapter describe limited parts of two-dimensional real world. Several other data models used in GIS can extend the real world to include the terrain surface, the time factor, and movable objects. The digital representation of a terrain surface is called either a digital terrain model (DTM) or a digital elevation model (DEM). In GIS disciplines, the term DTM is often used not just for the model itself, but also for the software used to manipulate the relevant data.

The terrain surface can be described as comprising two basically different elements. The random (stochastic) elements are the continuous surfaces with continuously varying relief. It would take an endless number of points to describe exactly the random terrain shapes, but these can be described in practice with a network of points. It is usual to use a network that creates sloping triangles or regular quadrants.

DIGITAL ELEVATION MODELS (DEMs)

A **B**

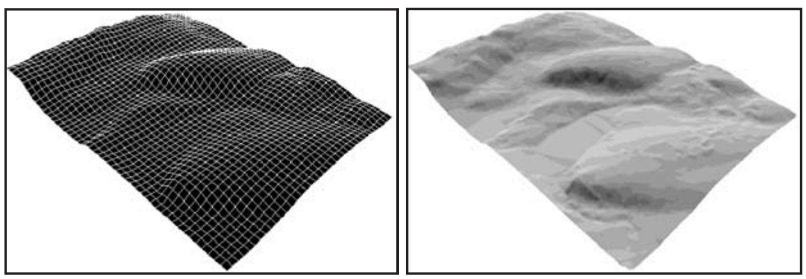


Figure 8.1: A DEM is an essential layer in the representation or analysis of any area with variable terrain.

The systematic part of the terrain surface is characterized either by sharp cracks in the terrain, such as the top or bottom of a road cut, or by characteristic points such as spot depression and spot height. The systematic part is thus best represented by lines and typical single points. Prominent terrain features can be verbally described using many terms, such as smooth slope, cliff, saddle, and so on. Geometry, however, has only three terms: point, line, and area. One cannot describe continuously varying terrain using only three discrete variables, so all descriptions are necessarily approximations of reality.

Essentially, DTMs comprise various arrangements of individual points in $x - y - z$ coordinates. Often, their purpose is to compute new spot heights from the originals. A terrain model can be realized by linking height as an attribute to each point (x, y). This type of elevation model can only describe a surface and cannot handle more z values to the same point. Therefore, the term 2.5 – dimensional is often used to describe the DTM dimension. This model is most suited to visualization. In a three-dimensional elevation model, elevation is an integral part of position (x, y, z) and the model can handle several z values for the same x, y pair. That is, it can handle different geological layers, roof heights on buildings, roads that cross each other, together with the terrain surface. A three-dimensional model is also suited to volume calculations.

The *z* value of a new point is calculated by interpolation from the *z* value to the closest existing points. If the points are stored in an unstructured way, all registered points will have to be searched to be able to calculate the *z* values to a new point. This can be very time consuming even for a powerful computer. It is therefore usual to use data structures which also describe the contiguity between the points. This is achieved by using data structures based on single points in a raster (grid) or triangles covering a surface.

Grid Model

A systematic grid, or raster, of spot heights at fixed mutual spaces is often used to describe terrain (Figure 8.1 A). Elevation is assumed constant within each cell of the grid, so small cells detail terrain more accurately than large cells. The size of cells is constant in a model, so areas with a greater variation of terrain may be described less accurately than those with less variation. The grid model is most suitable for describing random variations in the terrain, while the systematic linear structures can easily disappear or be deformed. One possible solution can be to store the data as individual points and generate grids of varying density as required. It is debatable whether the grid model represents samples on a grid and can therefore be called a point model, or represents an average over raster cells.

Elevation values are stored in a matrix and the contiguity between points is thus expressed through the column and line number. When the data points are dispersed, the averages of the elevations of those closest to grid points, within a given circle or square, are assigned to the grid points with inverse weighting in proportion to the intervening distances involved. When the data relate to profiles or contours, grid point elevations are interpolated from the elevations at the intersections of the original data lines and the lines of the grid.

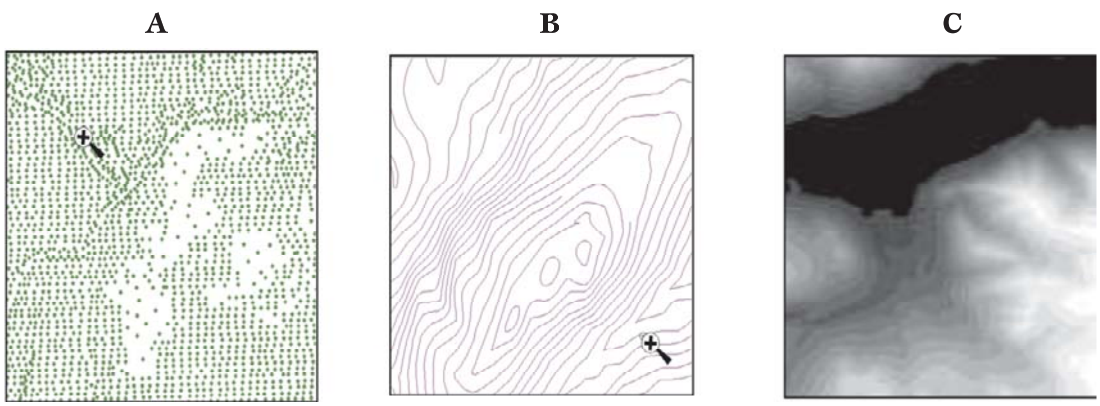


Figure 8.2: Elevation data are acquired through:

a. Mass points (lattices)
b. Contour lines
c. GRIDs (interpolated from points or lines; or created currently from digital imagery)

Terrain may also be described in terms of chosen or arbitrarily selected individual points (*i.e.,* a point cloud). In principle, the characteristics of the terrain between points are unknown, so it follows that point densities should be greatest in areas where terrain features vary the most. Only elevations are stored for the points of a regular grid, but both point coordinates and elevations must be stored for point clouds. So for given terrain coverage, the amount of memory storage required for the two point arrangements differs. For describing abrupt terrain variations, such as the top and bottom of a road cut, point models are inferior.

TIN Model

An area model is an array of triangular areas with their comers stationed at selected points of most importance, for which the elevations are known. The inclination of the terrain is assumed to be constant within each triangle. The area of the triangles may vary, with the smallest representing those areas in which the terrain varies the most. The resulting model is called the triangulated irregular network.

Insofar as possible, small equilateral triangles are preferable. To construct a TIN, all measured points are built and the model thus represents lines of fracture, single points, and random variations in the terrain. The points are established by triangulation and in such away that no other points are located within each triangle's converted circle. In the TIN model, the $x - y - z$ coordinates of all points, as well as the triangle attributes of inclination and direction, are stored. The triangles are stored in a topological data storage structure comprising polygons and nodes, thereby preserving the triangle's contiguity.

Box 16: *Make your own* TIN *model*

A simple experiment using a piece of paper can give you an idea of TIN model. If you take a sheet of paper squeeze it in your palm. Now if open it again you would find many irregular sized creases. *This is a good and simple example of* TIN MODEL. The creases are the ridges and valleys and the intersections of the creases are the peaks, depressions and passes. The areas of flat paper between the creases are the irregular triangles of the TIN model, which may be assigned area slope and aspect values. Certainly, this is not a perfect model as not all of the facets on your piece of paper will be triangles, but it gives a rough idea to illustrate TIN principle. Further if you hold the paper level with a light source you could get a fair idea of a miniature terrain of peaks and valleys casting shadows in front of you. The tighter you squeeze the paper more complex terrain you produce. This experiment also demonstrates the *'two and a half'* dimensional nature of terrain models in GIS. This model is the surface with no depth.

Various algorithms are available for selecting representative points from the basis data (grid, contours, point clouds) and for creating appropriate triangles. Should the basis data be available in grid form, it is possible to move a window (one point and its eight contiguous points) step by step over the data and remove the points that are least characteristic in relation to their contiguous points. Triangles can be formed by laying circles through three points and testing whether there are other points within the circle. If other points are not available, a new triangle will be formed. This method produces triangles with a low variance in length; it is known as the Delaunay triangle (Figure 8.3).

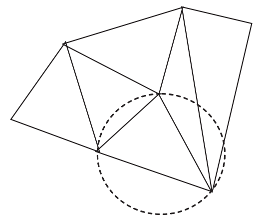
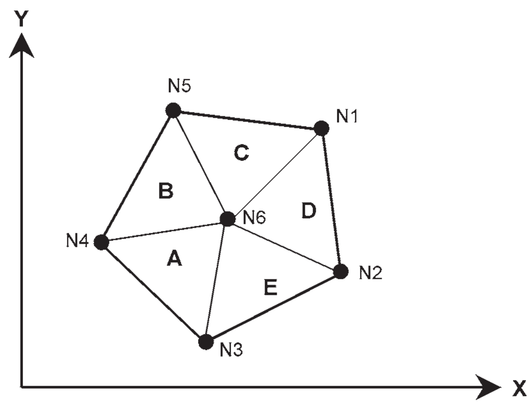


Figure 8.3: Delaunay triangulation is a method used to fit triangles in a point cloud. The circle described ensures that the triangles have good geometry with least possible variation in page lengths.



Triangle Table and Node

Triangle	Adjacent	Nodes
A	B, E	N3, N4, N6
B	A, C	N4, N5, N6
C	B, D	N1, N5, N6
D	C, E	N1, N2, N6
E	D, A	N2, N3, N6

Coordinate Table

Node	Coordinates
N1	X1, Y1, Z1
N2	X2, Y2, Z2
N3	X3, Y3, Z3
N4	X4, Y4, Z4
N5	X5, Y5, Z5
N6	X6, Y6, Z6

Figure 8.4: TIN model: the triangles are stored in a topological structure.

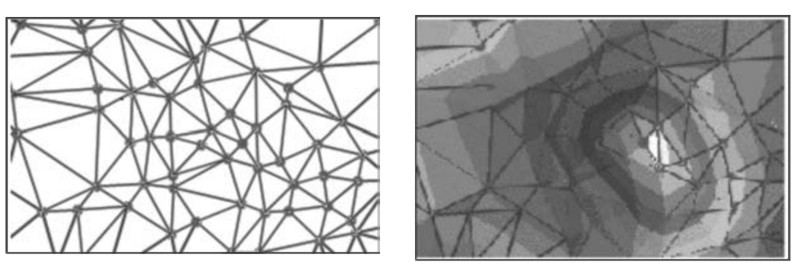


Figure 8.5: An example of topological structure in TIN model.

Compared to the grid model, the TIN model (Figure 8.4) is cumbersome to establish but more efficient to store because areas of terrain with little detail are described with fewer data than similar areas with greater variation. However, the TIN model normally requires considerably larger storage capacity than the grid model. TIN models are good for describing terrain because the sharp breaks of slope between uniform-slope facets fit certain types of terrain well.

Other Models

Isolines – continuous lines connecting points of the same elevation may represent terrain in much the same way as contour lines depict terrain on conventional maps (Figure 8.6). The point densities should be greatest in those areas in which the terrain varies the most. As the intervening terrain between successive isolines are unknown, smaller elevation increments between isolines result in greater accuracy of description.

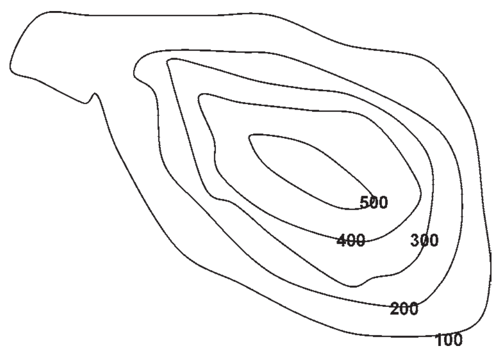


Figure 8.6: Lines that link points with the same terrain height are used to represent terrain surface, which corresponds to traditional elevation contours. However, this structure is poorly suited to the calculation of elevation values for new points.

Although an isoline model may be compiled readily, amending its data is involved. In practice, the methods used are determined by the data compilations. Parallel profile lines connecting points of varying elevation may be used to describe terrain. The density of points along profile lines should be increased in areas where there are major variations in the terrain. In principle, the terrain between successive profile lines is unknown, so the closer the lines, the greater the accuracy of description.

A combination of isolines and individual points may also be used to describe terrain, especially when specifying such point features as peaks and valley floors, or vital terrain lines, such as the top and bottom of a fill. As mentioned previously, the grid and TIN models are best suited for calculation of the *z* value of new points.

Practical Observations

Grid models and TIN models are always less accurate than the original data from which they are derived. In some GIS/DIM, therefore, original data are stored in point clouds. Models with accuracy suiting specific tasks are compiled from these as required. For example, a grid model for estimating road construction excavations might be more accurate than one intended to detail the general vegetation of the region.

GIS based on vector models can easily manipulate elevations stored as spaghetti data, but can handle elevation grid data less easily. Only a GIS based on a topological model can manipulate TIN data. Terrain data are usually compiled from survey point elevations, from isolines digitized from existing maps, or from photogrammetric point and/or line (contour or profile) registration.

Various interpolation programs compute new z coordinates for new $x - y$ coordinates, thus facilitating specific computations such as estimating cut – and – fill volumes in road planning, or assessing reservoir volumes for hydroelectric plants. Various GIS may also be implemented with functions for calculating slopes, drawing in perspective in order to visualize the impact of works, computing the runoff, or perhaps, draping in colours to enhance visualization.

The ways in which data are represented and stored are decisive in determining the type and efficiency of the computations. For example, digital isolines are ill suited to calculating slopes and relief shadowing; draping and runoff calculations are most expediently performed using a TIN model; the TIN model is ill suited to visualization without draping; and so on.

The methods used to describe terrain surfaces may also be used to describe other continuously varying phenomena. Thus population density, prevailing temperatures, or biomass production can be described quite simply by assigning the parameter involved to the z axis of all the observation points located in $x - y$ coordinates.

Accuracy

High accuracy is required in all terrain models to be used for engineering purposes. The accuracy of terrain descriptions is determined primarily by random variations in the terrain, spreading of measured points, distance between measured points, and accuracy of points by the method of generating the model grid and triangular surfaces; and by the method used to interpolate between points in the model. In DIM, an error in the x and y coordinates result in errors in the elevation.

For a grid model, the following degrees of accuracy are typical:

Source	Accuracy in elevation and ground plan
Surveying	±5cm
Photogrammetric data from	
1: 6000 images	±20 to 30 cm
Digitized 1: 1000 maps	±50 cm

In those models in which cells and profiles are recreated from a point cloud with each computation, accuracy depends on the cell or profile density. Profiles must be closer to each other to represent more rapid terrain variations, but greater profile densities naturally call for the processing of greater amounts of data.

Three-Dimensional Objects

All physical phenomena are located in space; thus the world can be described as three-dimensional. A complete data model should be based on these three dimensions: ground, position, and elevation. This applies not only to terrain surfaces but also to buildings, borders, addresses, accidents, and all manner of data; a complete data model should manipulate georeferenced data in three dimensions.

The realization of three-dimensional objects in GIS still has theoretical and practical limitations. Topological data are needed for such procedures as colour filling (and photo texture) of vertical areas and for data search. It is a theoretical and mathematical problem to establish topology for three-dimensional objects. The topology will be very complex and present opportunities to establish objects which are illegal (objects that cannot be oriented). It is also difficult to establish satisfactory routines for checking whether declared data exist in three-dimensional topology.

Specification of all objects in three dimensions can easily increase the amount of data collected beyond that which is needed; the amount of data will in any case be considerably larger than with the use of two and 2.5 dimensions. It may also influence the techniques used to collect data. The collection of photogrammetric data provides free elevation data in addition to the northing and easting data of the ground plane. When existing maps are digitized, however, elevation data must be entered manually (and sometimes inaccurately, as exact elevations may not be available for all objects). Today, users have an increasing need for digital three-dimensional map data. This applies in particular for applications connected to urban areas. At present, commercial GIS is still only capable of handling two-dimensional topology. Even though relational databases support binary large objects (BLOBs) for storage of texture (building facades or similar), this type of data cannot be searched for as with other data. Models can be constructed, but should in this case be carried out in systems for computer-aided design (CAD).

Box 17: Dimension of time in GIS

Integrating the dimension of time into GIS presents challenges. The main reasons for it is that, data about spatial object are not easily available for a continuous period, or data models and structures that allow us to record store and visualize in different temporal states are in infancy. This problem is bad enough when the geographic entity under investigation is fixed with respect to location, but it is more complex when the object is either mobile or changes its entity type through time. There are four type of temporal event. This provides an indication of the types of changes that may affect an entity:

- Continuous – these events go on throughout some interval of time.
- Majorative – these events go on most of the time.
- Sporadic – these events occur some of the time.
- Unique – events that occur only once.

Handling time in GIS: In a raster or vector layer based GIS, one option for handling time is to store multiple layers for the theme under investigation. The problem with this approach is that it generates lots of duplicate data. One solution is to store only information that changes to reduce the data storage requirements.

In a GIS using an object-oriented data model, a different approach is used. The various elements and attributes which make up an object can each be assigned a time tag.

Representation of Time

In the real world, time is a factor that concerns us deeply. It was studied by ancient Greek philosophers such as Plato and Aristotle, and also plays an important role in Einstein's theory of relativity. Most things change with time. The same applies to geographical data. For example, land use data. New land parcels are continuously under development because of the division of existing land parcels; therefore, new geometry occurs. However, most changes are related to title ownership, with the resulting changes in attribute values. One example of an area in which extreme changes occur is that of the transport sector with vehicles continuously changing position and where frequent on-and off-loading of goods change the attribute values that are linked to the vehicle. In addition to the fact that both the geometry and attributes of objects change over time, the reference system can also be changed. Earlier, it was discussed how this occurs when roads are changed in a distance-based reference system, but it also happens when the geometry of administrative reference units is changed (*e.g.,* with changes in ward borders and postal zones). In addition, topological changes often occur as a consequence of geometrical changes.

In practice, it is difficult to create a data model that is capable of incorporating all imaginable changes. The time factor is relatively often neglected in GIS, probably because we are more concerned with documenting our current situation than we are with historical changes. Databases are updated continuously, so unless special measures are taken, the time picture will be fleeting. The use of analog technology can document time changes via different printed versions of maps. To the extent that maps have been archived, it is also possible to preserve special versions of history, although they might be unsystematic. If the time factor is not incorporated in the data model for GIS, we run the risk of losing important historical data.

The most usual way of handling the time factor in GIS is to look on time as an attribute to the objects in the same way as for other attributes. This view corresponds with the usual way of presenting geographical data (*i.e.,* as two-dimensional time overlays) and can thus be realized for both vector and raster data. However, this simple approach will not necessarily create a logical connection between the various time layers. It can therefore be extremely difficult to assess what the situation might have been between two time layers.

Possible practical solutions will therefore be:

1. The attributes of the objects will be changed.
   a. Historical data are stored only in fixed or variable time intervals (*e.g.,* every second year, every fifth year, etc.). The attribute values between these intervals may have to be interpolated.
   b. All changes are registered and stored for selected types of objects (*i.e.,* historical data have to be preserved by date stamping).

The type of object will decide whether all changes should be registered or whether time intervals are sufficient. For example, time intervals would be most suited for registration of changes in population density, while changes in the form of new property title holders should have complete registration.

2. The geometry of objects is changed.
   a. Historical data are stored only in certain time intervals (*e.g.,* every second or tenth year). The geometry between these intervals may have to be calculated. As in the case of attribute values, the object type will decide where it is possible to interpolate new geometry.
   b. All changes are registered and stored for selected object types (*i.e.,* all historical data have to be preserved by date stamping).

Changes in geometry can lead to changes in the relationships between objects and the resulting changes in topology, which also have to be preserved.

Registration in time intervals is more of a practical solution, where the main aim is to maintain rapid access to data and limited data volumes rather than realization of a basic data model, of which time is an integral part (Figure 8.7). Time models specify how changes in terrain over time can be preserved. Updating comprises the routines to be followed for registration of changes and the speed at which changes can be loaded into the database(s) and, to a certain extent, can be viewed independent of the models.

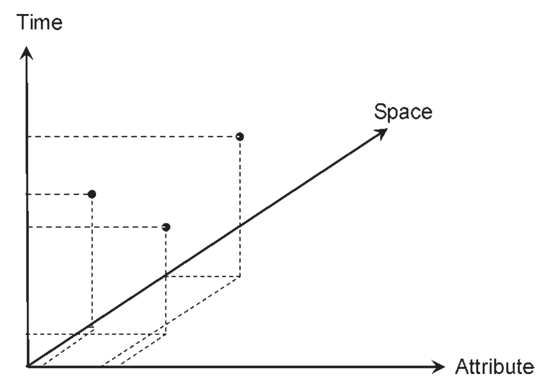


Figure 8.7: The real world changes occur in time, attribute and space, of which time is an integral part for realization of a basic data model.

Several prototype systems and even a few commercial systems are available which provide some temporal support, but it would seem that there is a fair amount of research and development work remaining before a complete data model for the time factor can be realized in GIS. Even though a data model can be created that can handle the time factor satisfactorily, we are, in practice, reliant on changes being registered and stored in the database within a reasonable time.

Models for Moveable Objects

A considerable part of the real world consists of moveable objects: vehicles on a road network that carry passengers or goods or water running over the terrain surface. One should also be able to realize this aspect of the real world in GIS. Special models have therefore been developed to handle these conditions.

Network Model

The network model comprises road systems, power grids, water supply, sewerage systems, and the like, all of which transport movable resources. The most usual type of network for which GIS models are developed is road systems; the following description therefore refers to road systems. For most purposes, reality can be simplified to a model that can handle two different situations:

I. Displacement of resources or objects from one place to another
II. Allocation of resources or objects from or to a center.

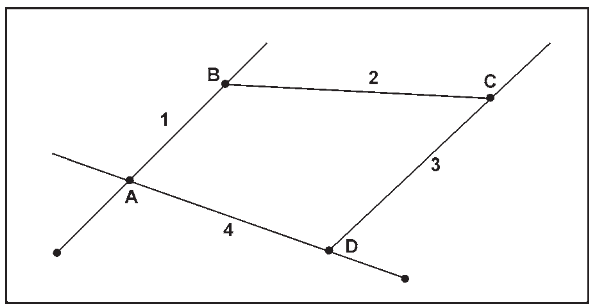


Figure 8.8: A road network splits into links and nodes, here link is a line without logical intermediate intersection and a node is an intersection point where two or more links meet or a start/end point.

As with other GIS data models, this model is based on geometry and attributes (Figure 8.8). The geometry in the network is represented by lines consisting of connected lines of vector data. The geometry of a road system will be represented by the center line of the road. This model assumes topological data built up of links and nodes. Every link and every node in a network must have a unique identity. The attributes are tied to links and nodes in a linear system through these IDs and are intended to describe the total accessibility of the system. We can thus state that the model is based on three basic relationships:

I. Continuous, connected networks
II. Rules for displacements in a network
III. The possibility of attribute value accumulations due to displacements

Attributes are connected to links and nodes and consist of two main categories. One sets conditions for transfer in the network, while the other specifies which resistance occurs at different locations in the network (Table 8.1). Attributes that determine how objects can be moved in the network can have direction predetermined (one-way streets, closed roads, weight limits etc.). Attributes that specify resistance in the network can be speed limits, road works, peak traffic, traffic lights, bus stops, sharp curves, and so on. The accumulation of all resistance occurs along a route in the system and indicates the transfer speed from the start to the finishing point.

Table 8.1: Attributes are attached to links and nodes and the resulting data are displayed in tables.

Link	Distance	Restriction	Node	Resistance
1	5 kilometers	25 (km/h)	1	2 minutes
2	4 kilometers	60 (km/h)	2	1 minute
3	5 kilometers	40 (km/h)	3	1.5 minutes
4	3 kilometers	30 (km/h)	4	2 minutes

Once the model has been constructed, it is possible to simulate the quickest and/or the shortest route between points A and B based on the route with lowest accumulated resistance. For example, it is possible to define areas that are covered within a driving distance of 5 km and 15 km, respectively, from, say, a school and to simulate personnel transport by different means. Certain relationships need to be taken into consideration when establishing network models. Links are customarily selected to carry information, which may complicate the task because data volume and complexity increase in proportion to the number of links. The initial step, how a road is divided into links, determines the nature of the nodes. If all intersections, events, and features along the road result in nodes, the number of links may be enormous, resulting in the need for large storage capacity and slow data retrieval.

The network model represents a real-world model, since it is based neither on entities/layers nor is it fields/object oriented. Information in the network model is based on links and nodes, which are not found in the entity model. Nor is it object oriented, since new nodes and links will not be established wherever attributes change; they are established only where it is practical to measure the flow of resources through the network.

Model for Movement Over Surfaces

In the network model, movements are limited to the network. There are, however, some situations where access is otherwise in the terrain, such as water that flows on the surface (drainage). The free flow of resources in the terrain can also be modeled by using geometry and attributes, but in this case, it is practical to use the full-coverage raster model instead of vector data. The geometry is thus represented with regular cells, and attributes are represented with coded values for each cell (Figure 8.9).

The cells are coded with attribute values which characterize the terrain in relation to the phenomenon to be studied. In the case of drainage, direction of slope can be one theme and angle of slope another. The accumulation of cell values based on certain rules will thus give the total drainage values for different parts of the terrain.

Connective models of raster data may also be used to determine travel distances, to identify areas of given shapes and sizes, and so on. The raster model for movable objects is in many ways not unlike standard raster models. It is a question of the suitable coding of cells relating to that part of the real world which the model is intended to reflect.

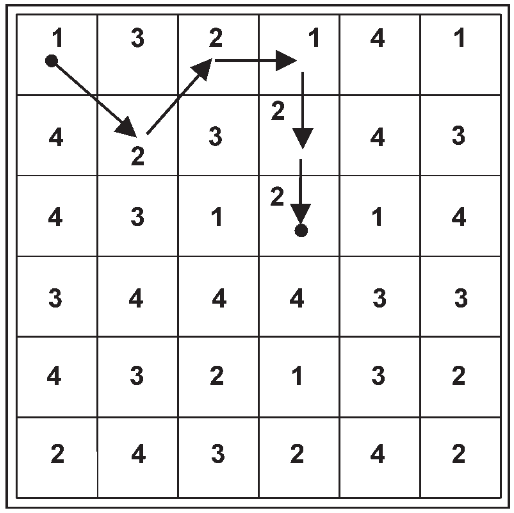


Figure 8.9: An example of optimizing route location on raster data.

Combination of Models

No models, of any degree of complexity, are perfect in relation to the real world; they are only more or less successful approximations. However, they can be better in combination than singly. This can be utilized in the creation of hydrological data models by combining three dimensions, the network model, and the raster model.

The technique of multimedia integrates several types of models: vector, raster, 3D, time, and so on. The multimedia technique helps the user to develop complete mental models of spatial problems and gives the user the ability to navigate in a GIS-derived information space. However, there is still a good deal of research and development work to be carried out before the time factor and three dimensions are fully integrated in commercial GIS.

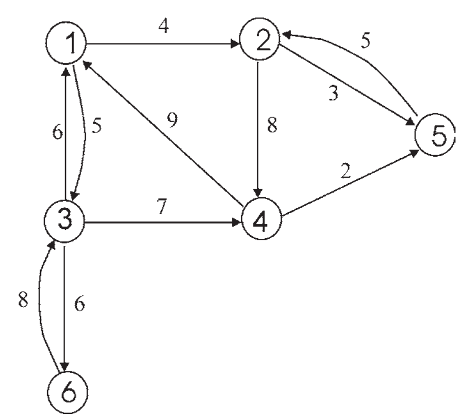


Figure 8.10: An example network.

An example network is given in figure 8.10. In this example, nodes are represented by the small circles and the arcs are represented by the lines connecting the nodes. The number in each circle is the node identifier (ID). The numerical value next to each arc can be considered the distance (time) of traversing that arc. A direction is also given for each arc.

The Representation of Networks

The tradeoff of choosing a particular data structure is often between speed and storage space. In network analysis, commonly used representations of a network include:

⇒ Node-Arc Incidence Matrix
⇒ Node-Node Adjacency Matrix
⇒ Adjacency Lists
⇒ Forward and Reverse Star Representation

The Forward and Reverse Star representation is the most efficient among all existing network data structures for representing a network. Therefore, here we describe Node-Node Adjacency Matrix and Forward and Reverse Star representation. The reason for this choice is that the Node-Node Adjacency Matrix is the most basic form of representing network topology and the Forward and Reverse Star representation is the most efficient.

Node-Node Adjacency Matrix

Following is the Node-Node Adjacency Matrix representation of the network given in table 8.2:

- The rows and columns in the matrix correspond to the nodes on the network.
- A non-zero element in the ith row and jth column in the matrix represents the numerical value associated with arc (i, j).
- A zero element in the matrix in the ith row and jth column in the matrix indicates that there exists no arc going from node i to node j.

Table 8.2: Node-Node Adjacency Matrix of the example network

0	4	5	0	0	0
0	0	0	8	3	0
6	0	0	7	0	6
9	0	0	0	2	0
0	5	0	0	0	0
0	0	8	0	0	0

The storage space required for the Node-Node Adjacency Matrix representation is *an^2* for a network with n nodes, where a is a constant. The advantages of the Node-Node Adjacency Matrix representation are that it is very easy to implement and is suitable for dense networks.

FORWARD AND REVERSE STAR REPRESENTATION

The Forward and Reverse Star representation stores the arcs eminating from the nodes in a single array. In constructing the Forward Star representation of a network, a unique sequence number is assigned to each arc to obtain the ordering of the arc list. Arcs are numbered in the following order:

⇒ first arcs eminating from node 1 are numbered, then those from node 2, and so forth.

⇒ Arcs eminating from the same node are numbered in an arbitrary fashion.

Once this list of ordered arcs is obtained, data associated with the arcs are stored in single arrays sequentially. For example, for any arc (i, j), if it is numbered arc 10, then the starting-node, ending-node, and length of this arc are stored in the array positions starting-node (10), ending-node (10), and length (10). In addition to the list of ordered arcs, a pointer is also maintained for each node i, denoted by $pointer(i)$. The numerical value associated with $pointer(i)$ is the smallest-numbered arc eminating from node i. If there exists no arc going out from node i, then $pointer(i)$ is set to be equal to $pointer(i+1)$.

For consistency, we set $pointer(1) = 1$ and $pointer(n+1) = m+1$.

Tables 8.3 and 8.4 shows the Forward Star Representation of the network example given in Figure 8.10.

Table 8.3: A list of order arcs in the Forward Star Representation.

Arc No.	Starting-node	Ending-node	Arc-length
1	1	2	4
2	1	3	5
3	2	4	8
4	2	5	3
5	3	1	6
6	3	4	7
7	3	6	6
8	4	5	2
9	4	1	9
10	5	2	5
11	6	3	8

Table 8.4: Pointer to each node in the Forward Star Representation.

Corresponding node	Element value
1	1
2	3
3	5
4	8
5	10
6	11
(7)	12

Table 8.5: A list of order arcs in the Reverse Star Representation.

Arc No.	Starting-node	Ending-node	Arc-length
1	3	1	6
2	4	1	9
3	1	2	4
4	5	2	5
5	1	3	5
6	6	3	8
7	2	4	8
8	3	4	7
9	4	5	2
10	2	5	3

Table 8.6: Pointer to each node in the Reverse Star Representation.

Corresponding node	Element value
1	1
2	3
3	5
4	7
5	9
6	11
(7)	12

The Forward Star representation is a data structure that can be used to efficiently determine the set of arcs outgoing from any node. On the flip side, the Reverse Star representation is a data structure that provides an efficient means to determine the set of incoming arcs for any node. The Reverse Star representation of a network can be constructed in a manner similar to the Forward Star representation. The only difference is that *incoming* arcs at each node are numbered sequentially. Tables 8.5 and 8.6 are the Reverse Star Representation of the example network shown in Figure 8.10.

There is a significant amount of duplicate information when both the forward star and reverse star representations are stored in a computer. To avoid the duplication, we only maintain a single array called *trace* which stores the arc numbers in the forward star representation. The sequence in which the arc numbers are stored corresponds to the sequence in the reverse star representation. For example, the first arc in the reverse star representation is arc (3, 1) whose arc number in the forward star representation is arc number 5. Therefore, we have *trace (1) = 5.*

Similarly, we have *trace* (*8*) = *6*, and so on. The size of the single array *trace* is *m*.

A compact forward and reverse star representation of the example network is given in Figure 8.11.

	Pointer
1	1
2	3
3	5
4	8
5	10
6	11
7	12

Arc No.	Starting Node	Ending Node	Arc Length
1	1	2	4
2	1	3	5
3	2	4	8
4	2	5	3
5	3	1	6
6	3	4	7
7	3	6	6
8	4	5	2
9	4	1	9
10	5	2	5
11	6	3	8

	Trace
1	5
2	9
3	1
4	10
5	2
6	11
7	3
8	6
9	8
10	4
11	7

	R-Pointer
1	1
2	3
3	5
4	7
5	9
6	11
7	12

Figure 8.11: Compact forward and reverse star representation of the example network.

The storage space required for the Forward and Reverse Star Representation is $an + bm$ for a network with n nodes and m arcs, where a and b are constants.

The advantages of the Forward and Reverse Star Representation are that it saves space, it is efficient to manipulate, and it is suited for dense as well as sparse networks.

Representation of Network Attributes

The key to network representation is to represent nodes, arcs and network topology efficiently. Once the nodes, arcs, and network topology are efficiently represented, other data and information associated with nodes, arcs, stops, centers, and turns can be represented as attributes either associated with nodes or arcs.

Computation of Shortest Paths on a Network

The computation of shortest path algorithms is a vital component of any network analysis task. For many network analysis tasks, the computation of shortest (fastest, least cost) paths is almost always the first step because shortest paths are often needed as input to 'higher level' models. When the network involved is large, the computation of shortest paths is a computationally intensive process. Therefore, the choice of the fastest and most efficient shortest path algorithm is a very important task in network analysis.

Existing shortest path algorithms can be categorized into two groups: *label-setting* and *label-correcting*. Both groups of algorithms are iterative and both employ the *labelling method* in computing one-to-all (one node to all other nodes) shortest paths. These two groups of algorithms differ in the ways in which they update the estimate (*i.e.,* upper bound) of the shortest path distance associated with each node at each iteration and in the ways in which they converge to the final optimal one-to-all shortest paths.

In *label-setting* algorithms, the final optimal shortest path distance from the source node to the destination node is determined once the destination node is scanned. Hence, if it is only necessary to compute a one-to-one shortest path, then a label-setting algorithm can be terminated as soon as the destination node is scanned, and there is no need to exhaust all nodes on the entire network. In contrast, a *label-correcting* algorithm treats the shortest path distance estimates of all nodes as temporary and converges to the final one-to-all optimal shortest path distances until its final step when the shortest paths from the source node to all other nodes are determined. A key operation in many shortest path algorithms is the *labelling method.*

Common network operations

- ***PATHFINDING*** is the process to find the shortest, least cost, or most efficient path or tour on a network.
- ***TRACING*** is the process to determine a connected portion of a network that are either flow from this connected portion of the network to a given node or flow from a given node to this connected portion of the network.
- ***ALLOCATION*** is the process to assign portions of a network to a location (*e.g.,* a center) based on some given criteria.

Common network applications

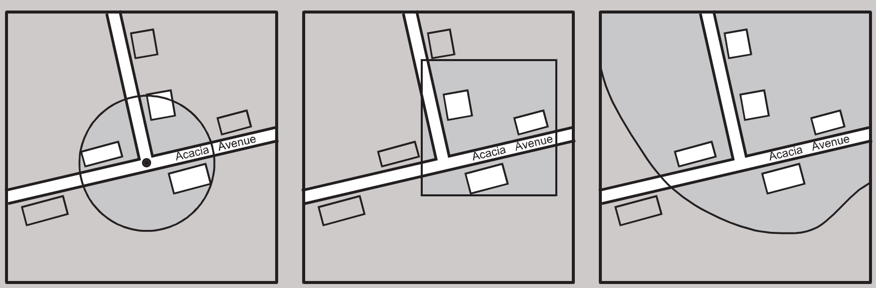
- ***GEOCODING*** is the process for building a relationship between locational data in a database and street address data that are normally in a tabular format. In many applications, there are only tabular address data available. Thus, geocoding provides a very convenient mechanism to establish a database relationship between geographic locations and addresses. There are many examples of geocoding. For example, in retail analysis, customers' addresses can be used to create maps showing locations of different customers with different shopping behaviours.
- ***LOCATION-ALLOCATION*** is the process of determining the optimal locations for a given number of facilities based on some criteria and simultaneously assigning the population to the facilities. The determination of locations for retail stores, restaurants, banks, factories, and warehouses or the choice of locations for libraries, hospitals, post offices, and schools can be supported by analysis results from location-allocation models.
- ***BUSINESS LOGISTICS:*** The optimization of vehicle routing and delivery scheduling is vital for many business operations. Business logistics is concerned with such an optimization. The combined power of GIS and network analysis makes GIS an ideal environment for analyses related to business logistics.

- **SPATIAL INTERACTION AND GRAVITY MODELLING:** The interaction between different locations in geographic space and the mathematical modelling of the interaction are important in application areas such as transportation and retail analyses. Gravity models are commonly used to support these analyses. Gravity modelling can be conveniently supported through network analysis in a GIS environment.
- **DYNAMIC SEGMENTATION:** Dynamic segmentation is a particular network model used to represent, analyze, query, and display linear features. The basic difference between dynamic segmentation and the network representations discussed above is that dynamic segmentation has the flexibility to associate an attribute to a portion of an arc or several arcs (*e.g.,* through the definition of a route). Dynamic segmentation is commonly used to model linear features such as highways, river networks, power lines, city streets, and telephone lines.

This page intentionally left blank

Chapter 9

Geographic Query and Analysis



How many houses are within 50 m of this junction?

How many children live in this 100 m grid square?

Which households fall within the floodplain?

Spatial analysis is in many ways the crux of GIS, because it includes all of the transformations, manipulations, and methods that can be applied to geographic data to add value to them, to support decisions, and to reveal patterns and anomalies that are not immediately obvious – in other words, spatial analysis is the process by which we turn raw data into useful information. If GIS is a method of communicating information about the Earth's surface from one person to another, then the transformations of spatial analysis are ways in which the sender tries to inform the receiver, by adding greater informative content and value, and by revealing things that the receiver might not otherwise see. Some methods of spatial analysis were developed long before the advent of GIS, and carried out by hand, or by the use of measuring devices like the ruler. The term analytical cartography is sometimes used to refer to methods of analysis that can be applied to maps to make them more useful and informative, and spatial analysis using GIS is in many ways its logical successor.

Spatial analysis is the crux of GIS. Spatial analysts can reveal things that might otherwise be invisible – it can make what is implicit explicit.

Here, we will look first at some definitions and basic concepts of spatial analysis. Further, we look at spatial analysis grouped into six distinct categories – queries and reasoning, measurements, transformations, descriptive summaries, optimization, and hypothesis testing.

Methods of spatial analysis can be very sophisticated, but they can also be very simple. A large body of methods of spatial analysis has been developed over the past century or so, and some methods are highly mathematical – so much so, that it might sometimes seem that mathematical complexity is an indicator of the importance of a technique. But the human eye and brain are also very sophisticated processors of geographic data, and excellent detectors of patterns and anomalies in maps and images. So the approach taken here is to regard spatial analysis as spread out along a continuum of sophistication, ranging from the simplest types that occur very quickly and intuitively when the eye and brain focus on a map, to the types that require complex software and sophisticated mathematical understanding. Spatial analysis is best seen as collaboration between the computer and the human, in which both play vital roles. Effective spatial analysis requires an intelligent user, not just a powerful computer. Spatial analysis helps us in situations when our eyes might otherwise deceive us.

There are many possible ways of defining spatial analysis, but all in one way or another express the basic idea that information on locations is essential – that analysis carried out without knowledge of locations is not spatial analysis. One fairly formal statement of this idea is: 'Spatial analysis is a set of methods whose results are not invariant under changes in the locations of the objects being analyzed'. The double negative in this statement follows convention in mathematics, but for our purposes we can remove it: 'Spatial analysis is a set of methods whose results change when the locations of the objects being analyzed change'.

On this test the calculation of an average income for a group of people is not spatial analysis, because it; in no way depends on the locations of the people. But the calculation of the center of the New Delhi's population is spatial analysis, because the results depend on knowing where all Delhi residents are located. GIS is an ideal platform for spatial analysis because its data structures accommodate the storage of object locations.

Spatial analysis can be used to further the aims of science, by revealing patterns that were not previously recognized, and that hint at undiscovered generalities and laws. Patterns in the occurrence of a disease may hint at the mechanisms that cause the disease, and some of the most famous examples of spatial analysis are of this nature, including the work of Dr. John Snow in unraveling the causes of cholera (Figure 9.1 and Box 18).

It is interesting to speculate on what would have happened today, if early epidemiologists like Snow had access to a GIS. The rules governing research today would not have allowed Snow to remove the pump handle, except after lengthy review, because the removal constituted an experiment on human subjects. To get approval, he would have to shown persuasive evidence in favour of his hypothesis, and it is doubtful that the map would have been sufficient, because several other hypotheses might have explained the pattern equally well. First, it is conceivable that the population of Soho was inherently at risk of cholera, perhaps by being comparatively elderly, or because of poor housing conditions. The map would have been more convincing if it had shown the rate of incidence, relative to the population at risk. For example, if cholera was highest among the elderly, the map could have shown the number of cases as a proportion of the population over 50 years. Second, it is still conceivable that the hypothesis of transmission through the air between carriers could have produced the same observed pattern, if the first carrier lived in the center of the outbreak. Snow could have eliminated this alternative if he had been able to produce a sequence of maps, showing the locations of cases as the outbreak developed. Both of these options involve simple spatial analysis of the kind that is readily available today in GIS. Spatial analysis in GIS provides tools that are far more powerful than the map at suggesting causes of disease.

Today the causal mechanisms of diseases like cholera, which results in short, concentrated out breaks, have long since been worked out. Much more problematic are the causal mechanisms of diseases that are rare and not sharply concentrated in space and time. This example is of inductive use of spatial analysis, to examine empirical evidence in the search for patterns that might support new theories or general principles. Other uses of spatial analysis are deductive, focusing on the testing of known theories or principles against data. A third type of application is normative, using spatial analysis to develop or prescribe new or better designs, for the locations of new retail stores, or new roads, or new manufacturing plant.

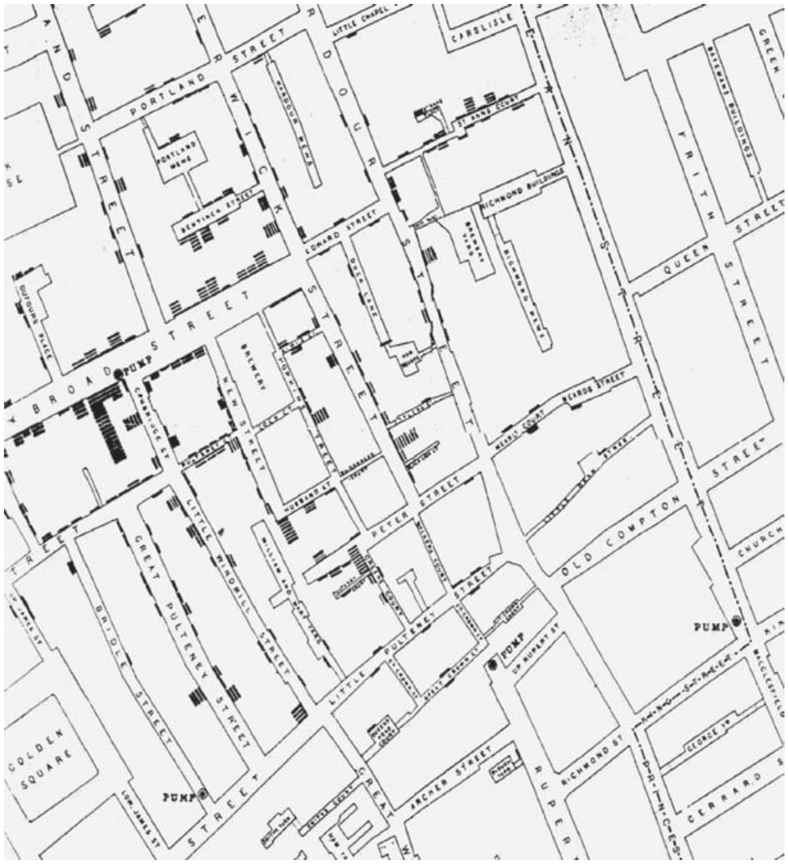


Figure 9.1: Dr. John Snow and the causes of cholera in London.

Box 18: Dr. John Snow and the causes of cholera

In the 1850s cholera was very poorly understood, and massive outbreaks were a common occurrence in major industrial cities. An outbreak in London in 1854 in the Soho district was typical of the time, and the deaths it caused are mapped in Figure 9.1. The map was made by Dr. John Snow, who had conceived the hypothesis that cholera was transmitted through the drinking of polluted water, rather than through the air, as was commonly believed. He noticed that the outbreak appeared to be centered on a public drinking water pump in Broad Street, and if his hypothesis was correct, the pattern shown on the map would reflect the locations of people who drank the pump's water. There appeared to be anomalies, in the sense that deaths had occurred in households that were located closer to other sources of water, but he was able to confirm that these households also drew their water from the Broad Street pump. Snow had the handle of the pump removed, and the outbreak subsided, providing direct causal evidence in favour of his hypothesis. This was perhaps the first use of cartographic techniques for solving a real world problem.

Box 19: Spatial relations and analysis on geometric objects

There are nine methods for testing spatial relations between geometric objects. Each takes as input two geometries and evaluates whether the relation is true or not.

Equals – are the geometries the same.

Disjoint – do the geometries share a common point

Intersects – do the geometries intersect

Touches – do the geometries intersect at their boundaries

Crosses – do the geometries overlap

Within – do the geometries within another

Contains – does one geometry completely contain another

Overlaps – do the geometries overlap

Relate – are the intersections between the interior, boundary or exterior of the geometries.

Seven methods support spatial analysis on these geometries:

Distance – determines the shortest distance between any two points in two geometries.

Buffer – returns a geometry that represents all the points whose distance from the geometry is less than or equal to a user defined distance.

Convex hull – returns a geometry representing the convex hull of a geometry (convex hull is the smallest polygon that can enclose another geometry without any concave areas).

Intersection – returns a geometry that contains just the points common to both input geometries.

Union – returns a geometry that contains all the points in both input geometries.

Difference – returns a geometry containing the points that are different between the two geometries.

SymDifference – returns a geometry containing the points that are in either of the input geometries, but not both.

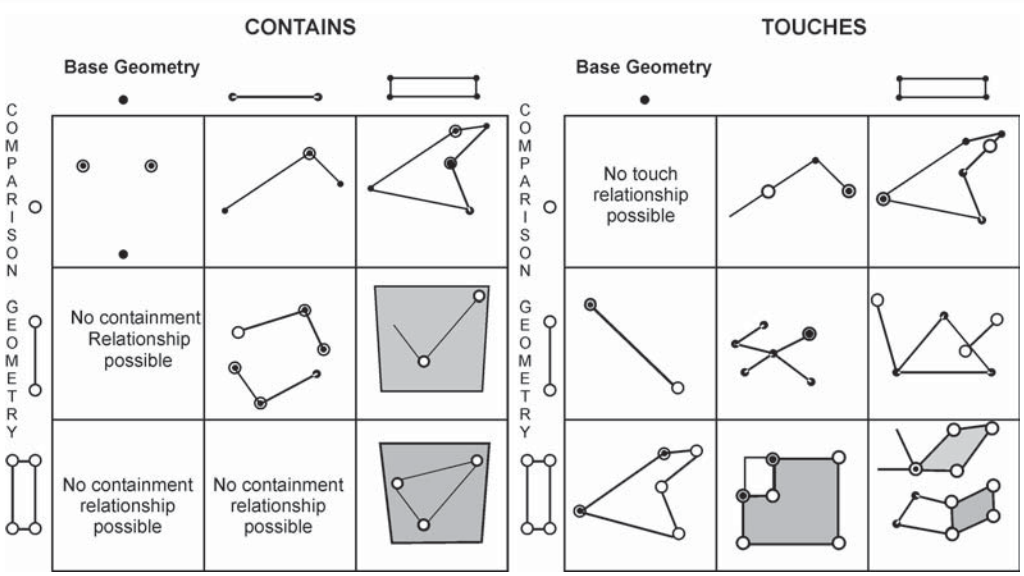


Figure 9.2: Examples of possible relations for two geographic database.

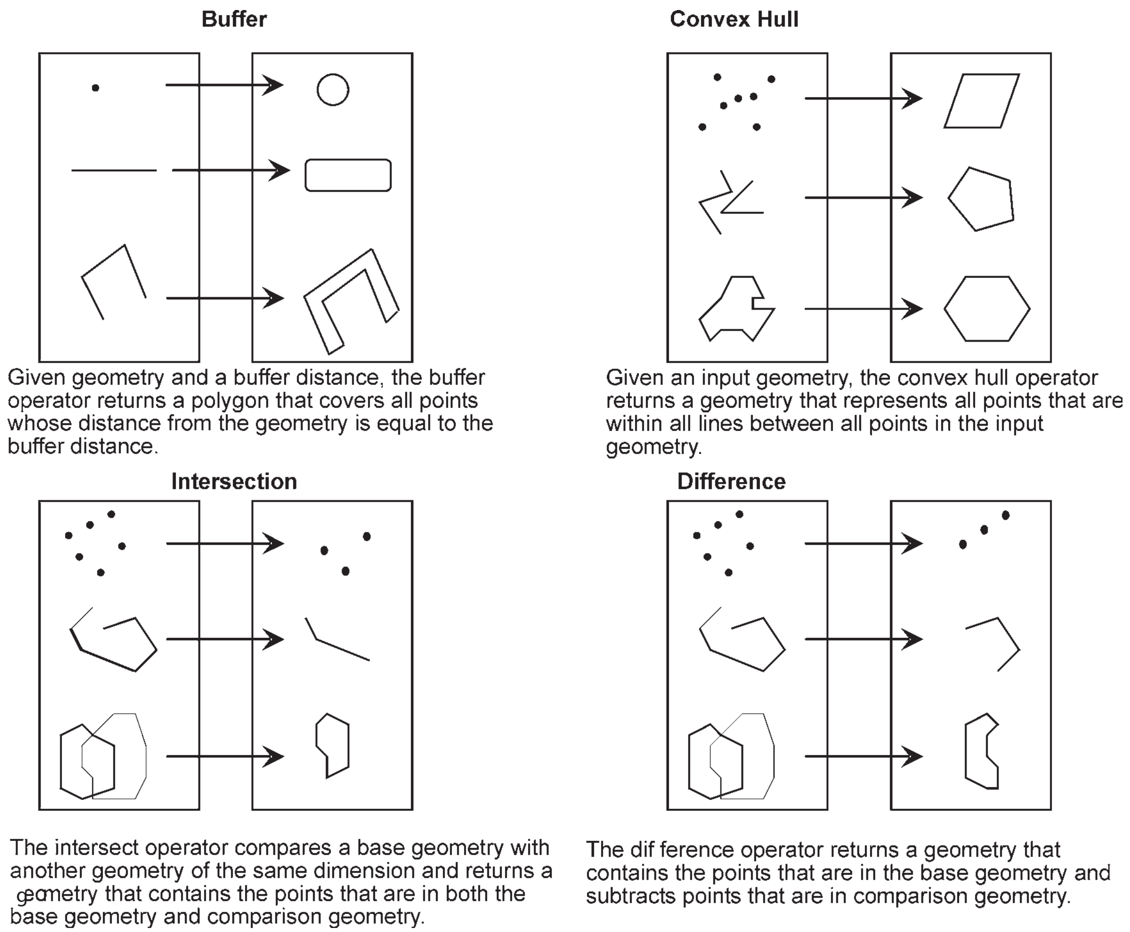


Figure 9.3: Examples of spatial analysis methods on geometries.

TYPES OF SPATIAL ANALYSIS

We would focus on methods of spatial analysis using six general headings:

i. *QUERIES AND REASONING* are the most basic of analysis operations, in which the GIS is used to answer simple questions posed by the user. No changes occur in the database, and no new data are produced. The operations vary from simple and well-defined queries like 'how many houses are found within 1 km of this point', to vaguer questions like 'which is the closest city to New Delhi going east', where the response may depend on the system's ability to understand what the user means by 'going east'.

ii. *MEASUREMENTS* are simple numerical values that describe aspects of geographic data. They include measurement of simple properties of objects, like length, area, or shape, and of the relationships between pairs of objects, like distance or direction.

iii. *TRANSFORMATIONS* are simple methods of spatial analysis that change datasets, combining them or comparing them to obtain new datasets, and eventually new insights.

Transformations use simple geometric, arithmetic, or logical rules, and they include operations that convert raster data into vector data, or vice versa. They may also create fields from collections of objects, or detect collections of objects in fields.

iv. *DESCRIPTIVE SUMMARIES* attempt to capture the essence of a dataset in one or two numbers. They are the spatial equivalent of the descriptive statistics commonly used in statistical analysis, including the mean and standard deviation.

v. *OPTIMIZATION TECHNIQUES* are normative in nature, designed to select ideal locations for objects given certain well-defined criteria. They are widely used in market research, in the package delivery industry, and in a host of other applications.

vi. *HYPOTHESIS TESTING* focuses on the process of reasoning from the results of a limited sample to make generalizations about an entire population. It allows us, for example, to determine whether a pattern of points could have arisen by chance, based on the information from a sample. Hypothesis testing is the basis of inferential statistics and lies at the core of statistical analysis, but its use with spatial data is much more problematic.

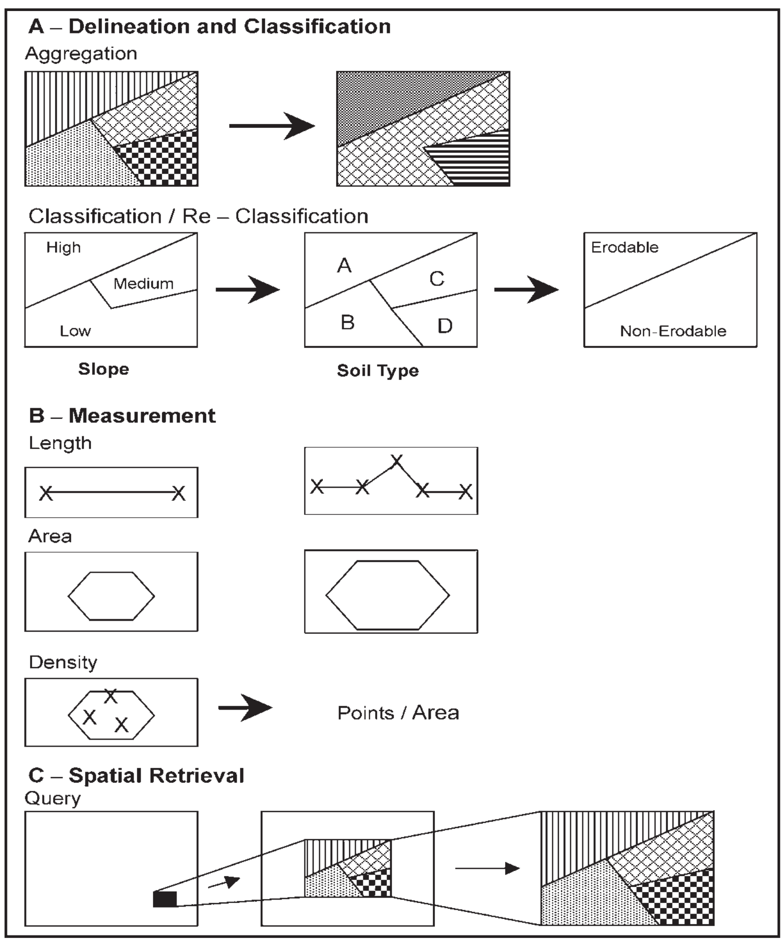


Figure 9.4: Spatial retrieval, delineation and classification, and measurement are separate functions, but are commonly used together.

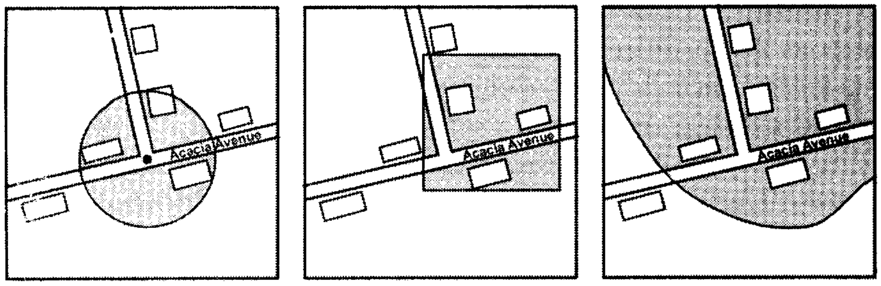


Figure 9.5: Some examples of spatial query.

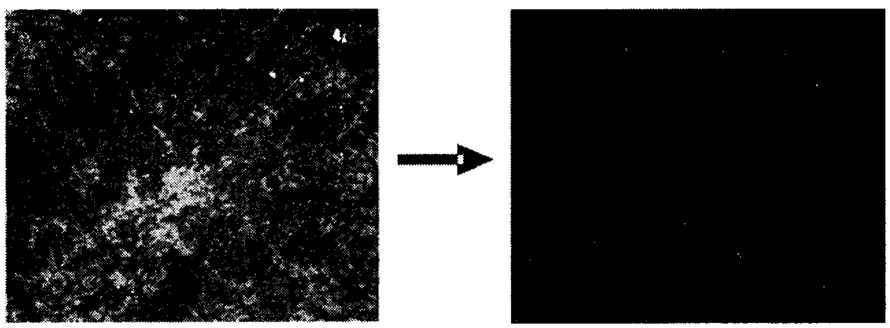


Figure 9.6: Example of *re-classification* where the modification in attribute values are made to produce new object data sets.

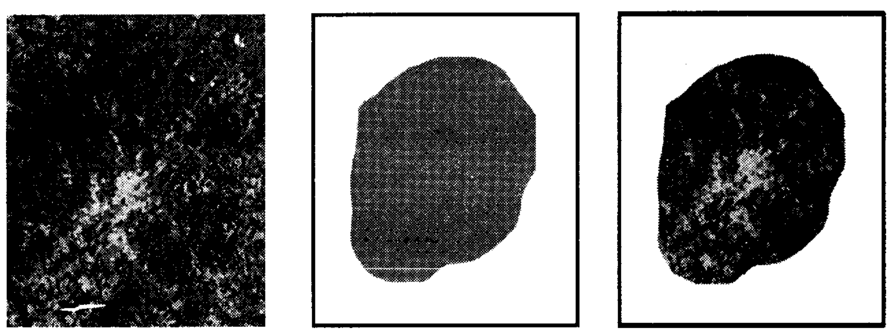


Figure 9.7: Example of *Cookie Cutting* where overlaying of datasets is made, using one dataset as a sieve or cookie cutter to select a subset of the other dataset.

QUERIES AND REASONING

In the ideal GIS it should be possible for the user to interrogate the system about any aspect of its contents, and obtain an immediate answer. Interrogation might involve pointing at a map, or typing a question, or pulling down a menu and clicking on some buttons, or sending a formal SQL request to a database. Today's user interfaces are very versatile, and have very nearly reached the point where it will be possible to interrogate the system by speaking to it – this would be extremely valuable in vehicles, where the use of more conventional ways of interrogating the system through keyboards or pointing devices can be too distracting for the driver.

The very simplest kinds of queries involve interactions between the user and the various views that a GIS is capable of presenting. A Catalog view shows the contents of a database, in the form of storage devices (hard drives, Internet sites, floppies, CDs, or ZIP disks) with their associated folders, and the datasets contained in those folders. The Catalog will likely be arranged in a hierarchy, and the user is able to expose or hide various branches of the hierarchy by clicking at appropriate points. Different types of datasets are symbolized using different icons, so the user can tell at a glance which files contain grids, polygons, points, etc.

In contemporary software environments, such as Microsoft's Windows or Windows NT, or the Macintosh or Unix environments, many kinds of interrogation are available through simple pointing and clicking. For example, in ArcCatalog simply pointing at a dataset icon and clicking the right mouse button exposes basic statistics on the dataset when the Properties option is selected. The metadata option exposes the metadata stored with the dataset, including its projection and datum details, the names of each of its attributes, and its date of creation.

The map view of a dataset shows its contents in visual form, and opens many more possibilities for querying. When the user points to any location on the screen the GIS display the pointer's coordinates, using the units appropriate to the dataset's projection and coordinate system. Today's GIS supports much more sophisticated forms of query than these. Suppose both the map view and the table views are displayed on the screen simultaneously. Linkage allows the user to select objects in one view, perhaps by pointing and clicking, and to see the selected objects highlighted in both views. Linkage is often possible between other views, including the histogram and scatterplot views. For example, by linking a scatterplot with a map view, it is possible to select points in the scatterplot and see the corresponding objects highlighted on the map. This kind of linkage is very useful in examining residuals, or cases that deviate substantially from the trend shown by a scatterplot. The term exploratory spatial data analysis is sometimes used to describe these forms of interrogation, which allow the user to explore data in interesting and potentially insightful ways. Exploratory spatial data analysis allows its users to gain insight by interacting with dynamically linked views.

Second, many methods are commonly available for interrogating the contents of tables, such as SQL. SQL is a standard language for querying tables and relational databases. The language becomes much more powerful when tables are linked, using common keys, and much more complex and sophisticated queries, involving multiple tables, are possible with the full language. More complex methods of table interrogation include the ability to average the values of an attribute across selected records, and to create new attributes through arithmetic operations on existing ones (*e.g.,* create a new attribute equal to the ratio of two selected attributes).

The term reasoning encompasses a collection of methods designed to respond to more complex forms of query and interrogation. Humans have sophisticated abilities to reason with spatial data, often learned in early childhood, and if computers could be designed to emulate these abilities then many useful applications would follow. One is in the area of navigation. Humans are very skilled at direction giving, and computer emulation of these skills would be useful in the design of in-vehicle navigation systems. The difference between the directions given orally to a person and in GIS is obvious that the human's are given in familiar terms, and they use many more landmarks and hints designed to make the driver's task less error-prone and to allow the driver to recover from mistakes. They also use gestures such as pointing that cannot be easily represented in digital form.

One major difference between the two sets is in the use of vague terms. Computers are generally uncomfortable with vagueness, preferring the precise terms (like, start from Geography department, turn right from V.C. lodge, stay straight up to University circle, turn left on Dodhpur road, turn left on to Medical road and stay straight up to Zakaria market). But the world of human communication is inherently vague and full of terms and phrases like 'near', 'north', 'too far', or 'watch out for' that defy precise definition. Very often the meaning of human terms depends on the context in which they are used. For example, Agra may be 'near' New Delhi in a conversation in Chennai, but not in a conversation in Aligarh.

Measurements

Many types of interrogation ask for measurements – we might want to know the total area of a parcel of land, or the distance between two points, or the length of a stretch of road – and in principle all of these measurements are obtainable by simple calculations inside a GIS. Comparable measurements by hand from maps can be very tedious and error prone. In fact it was the ability of the computer to make accurate evaluations of area quickly that led the Canadian government to fund the development of the world's first GIS, the Canada Geographic Information System, in the mid-1960s, despite the primitive state and high costs of computing at that time. Evaluation of area by hand is a messy and tedious job. The dot-counting method uses transparent sheets on which randomly located dots have been printed – all area on the map is estimated by counting the number of dots falling within it. In the planimeter method a mechanical device is used to trace the area's boundary, and the required measure accumulates on a dial on the machine.

Box 20: *What is an algorithm?*

Algorithm is a procedure consisting of a set of unambiguous rules which specify a finite sequence of operations that provides the solution to a problem, or to a specific class of problems. Each step of an algorithm needs to be unambiguous and precisely defined and the actions to be carried out must be rigorously specified for each case. An algorithm always arrives at a problem solution after a finite and reasonable number of steps. An algorithm that satisfies these requirements can be programmed as software for a computer.

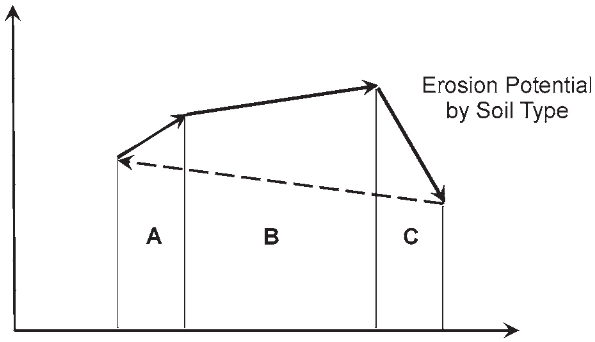


Figure 9.8: The algorithm for calculation of the area of a polygon given the coordinates of the polygon's vertices. The polygon consists of the three arrows and one arrow with dashed line forming the fourth side. Trapezia are dropped from each edge to the x axis and their areas are calculated as (difference in x) times average of y. The trapezia for the first three edges, shown in 'A' 'B' and 'C', are summed. When the fourth trapezia is formed from the dashed arrow its area is negative because its start point has a larger x than its end point. When this area is subtracted from the total, the result is the correct area of the polygon.

By comparison, measurement of the area of a digitally represented polygon is trivial and totally reliable. The common algorithm (Box 20) calculates and sums the areas of a series of trapezia, formed by dropping perpendiculars to the x axis as shown in Figure 9.8. By making a simple change to the algorithm it is also possible to use it to compute a polygon's centroid.

Distance and Length

A metric is a rule for the determination of distance between points in a space. Several kinds of metrics are used in GIS, depending on the application. The simplest is the rule for determining the shortest distance between two points in a flat plane, called the pythagorean or straight-line metric. If the two points are defined by the coordinates (X1, Y1) and (X2, Y2), then the distance D between them is the length of the hypotenuse of a right-angled triangle (Figure 9.9), and pythagoras's theorem tells us that the square of this length is equal to the sum of the squares of the lengths of the other two sides. So a simple formula results:

$$D = \sqrt{(x_2\ x_1)2\ (y_2\ y_1)^2}$$

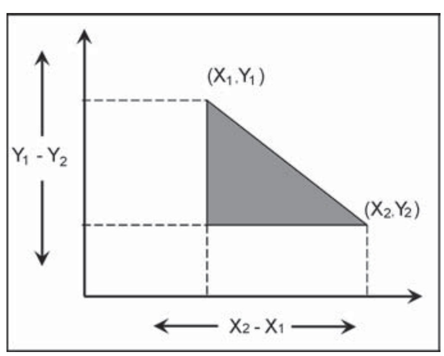


Figure 9.9: Pythagoras's theorem and the straight-line distance between two points.

The Pythagorean metric gives a simple and straightforward solution for a plane, if the coordinates X and Y are comparable, as they are in any coordinate system based on a projection, such as the UTM. But the metric will not work for latitude and longitude, reflecting a common source of problems in GIS – the temptation to treat latitude and longitude as if they were equivalent to plane coordinates.

Distance between two points on a curved surface such as that of the Earth requires a more elaborate approach. The shortest distance between two points is the length of a taut string stretched between them, and if the surface is spherical that is the length of the arc of the great circle between them (the circle formed by slicing the sphere through the center and through the two points).

Given latitude and longitude for two points, the length of this arc is:

$$D = R \cos^{-1} [\sin Ø1 \sin Ø2 + \cos Ø1 \cos Ø2 \cos(\lambda_1 - \lambda_2)]$$

where R is the radius of the Earth (6378 km to the nearest km and assuming a spherical Earth). In some cases it may be necessary to use the ellipsoid model of the Earth, in which case the calculation of distance is more complex.

In many applications the simple rules – the Pythagorean and great circle equations–are not sufficiently accurate estimates of actual travel distance, and we are forced to resort to summing the actual lengths of travel routes. In GIS this normally means summing the lengths of links in a network representation, and many forms of GIS analysis use this approach. If a line is represented as a polyline, or a series of straight segments, then its length is simply the sum of the lengths of each segment, and each segment length can be calculated using the pythagorean formula and the coordinates of its endpoints. But here two problems arise with this simple approach.

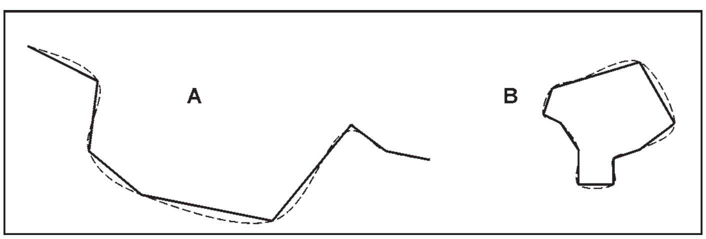


Figure 9.10: (A) – The polyline representations of smooth curves tend to be shorter in length.
(B) – But estimates of area tend not to show systematic bias because of the effects of overshoots and undershoots cancel out.

First, a polyline is often only a rough version of the true object's geometry. A river, for example, never makes sudden changes of direction, and Figure 9.10 shows, how smooth curves have to be approximated by the sharp corners of a polyline. Because there is a tendency for polylines to short-cut corners, the length of a polyline tends to be shorter than the length of the object it represents. There are some exceptions, of course – surveyed boundaries are often truly straight between corner points, and streets are often truly straight between intersections. But in general the lengths of linear objects estimated in a GIS, and this includes the lengths of the perimeters of areas represented as polygons, are often substantially shorter than their counterparts on the ground. Note that this is not similarly true of area estimates, because shortcutting corners tends to produce both underestimates and overestimates of area, and these tend to cancel out (Figure 9.10)

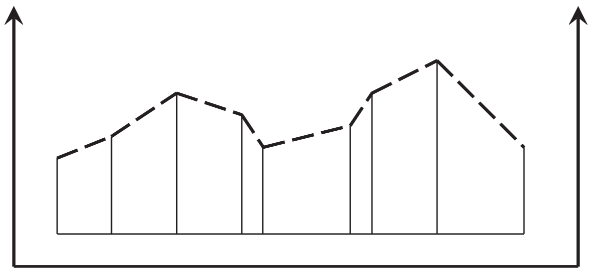


Figure 9.11: The length of a path on earth's surface (dashed line) remains longer than the length of its horizontal projection.

Second, the length of a line in a two-dimensional GIS representation will always be the length of the line's planar projection, not its true length in three dimensions, and the difference can be substantial if the line is steep (Figure 9.11). In most jurisdictions the area of a parcel of land is the area of its horizontal projection, not its true surface area. A GIS that stores the third dimension for every point is able to calculate both versions of length and area, but not a GIS that stores only the two horizontal dimensions.

Shape

GIS are also used to calculate the shapes of objects, particularly area objects. In many countries the system of political representation is based on the concept of constituencies, which are used to define who will vote for each place in the legislature. In the USA and also in India, and in many other countries that derived their system of representation from the UK, there is one place in the legislature for each district. It is expected that districts will be compact in shape, and the manipulation of a district's shape to achieve certain overt or covert objectives is termed Gerrymandering, after an early governor of Massachusetts, Elbridge Gerry (the shape of one of the state's districts was thought to resemble a salamander, with the implication that it had been manipulated to achieve a certain outcome in the voting; The construction of voting districts is an example of the principles of aggregation and zone design.

Anomalous shape is the primary means of detecting gerrymanders of political districts. An easy way to define shape is by comparing the perimeter length of an area to its area measure. Normally the square root of area is used, to ensure that the numerator and denominator are both measured in the same units. A common measure of shape or compactness is:

$$S = P/3.54\sqrt{A}$$

where P is the perimeter length and A is the area.

The factor 3.54 (twice the square root of π) ensures that the most compact shape, a circle, returns a shape of 1.0, and the most distended and contorted shapes return much higher values.

Slope and Aspect

The most versatile and useful representation of terrain in GIS is the digital elevation model, or DEM. This is a raster representation, in which each grid cell records the elevation of the Earth's surface, and reflects a view of terrain as a field of elevation values. The elevation recorded is often the elevation of the cell's central point, but sometimes it is the mean elevation of the cell, and other rules have been used to define the cell's elevation (the rules used to define elevation in each cell of the US Geological Survey's GTOPO30 DEM, which covers the entire Earth's surface, vary depending on the source of data.

Knowing the exact elevation of a point above sea level is important for some applications, including prediction of the effects of global warming and rising sea levels on coastal cities, but for many applications the value of a DEM lies in its ability to produce derivative measures through transformation, specifically measures of slope and aspect, both of which are also conceptualized as fields. Imagine taking a large sheet of plywood and laying it on the Earth's surface so that it touches at the point of interest. The magnitude of steepest tilt of the sheet defines the slope at that point, and the direction of steepest tilt defines the aspect. This sounds straightforward, but it is complicated by a number of issues. First, what if the plywood fails to sit firmly on the surface, but instead pivots, because the point of interest happens to be a peak, or a ridge? In mathematical terms, we say that the surface at this point lacks a well-defined tangent, or that the surface at this point is not differentiable, meaning that it fails to obey the normal rules of continuous mathematical functions and differential calculus. The surface of the Earth has numerous instances of sharp breaks of slope, rocky outcrops, cliffs, canyons, and deep gullies that defy this simple mathematical approach to slope, and this is one of the issues that led Benoit Mandelbrot to develop his theory of fractals, or mathematical functions that display behaviours of this nature. Mandelbrot argues in his books (Mandelbrot 1977, 1983) that many natural phenomena are fundamentally incompatible with traditional mathematics, and need a different approach.

A simple and satisfactory alternative is to take the view that slope must be measured at a particular resolution. To measure slope at a 30 meters resolution, for example, we evaluate elevation at points 30 meters apart and compute slope by comparing them. The value this gives is specific to the 30 meters spacing, and a different spacing would have given a different result. In other words, slope is a function of resolution, and it makes no sense to talk about slope without at the same time talking about a specific resolution or level of detail. This is convenient, because slope is easily computed in this way from a DEM with the appropriate resolution.

Second, there are several alternative measures of slope, and it is important to know which one is used in a particular software package and application. Slope can be measured as an angle, varying from 0 to 90 degrees as the surface ranges from horizontal to vertical. But it can also be measured as a percentage or ratio, defined as rise over run, and unfortunately there are two different ways of defining run. Figure 9.12 shows the two options, depending on whether run means the horizontal distance covered between two points, or the diagonal distance (the adjacent or the hypotenuse of the right-angled triangle respectively). In the first case (opposite over adjacent) slope as a ratio is equal to the tangent of the angle of slope, and ranges from zero (horizontal) through 1 (45 degrees) to infinity (vertical). In the second case (opposite over hypotenuse) slope as a ratio is equal to the sine of the angle of slope, and ranges from zero (horizontal) through 0.707 (45 degrees) to 1 (vertical). To avoid confusion we will use the term slope only to refer to the measurement in degrees, and call the other options tan(slope) and sin(slope) respectively.

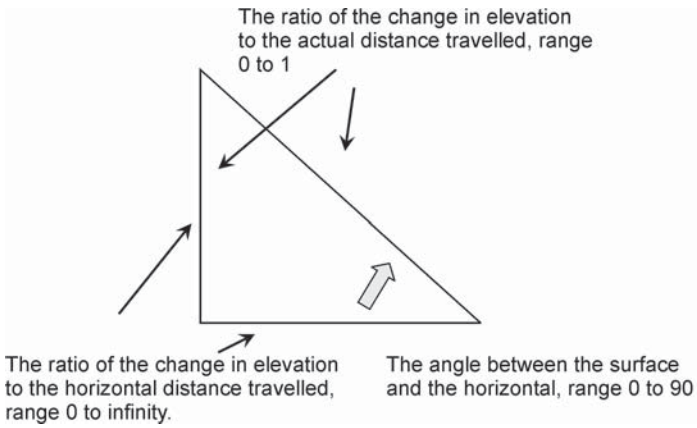


Figure 9.12: Three alternative definitions of slope.

When a GIS calculates slope and aspect from a DEM, it does so by estimating slope at each of the data points of the DEM, by comparing the elevation at that point to the elevations of surrounding points. But the number of surrounding points used in the calculation varies, as do the weights given to each of the surrounding points in the calculation.

Slope and aspect are the basis for many interesting and useful forms of analysis. Slope is an input to many models of the soil erosion and runoff that result from heavy storms. Slope is also an important input to analyses that find the most suitable routes across terrain for power lines, highways etc.

Transformations

In this section, we look at methods that transform GIS objects and databases into more useful products, using simple rules. These operations form the basis for many applications, because they are capable of revealing aspects that are not immediately visible or obvious.

Buffering

One of the most important transformations available to the GIS user is the buffer operation. Given any set of objects, which may include points, lines, or areas, a buffer operation builds a new object or objects by identifying all areas that are within a certain specified distance of the original objects. Figure 9.13 shows instances of a point, a line, and an area, and the results of buffering. Buffers have many uses, and they are among the most popular of GIS functions:

- A builder wishes to develop a residential colony, but is concerned of flooding in rainy season. He is required to avoid construction within 100 meters of streams – the builder could build buffers 100 meters wide around all streams to identify these flooding areas.

- A retailer is considering developing a new store on a site, of a type that is able to draw consumers from up to 4 km away from its stores the retailer could build a buffer around the site to identify the number of consumers living within 4 km of the site, in order to estimate the new store's potential sales.

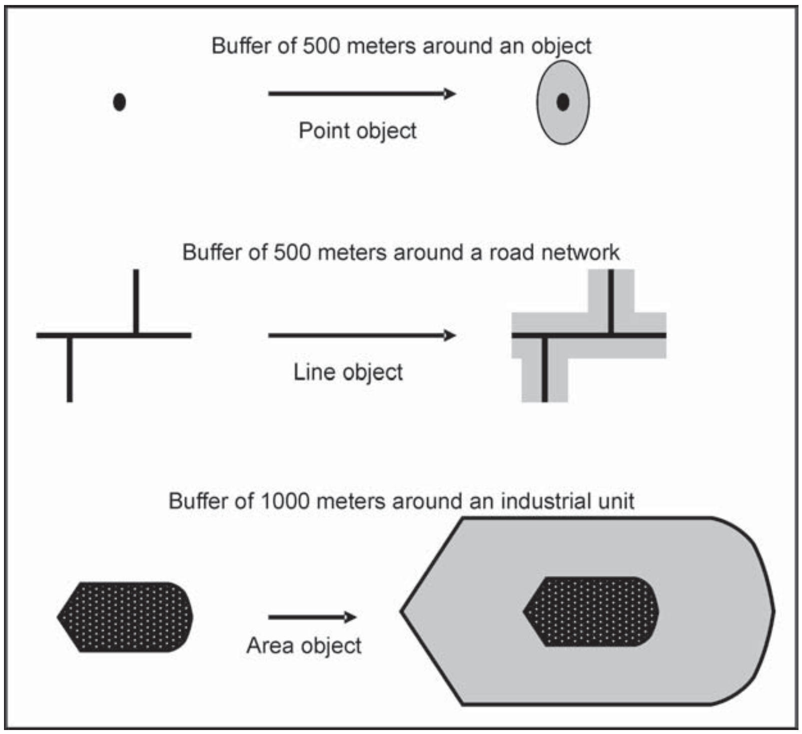


Figure 9.13: Buffers of constant width drawn around a point, line and a polygon.

Buffering is possible in both raster and vector GIS, in the raster case, the result is the classification of cells according to whether they lie inside or outside the buffer, while the result in the vector case is a new set of objects. But there is an additional possibility in the raster case that makes buffering more useful in some situations. Figure 9.15 shows a city; average travel speeds vary in each cell of the raster outside the city. Rather than buffer according to distance from the city, we can ask a raster GIS to spread outwards from the city at rates determined by the travel speed values in each cell. Where travel speeds are high the spread will extend further, so we can compute how far it is possible to go from the city in a given period of time. This idea of spreading over a variable surface is easily implemented in raster representations, but impossible in vector representations.

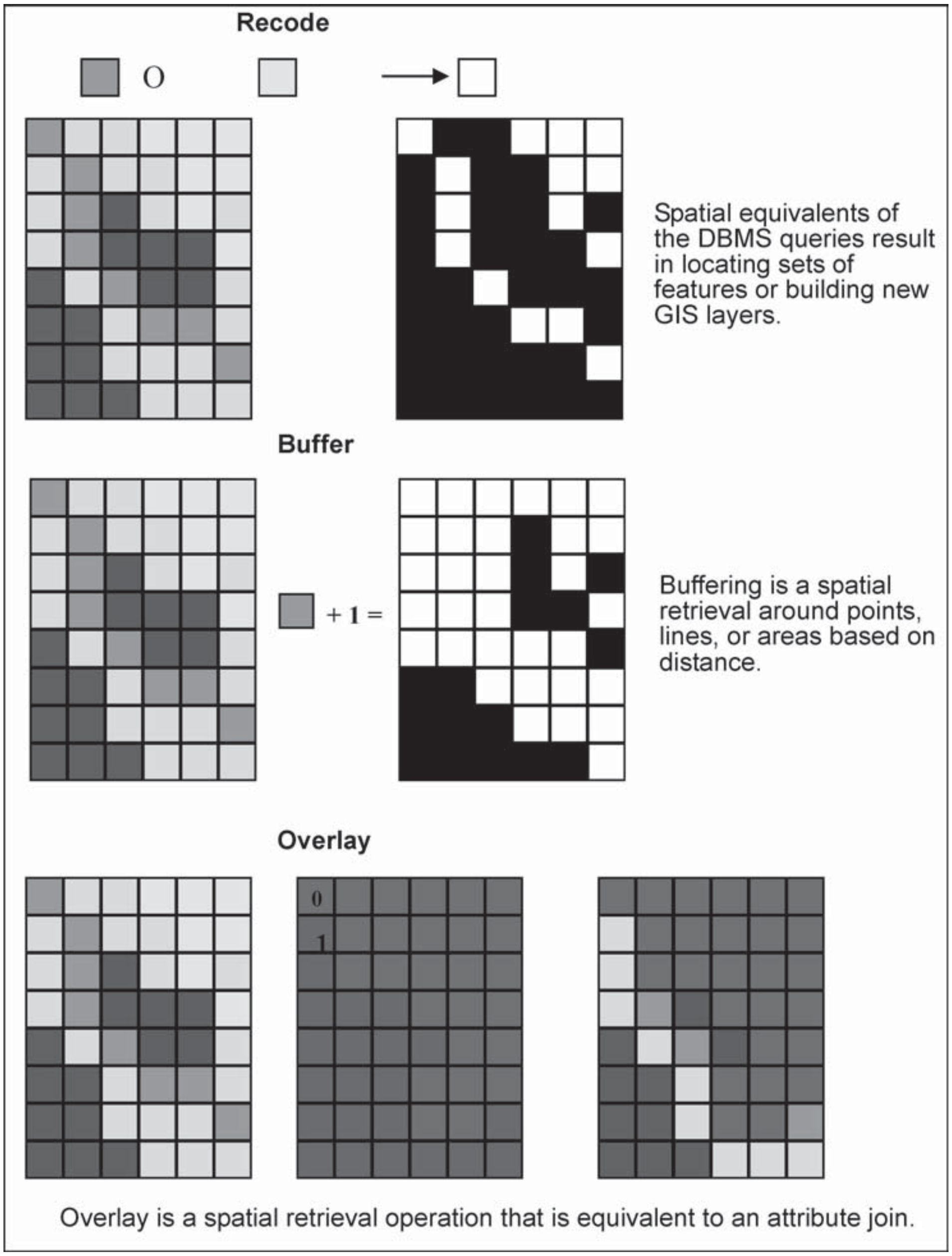


Figure 9.14: Spatial retrieval operations.

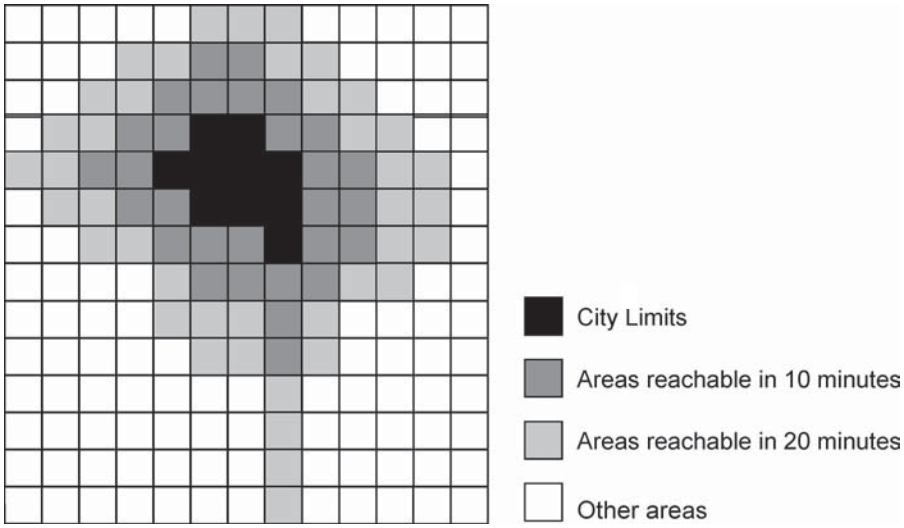


Figure 9.15: A raster generalization of the buffer function where changes may be controlled by some variable (example is of travel speed, whose value is recorded in every raster cell)

Point in Polygon

In its simplest form, the point in polygon operation determines whether a given point lies inside or outside a given polygon. In more elaborate forms there may be many polygons, and many points, and the task is to assign points to polygons. If the polygons overlap, it is possible that a given point lies in one, many, or no polygons, depending on its location. Figure 9.17 illustrates the task. The operation is popular in GIS analysis because it is the basis for answering many simple queries:

- The points represent instances of a disease in a population, and the polygons represent reporting zones such as wards-the task is to determine how many instances of the disease occurred in each ward (in this case the ward should not overlap and each point should fall into exactly one polygon).
- The points represent the locations of a tube-well owned by a person, and the polygons are parcels of land-the task is to determine the owner of the land where tube-well lies has necessary permission and owner of the land which is irrigated by tube-well has paid the necessary fees.
- The points represent the residential locations of an industry, and the polygons represent the entire settlement-the task is to ensure that each worker of the industry receives the invitation for a function by mail.

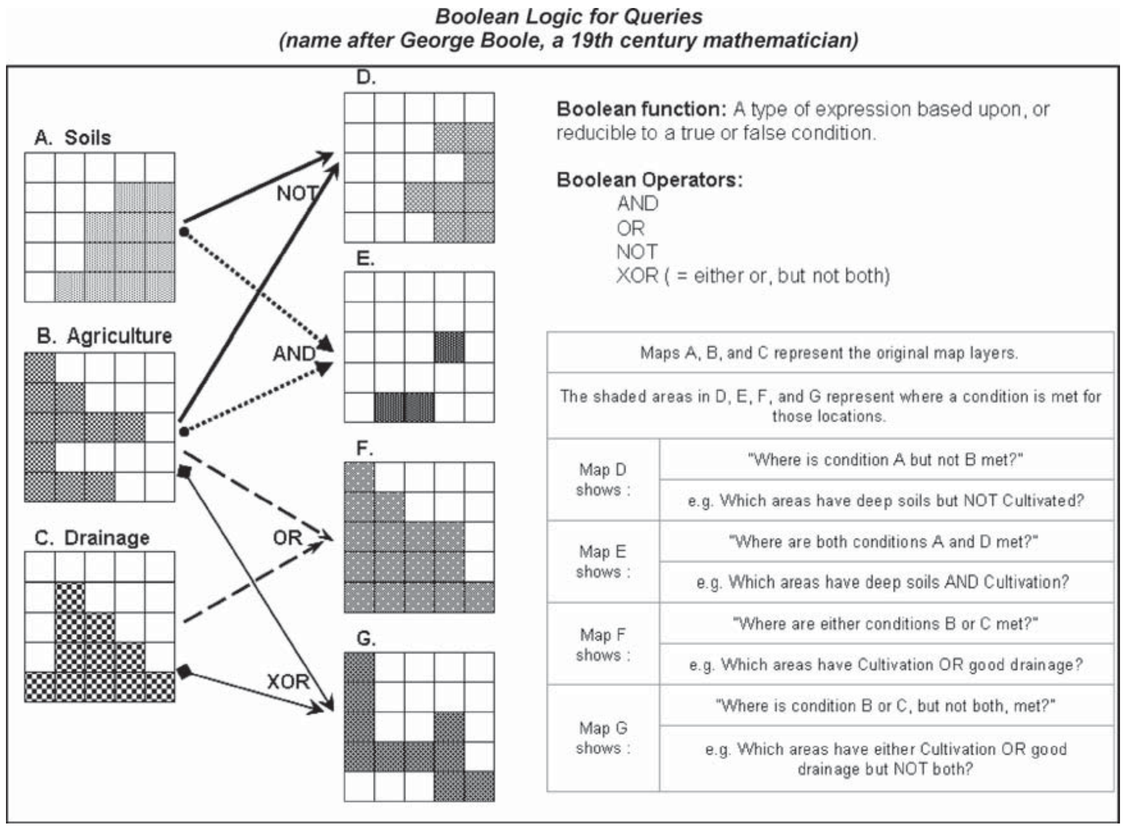


Figure 9.16: Examples of Boolean logic using Boolean operators.

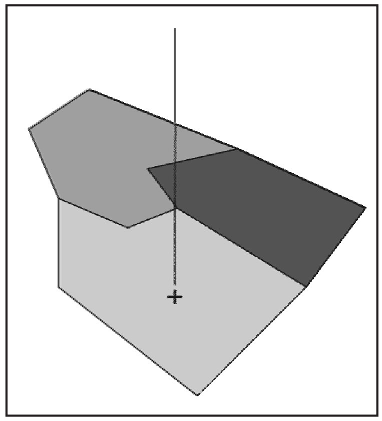


Figure 9.17: The point in polygon problem.

The point in polygon operation makes sense from both the discrete object and the field perspectives. From a discrete object perspective both points and polygons are objects, and the task is simply to determine enclosure. From a field perspective, polygons representing a variable such as land ownership cannot overlap, since each polygon represents the land owned by one owner, and overlap would imply that a point is owned simultaneously by two owners. Similarly from a field perspective there can be no gaps between polygons. Consequently, the result of a point in polygon operation from a field perspective must assign each point to exactly one polygon.

The standard algorithm for the point in polygon operation is shown in Figure 9.14. In essence, it consists of drawing a line vertically upwards from the point, and determining the number of intersections between the line and the polygon's boundary. If the number is odd the point is inside the polygon, and if it is even the point is outside. The algorithm must deal successfully with special cases, for example, if the point lies directly below a corner point of the polygon. Some algorithms extend the task to include a third option, when the point lies exactly on the boundary. But others ignore this, on the grounds that it is never possible to determine location with perfect accuracy, and so never possible to determine if an infinitely small point lies on an infinitely thin boundary line.

Polygon Overlay

Polygon overlay is similar to point in polygon transformation in the sense that two sets of objects are involved, but in this case both are polygons. It exists in two forms, depending on whether a field or discrete object perspective is taken. From the discrete object perspective, the task is to determine whether two area objects overlap, to determine the area of overlap, and to define the area formed by the overlap as one or more new area objects (the overlay of two polygons can produce a large number of distinct area objects, see Figure 9.18). This operation is useful to determine answers to such queries as:

- How much area lies in the shaded zone?
- How much of this land parcel is shaded but not the white polygon?
- What proportion of the land area outside the shaded but inside the white polygon?

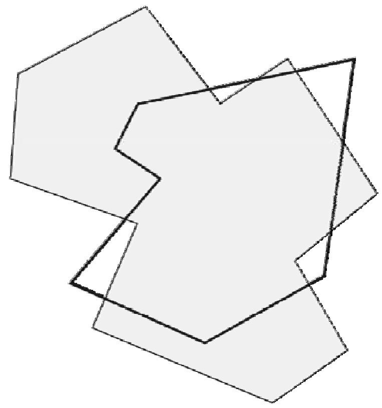


Figure 9.18: An example of polygon overlay, in the discrete object case. Here the overlay of two polygons produces nine polygons. One has the property of both, four have the properties of shaded but not the white polygon and four are outside the shaded but inside the white polygon.

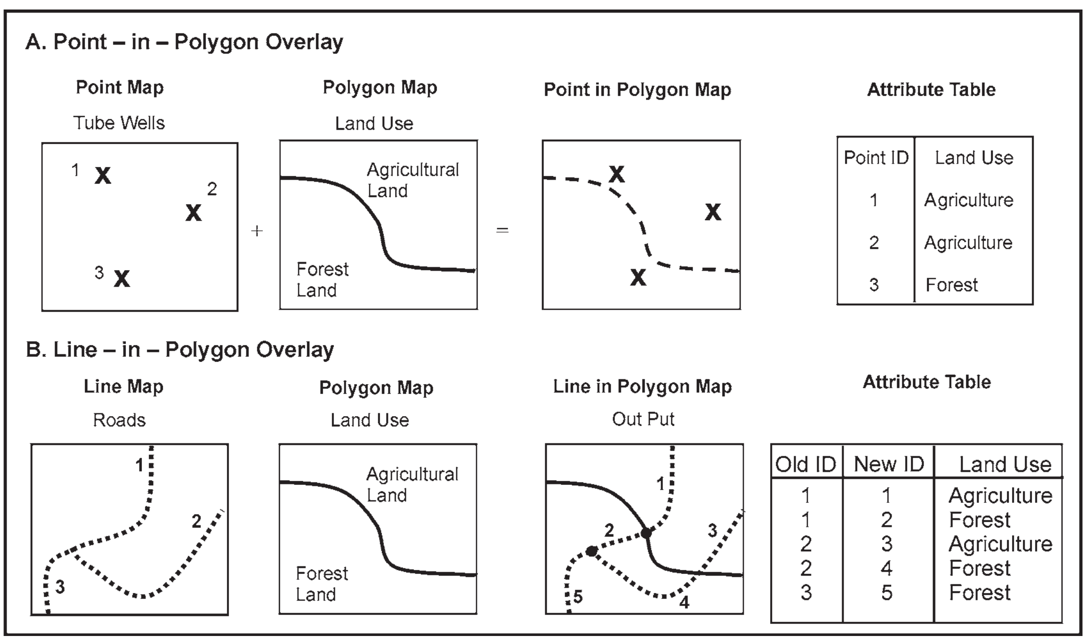


Figure 9.19: Vector overlays (point in polygon and line in polygon).

C. Polygon – on – Polygon Overlay

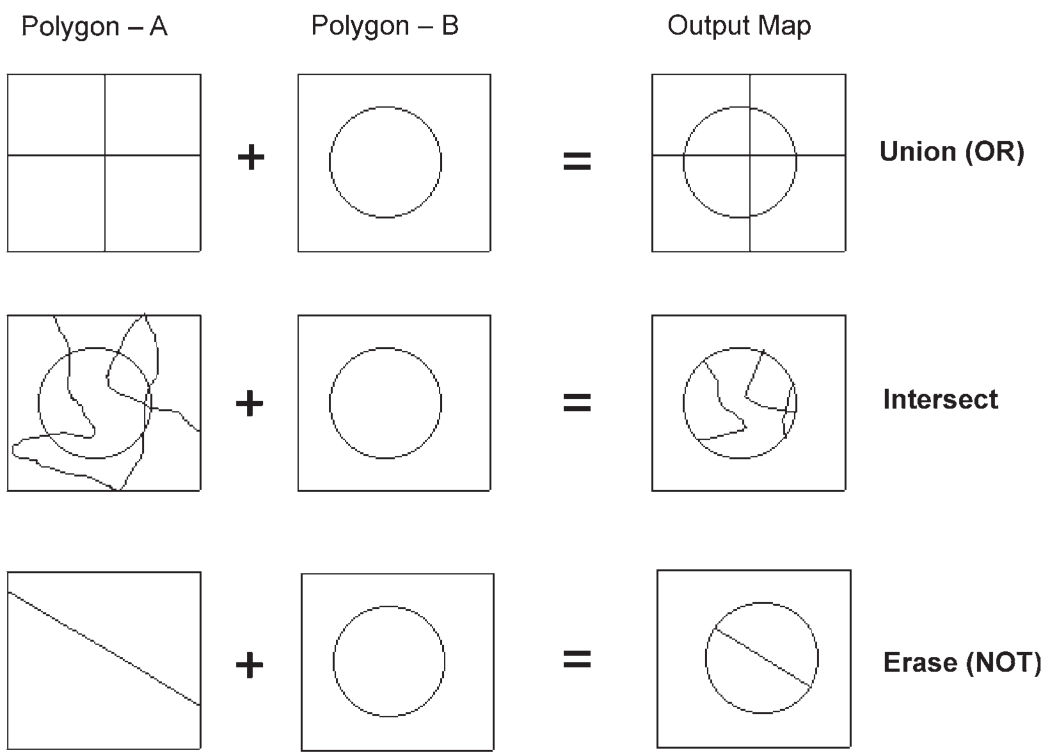


Figure 9.20: Vector overlays (polygon on polygon)

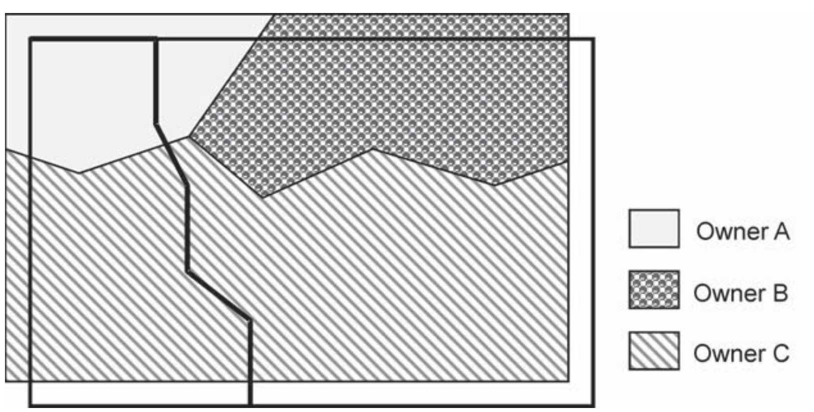


Figure 9.21: Polygon overlay in the field case. Where a dataset representing two types of land cover. (one on the left, say X and another in right, y). It is overlaid on a dataset showing three land parcels owned by three different persons. The overlay result will be a single dataset in which every point is identified with one land cover type and one ownership type. There will be five polygons, as land over X intersects two ownership types and land cover y intersects with three.

From the field perspective the task is somewhat different. Figure 9.21 shows two datasets, both representations of fields–one differentiates areas according to land ownership, and the other differentiates the same region according to land cover class. In the terminology of ESRI's Arc Info, both datasets are instances of area coverages, or fields of nominal variables represented by non-overlapping polygons. The methods discussed earlier in this chapter could be used to interrogate either dataset separately, but there are numerous queries that require simultaneous access to both datasets, for example:

⇒ What is the total area of land owned by A and with land cover class X?

⇒ Where are the areas that is owned by C and have land cover class Y?

⇒ What is the land cover class and who is the owner of the point indicated by the user?

None of these queries can be answered by interrogating one of the datasets alone the data sets must somehow be combined so that interrogation can be directed simultaneously at both of them. The field version of polygon overlay does this by first computing a new dataset in which the region is partitioned into smaller areas that have uniform characteristics on both field variables. Each area in the new dataset will have two sets of attributes – those obtained from one of the input datasets, together with those obtained from the other. All of the boundaries will be retained, but they will be broken into shorter fragments by the intersections that occur between boundaries in one input data set and boundaries in the other. Unlike the two input datasets, where boundaries meet in a junction of three lines, the new map contains a new junction of four lines, formed by the new intersection discovered during the overlay process. Because the results of overlay are distinct in this way it is almost always possible to discover whether a GIS dataset was formed by overlaying two earlier datasets.

With a single dataset that combines both inputs, it is an easy matter to answer all of the queries listed above through simple interrogation. It is also easy to reverse the overlay process-if neighbouring areas that share the same land cover class are merged. Polygon overlay is a computationally complex operation, and much work has gone into developing algorithms that function efficiently for large datasets. One of the issues that must be tackled by a practically useful algorithm is known as the spurious polygon or coastline weave problem. It is almost inevitable that there will be instances in any practical application where the same line on the ground occurs in both datasets.

Rivers and roads often form boundaries in many different datasets – a river may function both as a land cover class boundary and as a land ownership boundary, for example. But although the same line is represented in both datasets, its representations will almost certainly not be the same- They may have been digitized from different maps, subjected to different manipulations, obtained from entirely different sources (an air photograph and a topographic map), and subjected to different measurement errors. When overlaid, the result is a series of small slivers. Paradoxically, the more care one takes in digitizing or processing, the worse the problem becomes, as the result is simply more slivers, albeit smaller in size.

Today, a GIS offers various methods for dealing with the problem, the most common of which is the specification of a tolerance. If two lines fall within this distance of each other, the GIS will treat them as a single line, and not create slivers. The resulting overlay contains just one version of the line, not two. But at least one of the input lines has been moved, and if the tolerance is set too high the movement can be substantial, and can lead to problems later.

Overlay in raster is an altogether simpler operation, and this has often been cited as a good reason to adopt raster rather than vector structures. When two raster layers are overlaid, the attributes of each cell are combined according to a set of rules. For example, suppose the task is to find all areas that belong to owner A and have land use class X. Areas with these characteristics would be assigned a value, say 1, and all other areas would be assigned a value of 0. The important difference between raster and vector overlay in vector overlay there is no rule for combination, and instead the result of overlay contains all of the input information, rearranged and combined so that it can be used to respond to queries and can be subjected to analysis.

Spatial Interpolation

Spatial interpolation is a pervasive operation in GIS. Although it is often used explicitly in analysis, it is also used implicitly, in various operations such as the preparation of a contour map display, where spatial interpolation is invoked without the user's direct involvement. Spatial interpolation is a process of intelligent guesswork, in which the investigator attempts to make a reasonable estimate of the value of a field at places where the field has not actually been measured. Spatial interpolation is an operation that makes sense only from the field perspective. Spatial interpolation finds applications in many areas:

- In contouring, when it is necessary to guess where to place contours in between measured locations.
- In estimating the elevation of the surface in between the measured locations of a DEM.
- In estimating rainfall, temperature, and other attributes at places that are not weather stations, and where no direct measurements of these variables are available.
- In resampling rasters, the operation that must take place whenever raster data must be transformed to another grid.

In all of these instances spatial interpolation calls for intelligent guesswork, and the one principle that underlies all spatial interpolation is the Tobler Law-'all places are related but nearby places are more related than distant places'. In other words, the best guess as to the value of a field at some point is the value measured at the closest observation points - the rainfall here is likely to be more similar to the rainfall recorded at the nearest weather stations than to the rainfall recorded at more distant weather stations. A corollary of this same principle is that in the absence of better information, it is reasonable to assume that any field exhibits relatively smooth variation-fields tend to vary slowly, and to exhibit strong positive spatial autocorrelation, a property of geographic data.

Here we discuss two commonly used methods of spatial interpolation: inverse distance weighting (IDW), which is the simplest method; and Kriging, a popular statistical method that is grounded in the theory of regionalized variables and falls within the field of geostatistics.

INVERSE DISTANCE WEIGHTING (IDW): IDW is the workhorse of spatial interpolation, the method that is most often used by GIS analysts. It employs the Tobler law by estimating unknown measurements as weighted averages over the known measurements at nearby points, giving the greatest weight to the nearest points. IDW provides a simple way of guessing the values of a field at locations where no measurement is available.

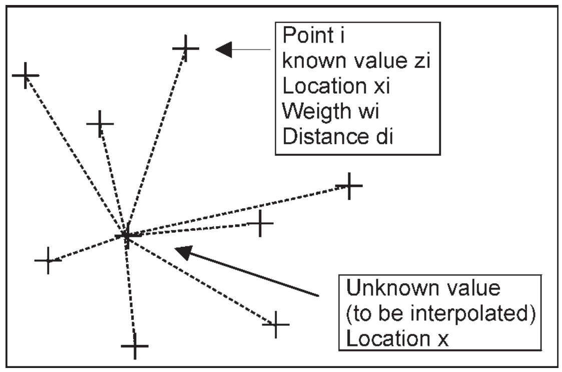


Figure 9.22: Notation used in the equations defining spatial interpolation.

IDW achieves the desired objective of creating a smooth surface whose value at any point is more like the values at nearby points than the values at distant points. If it is used to determine z at a location where z has already been measured it will return the measured value, because the weight assigned to a point at zero distance is infinite, and for this reason IDW is described as an exact method of interpolation because its interpolated results honour the data points exactly (an approximate method is allowed to deviate from the measured values in the interests of greater smoothness, a property which is often useful if deviations are interpreted as indicating possible errors of measurement, or local deviations that are to be separated from the general trend of the surface.

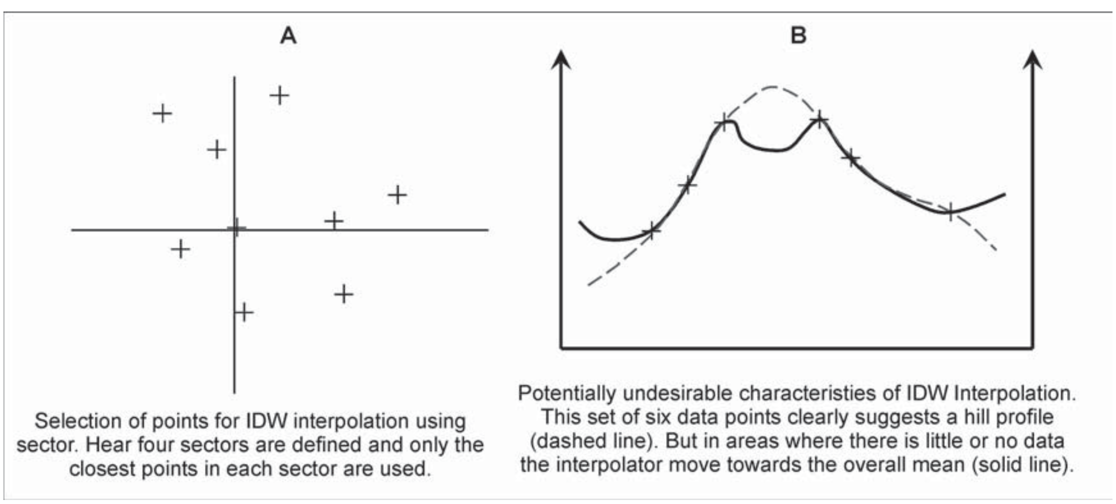


Figure 9.23: IDW interpolation.

But because IDW is an average it suffers from certain specific characteristics that are generally undesirable. A weighted average that uses weights that are never negative must always return a value that is between the limits of the measured values-no point on the interpolated surface can have an interpolated z that is more than the largest measured z, or less than the smallest measured z. IDW interpolation may produce counterintuitive results in the areas of peaks and pits, and outside the area covered by the data points.

In short, the results of IDW are not always what one would want. There are many better methods of spatial interpolation that address the problems that were just identified, but the ease of programming of IDW and its conceptual simplicity make it among the most popular.

Kriging: Of all of the common methods of spatial interpolation it is Kriging that makes the most convincing claim to be grounded in good theoretical principles. The basic idea is to discover something about the general properties of the surface, as revealed by the measured values, and then to apply these properties in estimating the missing parts of the surface.

Smoothness is the most important property, and it is operationalized in Kriging in a statistically meaningful way. There are many forms of Kriging, but all are firmly grounded in theory. Suppose we take a point x as a reference, and start comparing the values of the field there with the values at other locations at increasing distances from the reference point. If the field is smooth (if the Tobler law is true, that is, if there is positive spatial autocorrelation) the values nearby will not – very different-z(x) will not be very different from z(xi). To measure the amount, we take the difference and square it, since the sign of the difference is not important:

$$(z(x) - z(xi))^2$$

We could do this with any pair of points in the area.

As distance increases, this measure will likely increase also, and in general a monotonic (consistent) increase in squared difference with distance is observed for most geographic fields (z must be measured on a scale that is at least interval, though indicator Kriging has been developed to deal with the analysis of nominal fields). In Figure 9.24, each point represents one pair of values drawn from the total set of data points at which measurements have been taken.

The vertical axis represents one half of the squared difference (one half is taken for mathematical reasons), and the graph is known as the semivariogram (or variogram for short the difference of a factor of two is often overlooked in practice, though it is important mathematically). To express its contents in summary form the distance axis is divided into a number of ranges or buckets, as shown, and points within each range are averaged to define the heavy points shown in the figure. This semivariogram has been drawn without regard to the directions between points in a pair. Kriging responds both to the proximity of sample points and to their directions.

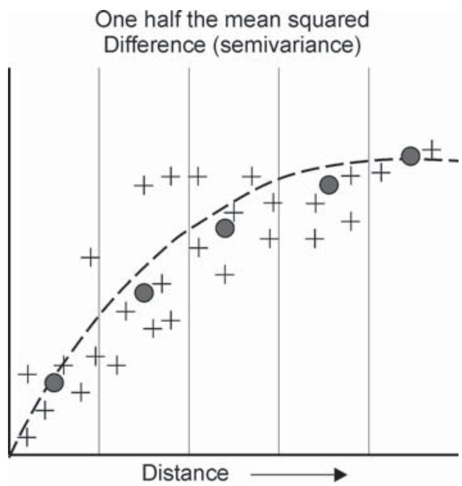


Figure 9.24: A semivariogram, here each cross represents a pair of points. The solid circles are obtained by averaging within the ranges of the distance axis. The dashed line is the best fit to the five points.

DENSITY ESTIMATION AND POTENTIAL

Density estimation is in many ways the logical twin of spatial interpolation – it begins with points, and ends with a surface. But conceptually the two approaches could not be more different, because one seeks to estimate the missing parts of a field from samples of the field taken at data points, while the other creates a field from discrete objects.

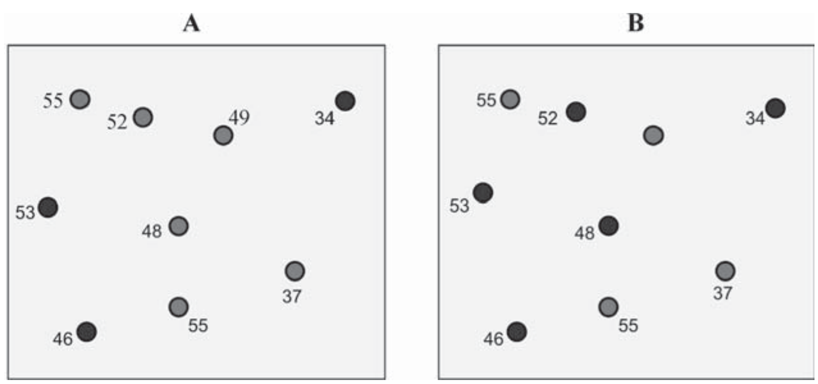


Figure 9.25: Two identical datasets but with different representations.
A – represents a field of atmospheric temperature recorded at nine sample points.
B – nine discrete objects representing population of different settlements in thousands.
Spatial interpolation is appropriate for case A, while for case B density estimation is suitable.

Figure 9.25 illustrates this difference. The two datasets in the diagram look identical from a GIS perspective – they are both sets of points, with locations and a single attribute. But one shows sample measurements from a field, and the other shows the locations of discrete objects. In the discrete object view there is nothing between the objects but empty space – no missing field to be filled in through spatial interpolation. It would make no sense at all to apply spatial interpolation to a collection of discrete objects – and no sense at all to apply density estimation to samples of a field. Density estimation makes sense only from the discrete object perspective, and spatial interpolation only from the field perspective.

Density estimation could be applied to any type of discrete spatial object, it is most often applied to the estimation of point density, and that is the focus here. The most obvious example is the estimation of population density, and but it could be equally well applied to the density of different kinds of diseases, or animals, or any other set of well-defined points.

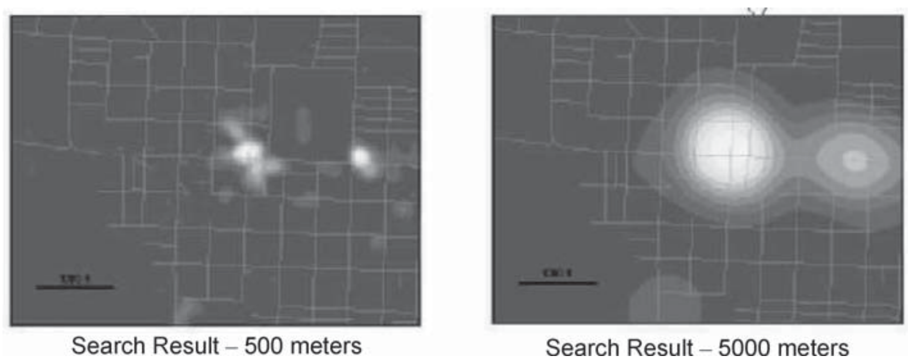


Figure 9.26: Density estimation using two different distance parameters in the respective kernel functions, displaying smoother and less peaked nature of the surface that results from the larger distance parameter.

Advanced Spatial Analysis

There are also some complex spatial analysis in GIS, which uses advanced conceptual frameworks. These spatial analysis are the outcome of advancement in technology. The advent and easy availability of large datasets and fast computing led new ways of thinking about spatial analysis. Now loads of datasets collected and archived everyday like continuous imaging of every corner of the earth or socio-economic information of population for every settlement or even the use of credit card all over the world. All this leads to thinking of interesting patterns, anomalies, truths – myths and many of these are captured in through ***data mining.*** Data mining is used to detect anomalies and patterns in vast archives of digital data. The objective of it is to find patterns that stand out from the normal in an area.

Descriptive Summaries

CENTERS: To analyze the numerical summaries generally we measure by methods of central tendency. Like **mean** is one method which is broadly citing the average of data series, similarly **median,** where the value is as such that one half of the numbers are larger and one half are smaller. Although mean can be computed only for numbers measured on interval of ratio scales, the median can be computed for ordinal data. For nominal data appropriate measure of central tendency is the **mode.**

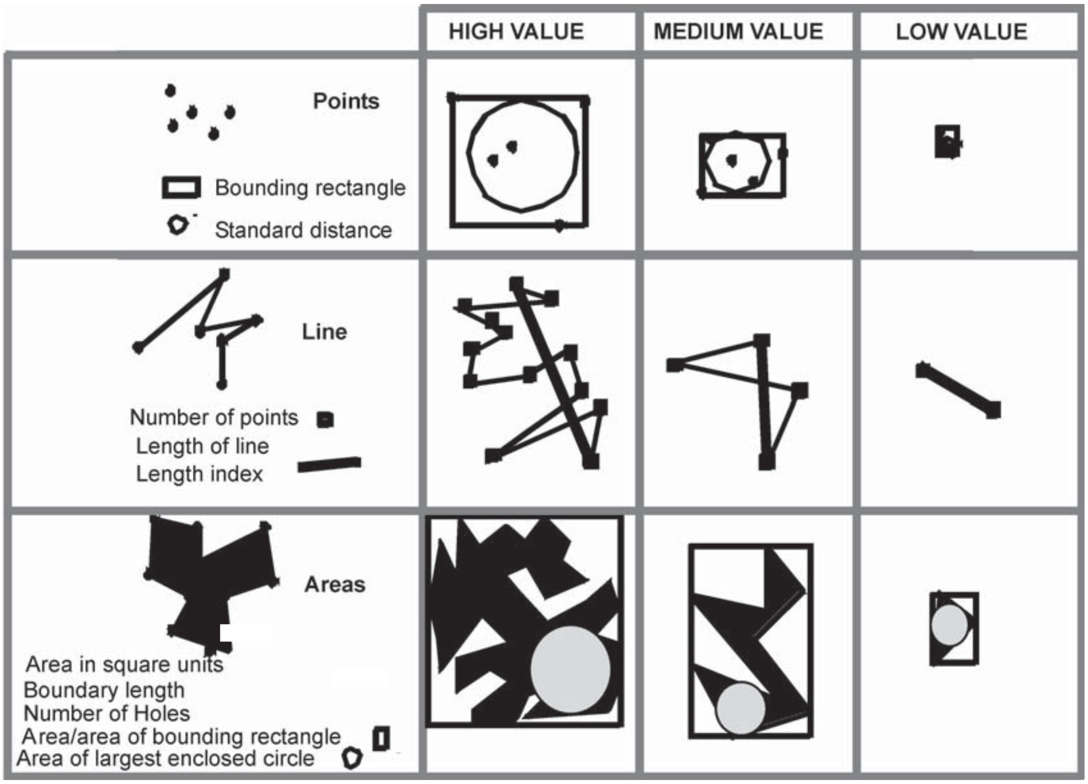


Figure 9.27: Statistics and features.

The spatial equivalent of the mean would be some kind of center, which is calculated to summarize the positions of a number of points in GIS. The center is the most convenient way of summarizing the locations of a set of points.

DISPERSION: Central tendency is the obvious choice if a set of numbers are to be summarized in a single value, but where there is opportunity for a second summary value, the measure of choice for numbers with interval or ratio properties is **standard deviation** or the **variance** is often used, which is the square of the standard deviation (the mean squared difference from the mean). But it is not convenient to measure for descriptive purpose. Standard deviation and variance are more appropriate measures of dispersion than the range because as averages they are less sensitive to the specific values of the extremes. Measures of dispersion are applied in many areas of GIS. A simple measure of dispersion in two dimensions is the mean distance from the centroid.

HISTOGRAMS AND PIE CHARTS: Histograms (bar graphs) and pie charts are two of many ways of visualizing the content of a geographic database. A histogram shows the relative frequencies of different value of an attribute by ordering them on the X axis and displaying frequency through the length of a bar parallel to the Y axis. Attributes should have interval or ratio properties, although ordinal properties are sufficient to allow the values to be ranked and histogram based on ordinal data is useful representation. A pie chart is useful for nominal data and is used to display the relative frequencies of distinct values, with no necessity for ranking. Pie charts are also useful in dealing with attributes measured on cyclic scales. Both take a single attribute and organize its values in a form that allow quick comprehension.

SCATTERPLOTS: We looked at descriptive summaries of single set of objects, further the power of GIS lies in its ability to compare sets of attributes – often thought of as the process of overlaying layers. Where we tend to explore vertical relationships (in GIS vertical refers to comparison of attributes) rather than horizontal ones. Scatterplots are useful visual summaries of relationships between attributes. It display the value of one attribute plotted against the other. If both sets of attributes belong to the same objects then the construction of a scatterplot is straight. Further if both are attributes of raster datasets, then scatterplot is built by comparing the datasets pixel by pixel. But if the attributes are from different sets of vector objects, which do not coincide in space then it is sorted by interpolating the datasets and inventing a geographic data.

SPATIAL DEPENDENCE: The fundamental problem of spatial analysis is selecting appropriate digital representations from the real world. The Tobler's first law of geography states that everything is related to everything else but near things are more related than distant things. The real world without spatial dependence is impossible to imagine. Thus, spatial dependence is crucial for GIS. It is inherently scale specific and can be measured at any spatial resolution. However, a dataset can exhibit positive spatial dependence at one scale but negative at another scale. Spatial dependence is a very useful descriptive summary of geographic data and a fundamental part of its nature. The **semivariogram** of a raster dataset elaborates how difference increases with distance and whether difference ceases to increase beyond a certain range. The computation of semivariogram in different directions, we can also determine whether a dataset displays marked anisotropy or distinct behaviours.

FRAGMENTATION AND FRACTIONAL DIMENSION: In GIS, maps may show many patches with each patch representing an area of uniform class and this may be bounded by patch of different class. For example, a soil map where we may be interested in the degree to which the landscape can be fragmented (meaning breaking in small or large patches). Fragmentation statistics provide the numerical basis for this purpose. Here we can analyze the number of patches, their shape or size etc, as a way of summarizing the geographic details. The concept of fractals is used as a way f summarizing the relationship between apparent length and level of geographic detail in the form of fractional dimensions. Smooth lines would indicate fractional dimension close to one while contorted lines would indicate towards higher values.

Optimization

Optimization is a prime example of GIS utility to support spatial decisions. It can be by many ways like optimum location of points, routing on a network, selection of optimum paths across continuous space, locating facilities etc. The methods also divide between those that are designed to locate points and routes on network and those designed to locate points and routes in continuous space without respect to the existence of roads or other links.

POINT LOCATIONS: It is an instance of location in continuous space and identifying location that minimizes total distance with respect to a number of points. The analogous problem on a network would involve finding that location on the network which minimize total distance to a number of points, also located on the network, using routes that are considered on the network using routes that are constrained to the network. Location allocation involves two types of decisions – where to locate and how to allocate demand for service.

ROUTING PROBLEMS: This is another area of optimization where routing and scheduling or decisions about he optimum tracks are considered. At the root of all routing problem is the shortest path, the path through the network between a defined origin and destination that minimizes distance or travel time. Attributes such as length, travel speed, restrictions on travel direction and level of congestion are taken into account. A GIS can be very effective at solving routing problems because it is able to examine vast numbers of possible solutions quickly.

OPTIMUM PATHS: Here the concern is for finding optimum path across continuous space for linear facilities like highways, pipelines or even airline path etc. Again emphasis would be on shortest route may be to save fuel, time or avoiding the restrictions if there are any. These are normally sorted in raster, where each cell may be assigned a friction value, equal to the cost or time associated with moving across the cell in the horizontal or vertical directions.

Hypothesis Testing

Another kind of complex spatial analysis deals with the testing of hypotheses and drawing of inference and its relationship to GIS. It is about methods of inference drawn from information about a sample to a more general information for a larger population. Hypothesis testing is based on two concepts – confidence limits and inferential tests, which are basically statistical testing. The focus here is on the issue of using these approaches with geographic data in a GIS context.

HYPOTHESIS TESTS ON GEOGRAPHIC DATA: Although inferential tests are standard practice in much of science, they are very problematic for geographic data. In GIS, we analyze all the data that are there in a given area rather than sample. The example can be of sampling topographic elevation. The ability to estimate is the base of spatial interpolation. So here on one side banking on spatial interpolation, we can not believe in independence of geographic samples (basic assumption of statistical tests). Another important issue in this context is about the earth's surface which is heterogeneous, making it difficult to take samples that are truly representative for any large region. So what an investigator do, when inferential tests on geographical data are unacceptable, certainly investigator cannot discard spatial data. Here rather investigator may abandon inferential approach. The results obtained from the data are descriptive of the study area but it need not to be generalized. This approach, using local statistics observes the differences in the results of analysis over space. It represents a compromise between nomothetic and idiographic positions. Generalization is very tempting but the heterogeneous nature of the earth's surface makes it difficult. If generalization is necessary, then it can be accomplished by appropriate experimental design, replicating the study in a sufficient number of distinct areas to ensure confidence.

This page intentionally left blank

CHAPTER 10

SELECTION OF A GIS

Here we look in to the background necessary to make an intelligent GIS selection. There is quite a history to learn from, including some excellent accounts of spectacular failures, but also many examples of clear statements of how things went right. The philosophy here is that the educated consumer is the best GIS user, and an effective user soon becomes an advocate and sometimes a GIS evangelist. This chapter is not intended to tell as to which GIS to buy or use. Rather, it is hoped that, it will help in deciding this.

The Evolution of GIS Software

GIS software did not suddenly appear, rather there was a lengthy period leading up to the first real GISs during which the breed evolved rather rapidly. The intellectual ancestry included the creation of a spatial analysis tradition in geography, the quantitative revolution, and dramatic technological and conceptual improvements in the discipline of cartography. An early GIS landmark was an international survey of software conducted by the International Geographical Congress in 1979 (Marble, 1980). This survey had three volumes, one of which was entitled ***Complete Geographic Information Systems,*** although in fact few true GIS packages were represented. This volume was influential in deciding on the name 'GIS' because many alternatives were in use at that time. Just as important were the two volumes – ***Cartography and Graphics*** and ***Data Manipulation Programs.*** Together, these three volumes encapsulated the state of geographic data processing in the 1970s (Brassel, 1977). Most cartographic programs were single-purpose FORTRAN programs to do individual GIS operations such as digitizing, data format conversion, plotting on a specific hardware device such as a pen plotter, map projection transformations, or statistical analysis of data. None of these packages were integrated; a typical use would be to apply a series of one-at-a-time geographic operations to arrive at a final result or map.

Some of the early computer mapping systems had already devised many GIS functions by this time, however. Among these were ***SURFACE II*** by the Kansas Geological Survey, which could do point-to-grid conversions, interpolation, surface subtraction, and surface and contour mapping; ***CALFORM***, a package that could produce thematic maps; ***SYMAP***, a sophisticated analytical package from the Harvard Laboratory for Computer Graphics and Spatial Analysis that nevertheless ran only on mainframe computers and gave line-printer plots; and the Central Intelligence Agency's ***CAM***, which made plots from the World Data Bank outline maps with different map projections and features. By 1980 the first computer spreadsheet programs had arrived, led by the VisiCalc program, a very early microcomputer software 'killer app'. VisiCalc contained only a few of the capabilities of today's equivalent packages, yet for the first time gave the ability to store, manage, and manipulate numbers in a simple manner. Above all, data could be seen as active in a spreadsheet rather than as a static 'report' that consisted of a pile of computer printout. The links to statistical graphics, now common in packages such as ***SASGRAPH*** and ***Harvard Graphics,*** were a natural extension of this capability. The ancestry of GIS is completed by the first advances in database management systems. Early systems for database management were based on the less sophisticated data models of the hierarchical and related data models. A landmark was the beginning of the relational database managers in the early 1970s. Relational database managers quickly became the industry standard, first in the commercial world of records management and later in the microcomputer world.

The Early GIS Software Programs

By the late 1970s all of the necessary parts of a GIS existed as isolated software programs. The largest gap to be filled was between the relational database manager and the programs that dealt with plotting maps. The specific demands of hardware devices from particular manufacturers kept this as a constantly evolving field, with frequent rewrites and updates as systems and hardware changed. Later, the device independence attributable to common operating systems such as Unix and computer graphics programming standards such as **GKS, Core,** and **PHIGS** led to a narrowing of this chasm, to the point where today it remains as barely a discernible dip in the GIS ground. The scene was set for the arrival of the first true GISs. One of the earliest civilian systems to evolve all the capabilities of a true GIS was the **CGIS** (Canadian Geographical Information System), mostly because this system was the first to evolve from an inventory system toward doing analyses and then management. Essential to the emergence were the georeferencing and geocoding of the data, database management capability, a single integrated software package without separate, stand-alone elements, and a single user interface.

At first, GIS packages had unsophisticated user interfaces, and many actually made the user write short computer program-like scripts or to type highly structured formatted commands one at a time into the computer in response to prompts. As the GIS software evolved, the need for upward compatibility-that is, the need for existing users to be satisfied with a new version because things still work in much the same way as before-meant that many systems preserved elements of these older user interfaces long after they had been replaced by better tools. The second generation of GIS software included graphical user interfaces, usually involving the use of windows, icons, menus, and pointers. In the typical configuration today, the windows are standardized by the operating system and function in the same way that it does, 'inheriting' its characteristics. A first generation of GIS software used windows custom-built by the vendor. Later, after the broad distribution of windowing systems such as X-Windows and Microsoft Windows, the graphical user interface (GUI) tools that are part of the operating system became accessible to software designers and programmers. The typical system has pop-up, pull-down, and pull-right menus for selecting choices. Choices and locations are indicated with a mouse, although some systems use track balls or light pens. Similarly, the typical GIS can support multiple windows-for example, one for the database and one to display a map-and the tasks can be opened and closed as needed. While closed, they function in the background while they are graphically represented on the screen as an icon or small picture.

Operating Systems and GIS

Early GIS was heavily influenced by the types of operating systems in use. Early operating systems were quite unsophisticated but were used with GIS nevertheless. Among these were

IBM's mainframe operating systems, **MSDOS** by Microsoft, and DEC's **VMS**. These were rapidly replaced as the various GUI-based operating systems came into operation and as the microcomputer and workstation took over from the minicomputer and mainframe. In the microcomputer environment, the GUI-based operating systems include windows. The unified user interface, revolutionized by the Apple Macintosh's GUI and desktop metaphor, quickly took over as the dominant microcomputer operating environment, although others, such as IBM's **OS/2,** have remained popular also. These operating systems added two critical elements to the microcomputer's capabilities: multitasking (allowing many simultaneous work sessions) and device independence, meaning that plotters and printers could be taken out and assigned to the operating system instead of the GIS package, in somewhat the way that printing and screen fonts are handled centrally, rather than duplicated in every Windows package. One system that had encompassed these capabilities since its inception, and that swept the workstation environment, was Unix. Unix is a very small and efficient central operating system that is highly portable across computer systems. It has been the dominant workstation environment for two reasons: first, because it has complete integrated network support, and second, because several full GUIs exist for Unix in the public domain, the most important being the X-Windows system. X-Windows implementations of most leading GUIs exist, including OpenLook and the Open Software Foundation's MOTIF interface. In many Unix systems, the user can switch the GUI to suit particular needs or applications. As a final benefit, several versions of Unix and all of the GUI systems run extremely efficiently on microcomputers, including shareware Unix releases such as Linux, not only out performing the Windows-type GUIs, but being available free or as shareware on the Internet or from inexpensive suppliers on CD-ROM. A key element here has been the Free Software Foundation's releases.

Thus, two main avenues for GISs have evolved as far as operating systems are concerned. On the microcomputer platform a lingering set of DOS applications is rapidly being rewritten for the updated versions of Microsoft's Windows. In this GIS environment, the number of systems installed, the mobility of laptop and sub-notebook computing, and the low cost of software have been major strengths. On the workstation platform, Unix and X – Windows, often with MOTIF as the GUI, reign supreme. This work environment has led to high-end applications, large data sets, networking, depth of software, and high-quality graphics. Both are healthy and prospering workplaces for GIS.

GIS Functional Capabilities

A GIS is often defined not for what it is but for what it can do. This functional definition of GIS is very revealing about GIS use, because it shows us the set of capabilities that a GIS is expected to have. A minimal set of capabilities can be outlined and each GIS package held up to see whether it qualifies. A thorough examination of GIS capabilities is the critical step in how to select a GIS, because if the GIS do not match the requirements for a problem, no GIS solution will be forthcoming. In contrast, if the GIS have a large number of functions, the system may also need to be equally sophisticated or elaborate for efficient processing.

The functional capabilities can be grouped by the categories we have used earlier in this book, which are capabilities for data capture, data storage, data management, data retrieval, data analysis, and data display. These 'critical six' functions must always be present for the software to qualify as a GIS.

DATA CAPTURE: Data capture is getting the map into the computer. This is a critical first step in GIS. Geocoding must include at least the input of scanned or digitized maps in some appropriate format. The system should be able to absorb data in a variety of formats, not just in the native format of the particular GIS. For example, an outline map may be available as an Auto CAD DXF format file. The GIS should at a minimum be capable of absorbing the DXF file without further modification. Similarly, attributes may already be stored in standard database format (DBF) and should be absorbable either directly or through the generic ASCII format. Before a map can be digitized, however, it needs to be prepared. Different GIS packages handle the amount of preparation required in quite different ways. If the package supports scanning, the map needs to be clean, fold-free, free of handwritten annotation and marks, and on a stable base such as Mylar. If the map is digitized by hand it may need to be cut and spliced if the package does not support mosaicing, and control points with known locations and coordinates need to be marked for registering the map onto the digitizing tablet. Some GIS packages have extensive support for digitizing and sophisticated editing systems for detecting and eliminating digitizing errors. Others have few or none. Equally essential is to edit the maps after they have been captured. This requires the software to have an editing package or module of some kind. For a vector data set, at the minimum we should be able to delete and reenter a point or line. For a raster, we should be able to modify the grid by selecting subsets, changing the grid spacing, or changing a specific erroneous grid value. Other functions typical of an editor are node snapping, in which points that are close to each other and that should indeed be the same point, such as the endpoints of a line segment, are automatically placed into the graphic database with the identical coordinates; dissolve, when duplicate boundaries or unnecessary lines (*e.g.,* the digitized edges of adjacent category-type maps) are eliminated automatically or manually; and mosaicing or 'zipping,' in which adjacent map sheets scanned or digitized separately are merged into a seamless database without the unnecessary discontinuities caused by the lack of edge matching of the paper maps. For example, a major road that crosses two map sheets does not need to be represented as two separated features in the final GIS database. Another important editing function is the ability to deal with map generalization. Many digitizing modules of GIS systems, and certainly scanning, generate far more points than are necessary for the use of the GIS. This extra detail can complicate data reformatting and display, slow the analysis process, and lead to memory problems on the computer. Many GIS packages allow the user to select how much detail to retain in a feature. Most will retain points that have a minimum separation and snap together all points within a fuzzy tolerance. For point data sets, most GIS packages will eliminate or average duplicate points with the same coordinates. Some will allow line generalization, using anyone of many algorithms that reduce the number of points in a line. Common methods include extracting every nth point along the line (where n can be 2, 3, etc.), according to the amount of generalization required, and Douglas-Peucker point elimination, which uses a displacement orthogonal to the line to decide whether a point should be retained. Area features can be eliminated if they become too small, or can be grouped together, a process many GIS packages call clumping. It is also possible to generalize in the attributes, joining classes together.

To be useful, a GIS must provide tools above and beyond the editor to check the characteristics of the database. Checking the attributes is the responsibility of the database manager. The database system should enforce the restrictions on the GIS that are specified during the data definition phase of database construction and stored in the data dictionary. Most of this checking is done at data-entry time. It checks to determine that values fall within the correct type and range (a percentage numerical attribute, for example, should not contain a text string and should have a record of less than or equal to 100). More intricate and demanding are checks on the map data. Some GIS packages, which do not support topological structuring, do not enforce any restrictions on the map. Some simply check ranges; for example, every grid cell should have a data value between 0 and 255 in an image map. These systems run the risk of lacking a match between the attributes and the space they represent. No part of the map, for example, should fall into two separate areas-that is, the areas on a polygon map should not overlap or leave gaps. This happens when maps are captured at different scales or from inaccurate sources. Topological GIS systems can check automatically to ensure that the lines meet at nodes and that the entire map area is covered by polygons without gaps or overlaps. Beyond simply checking, many GIS packages allow automatic cleaning of topology, snapping nodes, eliminating duplicate lines, closing polygons, and eliminating slivers. Some systems simply point out the errors and ask the user to eliminate them with the editor. Some go ahead and make the corrections without user intervention. The GIS user should be careful when using automatic cleaning, for the tolerances may eliminate important small features or move the features around in geographic space without accountability.

A specific GIS package mayor may not be able to deal directly with GPS data conversion, with survey-type data from COGO (coordinate geometry) systems, or with remotely sensed imagery. Some GIS packages have both functions-that is, they serve as GIS and image processing systems. Among these are Idrisi, GRASS, and ERDAS. Essential to geocoding capabilities, because GIS allows maps from many sources to be brought into a common reference frame and to be overlaid, is the geocoding software's ability to move between coordinate systems and map projections. Most GIS packages accomplish this using affine transformation. Affine operations are plane geometry; they manipulate the coordinates themselves by scaling the axes, rotating the map, and moving the coordinate system's origin. In some cases, when no good control is available, maps must be statistically registered together, especially when one layer is a map and one an image or photograph. The statistical method known as rubber sheeting or warping is used for this and is a function inside many GIS packages.

DATA STORAGE: Data storage within a GIS has historically been an issue of both space-usually how much disk space the system requires-and access, or how flexible a GIS is in terms of making the data available for use. The massive reductions in the cost of disk storage, new high-density storage media such as the CD-ROM, and the integration of compression methods into common operating systems have made the former less critical and the latter more so. Current emphasis, therefore, is upon factors that improve data access. This has been a consequence also of the rise of distributed processing, the Internet, and the World Wide Web. As a result, many GIS packages are now capable of using metadata, or data about data, in an integrated manner. Metadata support might include a system for managing a single project as a separate entity, to managing many projects with multiple versions, to full support for exchangeable metadata stored in common formats and searchable through online 'clearinghouses'. Participation in the common library entails both standardizing the metadata to make it searchable and agreeing to make the data available either on or offline.

Other larger issues around GIS use, most essential to the degree of user friendliness of the system, concern the mechanism for user interaction with the software's functionality. Virtually all GIS software allows user interaction via command lines and/or windows within a GUI. The GUI interface is tedious, however, without some way of 'hatching' commands so that they can be executed either at another time, as a background task while the user gets on with another job, or for design-loop editing to change minor aspects of the process. Most systems, therefore, also contain a 'language' for the user to communicate with the system. This allows users to add their own custom functions, automate repetitive tasks, and add features to existing modules. These languages are usually command-line programs or macros, but they can also be enhancements of existing programming languages such as Basic and Smalltalk.

Although disk storage is less critical than in the past, it can still be a constraint. GIS software on a microcomputer can occupy tens of megabytes even without data, and on a workstation perhaps hundreds of megabytes. As data become higher resolution, as more raster layers are used, and as finer and finer detail becomes available, many GIS data sets can easily move into the gigabyte range in size. This implies that not only is supporting multiple resolutions important – for example, using coarse browse images as samples of the real thing – but also that data compression should be supported. This can vary all the way from partitioning data sets to meet constraints (such as a maximum number of polygons) to supporting compressed data formats and structures such as JPEG, run-length encoding, or quadtrees.

Also of great importance from a user perspective is the degree to which the system itself provides help to users, either via the operating system or as part of the software. Integration with online manuals, such as in Unix versions, support for context-sensitive hypertext help systems, such as the Windows help feature, and, ideally, an online interactive hypertext help system can be critical for the new user. These help systems can be used only when needed rather than encumbering the advanced user with unnecessary basic information. Support for data formats is important to a GIS when data are to be brought in from outside (*e.g.,* public-domain data from the Internet). Ideally, the GIS software should be able to read common data formats for both raster (DEM, GIF, TIFF, JPEG, Encapsulated PostScript) and vector (TIGER, HPGL, DXF, PostScript, DLG). Some GIS packages have import functions only into a single data structure, usually either an entity-by-entity structure or a topological structure. For three-dimensional data, these systems usually support only the triangular irregular network. Others support only raster structures based on the grid, including the quadtree, and either convert all data into this structure or just ignore it. A rather critical GIS function is the ability to convert between raster and vector data, an absolutely essential feature for the integration of multiple data sources such as GPS data and satellite images. In recent years interest has been in the development of GIS functions that support data in standard exchange formats. At the national and international levels, several data transfer standards have now been developed, such as the Spatial Data Transfer standard and DIGEST. As these standards become mandated, and as the role of data exchange increases, led by the Internet, most GIS systems will develop support for inputting and outputting data in these standard formats.

DATA MANAGEMENT: Much of the power of GIS software comes from the ability to manage not just map data but also attribute data. Every GIS is built around the software capabilities of a database management system (DBMS), a suite of software capable of storing, retrieving selectively, and reorganizing attribute information. The database manager allows us to think that all the data are available, that the data are structured in a simple flat-file format, and that they constitute a single entity. In fact, the database manager may have partitioned the data between files and memory locations and may have structured it in anyone of several formats and physical data models. A database manager is capable of many functions. Typically, a DBMS allows data entry, and data editing, and it supports tabular and other list types of output, sometimes independent of the GIS. Retrieval functions always include the ability to select certain attributes and records based on their values. For example, we can start with a database of India, and select out all records for states containing cities with over 1 million inhabitants, forming a new database that is wholly enclosed by the original and that duplicates part of it. We can also perform functions such as sorting data by value, and retrieving a selected record by its identification, such as a name or a number. Many operations on data are very important from a mapping perspective. For example, very often maps captured from different sheets must be merged together, or sometimes a mask must be placed over the data to exclude features entirely from the GIS. Examples of masks are restricted areas, water bodies, or military bases. Similarly, sometimes data must be assembled in one way, by topographic quadrangle, and then cookie cut into another region such as a state or a city boundary. Even more complex, sometimes line features such as the latitude/longitude grid, a river, or a political boundary must be sectioned up or have points added as new features or layers are introduced. This feature, called dynamic segmentation, can be done automatically by the GIS.

DATA RETRIEVAL: Another major area of GIS functionality is that of data retrieval. A GIS supports the retrieval of features by both their attributes and their spatial characteristics. All GIS systems allow users to retrieve data. Nevertheless, among systems some major differences exist between the type and sophistication of GIS functionality for data retrieval. The most basic act of data retrieval for a GIS is to show the position of a single feature.

This can be by retrieving coordinates as though they were attributes, or more commonly by displaying a feature in its spatial context on a map with respect to a grid or other features. For line features, the same goes, with the exception that line features have the attribute of length, and polygon features have the attribute of area. The GIS should be able to calculate and store these important basic properties as new attributes in the database. For example, for a set of districts we may want to take a polygon attribute such as an area of forest and divide it by the district area to make a percentage density of forest cover. Another common measurement we may want is to count features. For example, with the same database we could count the number of fire stations within the same districts by doing a point-in-polygon count from a separate database of municipal utilities and then relate the forest cover to the fire-prevention capabilities. A GIS has the critical capability of allowing the retrieval of features from the database using the map as the query vehicle. One way, indeed the most basic way, of doing this is to support the ability to point at a feature, using a device such as a mouse or a digitizer cursor, to see a list of attributes for that feature. Again, the ability to select by pointing to a location virtually defines a GIS. If it cannot do this, the system is probably a computer mapping system, not a GIS. Just as critical is the database manager select-by-attribute capability. This is normally a command to the database query language that generates a subset of the original data set. All GIS systems and all database managers support this capability.

GISs allow a set of retrieval operations based on using one or more map features as handles to select attributes of those features. Although some of them are very simple, these operations are also a real litmus test for establishing whether or not a software package is a GIS. A GIS should allow the user to select a feature by its proximity to a point, a line, or an area. For a point, this means selecting all features within a certain radius. For a line or a polygon, we have used the term buffering. Buffering allows the GIS users to retrieve features that lie within perhaps 1 kilometer of an address, within 1 kilometer of a river, or within 500 meters of a lake. Similarly, weighted buffering allows us to choose a non-uniform weighting of features within the buffer, favoring close-by instead of distant points, for example.

The next form of spatial retrieval is map overlay, when sets of irregular, non-overlapping regions are merged to form a new set of geographic regions that the two initial sets share. In the new attribute database it is possible to search by either set of units. A GIS should be able to perform overlay as a retrieval operation since to support the many spatial analyses based on map combination and weighted layer solutions. Vector systems usually compute a new set of polygons by adding points to and breaking up the existing sets, and in raster systems we allow map algebra, direct addition or multiplication of attributes stored in cells. Map overlay is an important part of a major GIS function, that of redistricting, in which new districts can be drawn and the data restructured into the regions so that tests and analyses can be performed by trial and error.

Another important set of retrieval options, especially in facilities mapping and hydrological systems, are those that allow networks to be constructed and queried. Typical networks are subway systems, pipes, power lines, and river systems. Retrieval operations involve searching for segments or nodes, adding or deleting nodes, redirecting flows, and routing. Not all GIS systems need these functions, but if the purpose is to manage a system usually abstracted as a network, such as a highway or rail system, a power supply system, or a service delivery system, obviously the GIS should then have this feature. Dana Tomlin (1990) has elegantly classified the operations that a raster GIS can perform into a structure called map algebra. In map algebra, the retrieval operations used are Boolean, multiply, recode, and algebra. Boolean operations are binary combinations. For example, we can take two maps, each divided into two attribute codes 'good' and 'bad' and find a binary AND solutions layer where both layers are 'good'. Multiply allows two layers to be multiplied together-for example, two sets of weights to be combined. In recode operations a range of computed attribute codes can be reorganized. An example is taking percentages and converting them to a binary layer by making all values greater than 70 % a '1' and all else a '0.' Map algebra allows compute operations, such as map-to-map multiplication for a binary AND over the space of a grid.

Two truly spatial retrieval operations are the ability to clump or aggregate areas, and to sift. For example, all areas of saturated soils surrounding swamps could be added to the swamps and recoded as wetlands, making a new, broader category of attribute. Sifting simply eliminates all areas that are too small, individual cells falling between two larger areas, or a tiny sliver polygon. Finally, some complex retrieval operations require the GIS to be able to compute numbers that describe shape. Common shape values are the length of the perimeter of a polygon squared, divided by its area, or the length of a line divided by the straight-line distance between the two endpoints.

DATA ANALYSIS: The analysis capabilities of GIS systems vary remarkably. Among the multitude of features that GIS systems offer are the computation of the slope and direction of slope (aspect) on a surface such as terrain; interpolation of missing or intermediate values; line-of-sight calculations on a surface; the incorporation of special break or skeleton lines into a surface; finding the optimal path through a network or a landscape; and the computations necessary to calculate the amount of material that must be moved during cut-and-fill operations such as road construction. Almost unique to GIS, and entirely absent in other types of information systems, are geometric tests. These can be absolutely fundamental to building a GIS in the first case. These are described by their dimensions, point-in -polygon, line-in-polygon, and point-to-line distance. The first, point-in-polygon, is how a point database such as a geocoded set of point samples is referenced into regions. Other more complex analytical operations include partitioning a surface into regions, perhaps using the locations of known points to form proximal regions, or by dividing a surface into automatically delineated drainage basins. Some of the most critical analytical operations are often the simplest. A GIS should be able to do spreadsheet and database tasks, compute a new attribute, generate a printed report or summarize a statistical description, and do at least simple statistical operations such as computing means and variance, performing significant testing, and plotting residuals.

DATA DISPLAY: GIS systems need to be able to perform what has become called desktop mapping, generating geographical and thematic maps so that they can be integrated with other functions. GISs typically can create several types of thematic mapping, including choropleth and proportional symbol maps; and they can draw isoline and cross-sectional diagrams when the data are three dimensional. Almost all GIS packages now either allow interactive modification of map elements-moving and resizing titles and legends-or allow their output to be exported into a package that has these capabilities, such as Adobe Illustrator or CorelDraw. A very limited few GIS packages include cartographic design help in their editing of graphics, defaulting to suitable colour schemes, or notifying the user if an inappropriate map type is being used for the data. This would be a desirable feature for many of the GISs on today's market and could avoid many tasteless or erroneous maps before they were created.

Data Structures and GIS Software

In the preceding discussion, the focus was on what functional capabilities the typical GIS offers. It should not be forgotten that many GIS features are predetermined by the GIS's particular data structure. At the very least the underlying data structure that the GIS uses, typically raster or vector but potentially also TIN, quadtree, or another model, such as object-based, determines what the GIS can and cannot do, how operations take place, and what level of error is involved. In general, the driving force for the choice of structure should be not only what type of system can be afforded, but more critically, what model is most suitable to a particular application, what retrieval and analysis functions will be used most, and what is the acceptable level of resolution and error. Some examples where particular structures are favoured include extensive land characterization applications such as land use/land cover study , where detailed data are not required (favours raster); applications involving irregular polygons and boundary lines, such as political units or census tracts (favours vector); applications that require the ability to register all features accurately to ground locations (favours vector); applications making extensive use of satellite or terrain data (favours raster); or applications where image processing functions and analyses such as slope and drainage analysis are to be conducted (favours raster). In many cases, the raster to vector conversion is done outside of the GIS in specialist conversion software, so that care can be taken to avoid the most common types of error, and so that the user can be brought in to resolve cases where the software is unable to solve a rasterization problem. Increasingly, many GIS systems allow the user to input and keep data in both raster and vector form. The GIS user should realize, however, that virtually all cross-structure retrieval and analysis requires one (or both) of the layers to change structure, and that this transformation often stamps itself irretrievably on the data's form, accuracy, and suitability for further use.

The Leading GIS Softwares

ARC GIS: ArcGIS, the latest version of Arc/Info, is a long-lived, full function GIS package that has been ported to the microcomputer, the workstation, and the mainframe. Arc/Info and ArcGIS are used to automate, manipulate, analyze, and display geographic data, and the software incorporates hundreds of sophisticated tools for map automation, data conversion, database management, map overlay and spatial analysis, interactive display and query, graphic editing, and address geocoding. The ArcInfo software includes a relational database interface for integration with commercial database management systems and a macro language called AML (ARC Macro Language) for developing customized applications. ArcGIS uses Visual Basic as its macro and programming language. ArcGIS uses a generic approach to geographic information systems that is not application specific, allowing the software to address virtually any geographic application. The software runs both on higher-end microcomputers and is available on several Unix workstations and for Windows NT, ArcGIS runs only on Windows NT. ESRI is broadly accepted as a market leader in GIS. Since its first release in 1999, it has a substantial modification of the program's user interface and functionality. Object-modelling capability and links to the Spatial Data Base Engine and other relational database management systems such as Oracle are included. With the latest versions of the software, the compatibility between ArcGIS and Arc View has been increased. The software uses the Windows COM component based software architecture, and is compatible with many other Window-based software tools.

ARC VIEW: ArcView is available for windows and a variety of Unix platforms. It is a desktop system for storing, querying, modifying, analyzing and displaying information about geographic space. An intuitive graphical user interface includes data display and a viewing tool. Support for spatial and tabular queries, 'hot links' to other desktop applications and data types, business graphics functions such as charting, bar and pie charts, and map symbolization, design, and layout capabilities are supported. Geo-coding and address matching are also possible. The Spatial Analyst tool kit makes working with raster data such as terrain and DEMs possible. Other extensions permit network analysis, allow Web activation of ArcView maps, and support advanced display features such as three-dimensional data visualization. ArcView GIS since version 8 has been more compatible with ArcGIS. ArcView is also a product of ESRI, which has developed ArcGIS. Compatibility exists between the two systems, with Arc View being more oriented toward map display than database management. Maps and data files are easily exchangeable between the formats used in the two systems, shape files, grid, images, and coverages.

MAPINFO: MapInfo was one of the first GIS programs to do desktop mapping. The vendor is MapInfo Corporation of Troy, New York. The software is well distributed and has many user groups and a broad variety of applications worldwide. The software runs under DOS, Windows, Macintosh, and on various Unix platfonns. MapInfo includes a link to the Basic programming language via a language called MapBasic. This development environment permits the creation of customized 'mapplications,' extending MapInfo's built-in functionality and allowing use of a common graphical interface. MapInfo has several GIS products aimed at different applications area, including MapInfo Professional, MapInfo MapX for programming GIS functionality, and specialist analytical modules such as MapXtreme for Web services, MapXSite for managing spatially enabled Web sites, and various database tools such as MapInfo Spatialware, Proviewer, and GIS Extension. MapInfo also supplies information products spanning geographic, economic, political, cultural, and industry application-specific content, each derived from leading worldwide sources to work the software. MapInfo also has an extensive training program, with classes at introductory and advanced levels for MapInfo and MapBasic.

GEOMEDIA: GeoMedia is a widely distributed layer-based GIS with a tradition in computer-assisted design by the Intergraph Corporation of Huntsville, Alabama. The software runs on workstations, PCs, and under the Windows NT system. An extensive set of add-on modules allow users to configure GIS capability around their specific needs. The set of modules includes GeoMedia, GeoMedia Professional, lntelliwhere Ondemand (for mobile systems), GeoMedia Webmap, and GeoMedia WebMap Professional. There are extensions aimed at applications in land information, parcel management, public works, and transportation. The layered implementation permits efficient storage structures for the geometry and linkages to relational database records. Geographic elements are represented in the GIS as features. Features are grouped into the same categories as the maps on which they appear. For the attribute data, GeoMedia incorporates use of the Oracle and SQL relational interface system, which facilitates client-server network communication to the relational DBMS so that multiple workstations communicate with the database server simultaneously. GeoMedia is fully integrated with Intergraph's traditional products, which include the MGE suite and tools for cartographic production. GeoMedia contains tools for building and maintaining topologically clean data without the processing and storage overhead of building and maintaining topology. In addition, it supports the open geodata interoperability specification and the spatial data transfer standard.

ILWIS: ILWIS is an acronym for the Integrated Land and Water Information System. It is a GIS software with image processing capabilities. ILWIS has been developed by the International Institute for Aerospace Survey an Earth Sciences (ITC), Enschede, The Netherlands. Since 1985, when it was released first, the software has undergone major improvements. It is a raster based software, and designed to be easy to use, yet provide professional-level GIS, image processing and spatial statistics analytical capability. It is intended to be affordable to all levels of users and to run on the most basic of common computer platforms As a GIS package, ILWIS, allows to input, manage, analyze and present geographic data. The newer version works in MS-Windows environment.

AUTODESK MAP: Autodesk Map is a GIS software suite built on the capabilities of the substantial AutoCAD software for automated drafting and design. Because this package is extensively used in planning, engineering, and architectural offices, many people can easily build upon their existing knowledge to enter the field of GIS. Autodesk Map uses AutoCAD 2002's drawing and plotting capabilities. Multiple data formats can be input, including those of AutoCAD (exchange format DXF and drawing format DWG) and also several other GIS packages. The software supports topology, query using Oracle and SQL, data management, and thematic mapping. The Autodesk Raster Design module supports grids and images and the Autodesk Onsite module handles all of the standard GIS data operations. There are extensive tools for coordinate conversion and specification, rubber-sheeting, and map editing and digitizing. The software uses the C^{++} programming language as a development tool. Output control and plotting support are strong, relying on AutoCAD's capability.

ERDAS: ERDAS (Earth Resource Data Analysis) is basically an image processing software but also has vector module and virtual GIS functions. It has different modules such as Imagine Advantage, Imagine Professional, Imagine Essentials and Virtual GIS etc. The GIS module of ERDAS, Imagine Vrtual GIS extends the powerful viewing and fast display of ERDAS Imagine with a range of superior 3D visual analysis capabilities, it also allows the creation of DEM, accurate terrain interpretations and rendering and adding of vector and image layers, symbols, annotation and 3D objects to create realistic views. The Imagine Vector module provides advantage of import and export of vector data, its cleaning and typology building. It also has ArcView extensions and ERDAS MapSheets module for gathering and analyzing geographic data.

GRASS: The U.S. Army Construction Engineering Research Laboratories (CERL) developed a public-domain software called the Geographic Resources Analysis Support System (GRASS). GRASS is raster based, was the first Unix GIS software, and has been considerably enhanced by the addition of user contributions – for example, in hydrologic modelling. The Web site states that GRASS is an open source, free software GIS with raster, topological vector, image processing, and graphics production functionality that operates on various platforms through a graphical user interface and shell in X-Windows. The source code for the program is available under the GNU General Public License. The software versions are available free over the Internet. Many users run GRASS on PCs under the Linux version of Unix, although a Windows port is now complete. Since 1985, CERL has released upgrades and enhancements to GRASS and provided technical user support. However, CERL terminated GRASS-related work in 1996. Public domain user support has been very strong, and highly international. Since 1996, the headquarters for GRASS support, research, and development has been at Baylor University, within the Department of Geology. The GRASS GIS uses a standardized command line input designed to resemble the Unix shell command language, but also uses a GUI under X-Windows. Unix compatibility allows users and programmers to create new applications and link GRASS to other software packages. Connections to the Unix shell and the C programing language allow simple extension and control.

IDRISI: The Idrisi GIS software has been developed, distributed, and supported on a not-for-profit basis by the Idrisi Project, Clark University Graduate School of Geography. To date, there are many thousands of registered users of Idrisi software worldwide, perhaps making it the most broadly used raster GIS in the world. Idrisi is designed to be easy to use, yet provide professional-level GIS, image processing and spatial statistics analytical capability on both DOS – and Windows – based personal computers. It is intended to be affordable to all levels of users and to run on the most basic of common computer platforms. Expensive graphics cards or peripheral devices are not required to make use of the analytical power of the system, which is designed with an open architecture so that researchers can integrate their own modules. Idrisi for Windows, first released in 1995, added a graphical user interface, flexible cartographic composition facilities, and an integrated database management system to the analytical tool kit. The more recent Idrisi32 is fully Windows and COM compliant and exploits object-oriented methods. Special routines for change and time – series analysis, spatial decision support, and uncertainty analysis and incorporation are included. A stand alone cartographic product, CartaLinx, allows topological editing and database development. Idrisi32 comes with a set of tutorial exercises and data that guide the new user through the concepts of GIS and image processing while also introducing the features of Idrisi.

GRAM++: GRAM++ is – (GeoReferenced Area Management) is a user-friendly GIS package developed indigenously by Center of Studies in Resource Engineering, IIT, Mumbai. GRAM is a modular DOS based integrated package designed for low cost computer configuration. The modules include standard spatial analysis, but again the software does not have much commercial application.

GEOSMART: It is developed by Department of Space, (DOS), Government of India and is exclusively used by Regional Remote Sensing Service Centers (RRSC). It is still in the process of development and is not used for commercial purposes and thus has very limited users.

MAPTITUDE: Maptitude is a GIS that works under the Windows operating system. The software is by Caliper Corporation, Newton, Massachusetts. Caliper has long been associated with the TransCAD and GIS-Plus GIS software packages. The latest version includes census data, a developer's toolkit and extended file support. The software comes with a considerable amount of geocoded and system-ready data on CD-ROM. The two CD-ROMs contain every street in the United States with the address information, state, county, zip codes, and census tracts as polygons with associated demographic data, and additional assorted U.S. and global data. Maptitude reads most standard PC file formats directly and can match each record against geographic data files using street address, zip code, and other features. Maptitude allows users to create and maintain geographic databases, analyze geographic relationships in data, and create highly professional map displays for presentations and reports. Maptitude runs under Windows NT, and with networks. The software uses the object linking and embedding of Windows, so that objects can be dragged and dropped into other applications.

This page intentionally left blank

Chapter 11

The Future of GIS

History has shown that how powerful GIS can be as a new mechanism for managing information. From humble origins, a set of simple ideas, and some rather inefficient software, GIS has grown into a sophisticated, a full-fledged giant industry in only few decades. GIS's dual role as a mainstream technology for the management of geographic information and as an effective tool for the use of resources is no longer a promise, but a reality.

Future Data

Easy Access to Digital Data: The blood of a GIS is the digital map data that runs through its software veins and hardware body. The future holds immense promise for new types of data, more complete data, higher-resolution data, and more timely data. Once the major obstacle to GIS development, data have now become GIS's greatest opportunity. Some of the types and sources of GIS data have already been described earlier in this book. The years ahead will bring us even more new types of data, and vast revisions of the existing types. As such, this summary of future data can be only a glimpse of what is still to come.

First, it should be stressed yet again that the entire mechanism for GIS data delivery has been revolutionized by the Internet and by the search tools built upon the structure of the World Wide Web. Most public-domain data, most shareware and freeware, and an increasingly large proportion of commercially produced GIS data use the Internet in place of computer tapes, diskettes, and the so-called sneaker-net *(i.e.,* hand delivery). This single trend has had, and will continue to have, the most impact on the field of GIS. Rarely does a new GIS project have to begin by digitizing or scanning geographic base maps. Instead, the majority of GIS work now involves bringing into the system a base layer of public-domain data and enriching it by capturing new layers pertinent to a particular GIS problem.

Remote Sensing and GIS: An additional increasingly high-resolution source of map data is that coming from aircraft and spacecraft in the form of remote sensing data. New spacecraft with the next generation of space instruments will provide an extremely rich set of both new and existing forms of data. Among the new programs are NASA's Earth Observation System (EOS), consisting of a huge variety of new instruments for mapping that will continue the NOAA polar orbiting programs and Landsat type data flows. NASA's Terra satellite, launched in 1999, has already begun to set the flow of Earth Science Enterprise data into the NASA databases. The IKONOS commercial satellite returns high-resolution data at about a 1-meter ground resolution. In addition, Landsat 7, also launched in 1999, has better spatial resolution as well as continuity with previous Landsat satellites. Our own IRS satellite provides a reasonably better spatial resolution images. The IRS 1A and 1B, where LISS 1 with 72.5 meters and LISS 2, 36.25 meters spatial resolution, but IRS 1 C and 1D, where Pan with 5.8 meters and LISS 3, with 23.5 meters spatial resolution stands far ahead among commercial remotely sensed data. Further, IRS P5 with 2.5 meters resolution and Cartosat with 1 meters spatial resolution have revolutionized the data quality. Several new commercial satellites, as well as a new generation of French SPOT satellites, will ensure that the diversity of instruments will increase. Similarly, the shuttle-carried radar mapping capabilities of SIR (shuttle imaging radar), as well as the Canadian RADARSAT, the

European Space Agency's ERS, and the Japanese JERS, all promise night-time and weather-invariant terrain mapping capabilities. The highly successful Shuttle Radar Topographic Mapping mission of Spring 2000 returned highly detailed topographic data and radar images for much of the world. Finally, the release of previously top-secret government spy satellite data, from the CORONA, LANYARD, and ARGON programs during the 1960s and 1970s, has allowed a significant amount of historical high-resolution imagery, much of it covering the United States, to be used for new mapping purposes. Evident after the release from the 'black' world of intelligence is the fact that this program and its successors have contributed significantly to the U.S. national mapping program, perhaps implying a higher degree of fidelity in these data than might have been imagined. As a historical record, these data are often able to show the 'before' image necessary to understand the 'after' of the present-day information.

Coupled with this plethora of new systems is a completely new infrastructure for data access, searching, and distribution. NASA's EOS program uses the EOSDIS, a program designed to make most satellite and other data, especially those of interest to scientists studying global change, available over the Internet. The USGS distributes land process data via the EROS Data Center in Sioux Falls, South Dakota, for EOS and many other programs, including the United Nations' GRID program. Even the CORONA data are distributed in this way, ensuring open and inexpensive public access to this map data. Landsat data are broadly available at reasonable costs, with up to l5-meter resolution. Most EOS data are also publicly available on the Internet.

The successful launch of Landsat 7 moved satellite data from the U.S. government back into the public domain, as a result, remotely sensed information finds its way back into the GIS mainstream, especially in the form of integrated GIS databases and GPS ground observations. Another major switch in policy will be the return to a continuous data stream. With commercialization, Landsat was moved over to a program that collected data only when a customer ordered it. As a result, much of the world remained uncovered, and searching back in time for data to show changes was impossible. Continuous coverage will allow far more images showing and contrasting changes, especially in the environment. The multi-agency Pathfinder program has attempted to demonstrate this capability, generating a U.S. coverage for three decades using historical multi-spectral scanner data. As a project of 'data mining' or searching existing data to extract products of value, another successful effort has been the AVHRR-based land-cover and vegetation index mapping conducted by scientists at the EROS data center and now released regularly on CD-ROM. In the future, this data set is planned for global coverage and periodic release, a massive boost to global-scale GIS use.

Clearly, remotely sensed data are highly structured around the raster data format. As much more data become available in this format, the demand upon software that converts between raster and vector data will increase, as will intelligent software for correcting lines and boundaries that come from pixel-based images. If this software becomes powerful and inexpensive, the possibility of having it work directly on the orbiting spacecraft becomes attractive, since the resulting vectors use far less data storage to either store or transmit to earth, allowing more efficient use of the orbit time. If the existing digital map could also be loaded, the spacecraft need only send back to earth revisions reflecting construction, natural changes, and so on. The prospects for automatic up-to-date maps seem bright.

GPS AS DATA SOURCE FOR GIS: Another critical step in data provision has been the ability, using the global positioning system (GPS), to go directly to the field to collect data rather than relying completely on maps. The GPS has also improved mapping significantly, because the geodetic control once only marginally available to mapping projects is now as easy as pushing a GPS receiver button and doing a differential correction to sub-meter accuracy. So precise is this new mechanism for data collection that existing GIS maps of cities, buildings, and other areas will have to be revisited for field verification. The ability to register a map quickly to a given map geometry (projection, ellipsoid, and datum) means that GIS layers can quickly and efficiently be brought into registration for overlay and comparative analysis. The field of GIS has greatly benefited, and the GIS-to-GPS link is now such that many GPS receivers and their data loggers can write data directly into GIS formats or include satellite images, air photos, or regular photographs directly in the field.

The flexibility of this system, when integrated with in-vehicle navigation systems that also use inertial navigation and stored digital street maps, has evolved a technology that is becoming standard equipment in public and private vehicles. The drivers in future may never again have to stop to ask the way to a destination. Now moving into large-scale production, these systems have already been incorporated into a car's dashboard. The rapid generation of street, highway, and city maps resulting from the growth of these systems-data that are by definition of great locational accuracy is greatly benefiting GIS. Although the data have so far been digitized almost exclusively by private companies, competition has led to a data price war in recent years, and costs have fallen remarkably. Hand-held receivers with map displays can now be purchased for much smaller amount than what need to be paid, a few years ago. These easily available data are now being used for variety of purposes like hunting, travel, and driving etc. GPS has also found use in fleet vehicles such as the trucking and moving industries, and in the delivery business. In each case, the common element is the need for moving around a street network efficiently.

IMAGE MAPS AND GIS: Another significant new data source now exists owing to the arrival of **digital orthophotoquads.** Digital orthophotoquads are geometrically corrected air photos with some cartographic annotation. Their historical use has been as sources of information for the U.S. Department of Agriculture. Recently, however, these data have been made available by the USGS on CD-ROMs in digital format in quarter-quadrangles; that is, one-fourth of a 1: 24,000 7.5-minute quadrangle as one data set, with an equivalent scale of 1: 12,000 and a ground resolution of 1 meter. Rather than being vector data, though, the raster nature of this layer and the fact that it is monochrome have resulted in its use as a background image for GIS, over which field and existing geocoded data are assembled. The primary function of the orthophoto will be to assure the same type of layer-to-layer registration discussed in the case of GPS above. Over the next few years, the coverage will expand to cover the entire United States, and a 10-year revisit will assure that city and other maps can be updated as required. In addition, new raster images of the entire United States, digital raster graphics, have also been made available in CD-ROM format.

The digital raster graphic (DRG) is a scanned image of a U.S. Geological Survey (USGS) topographic map, including all the information on the map edge, or 'collar.' The image inside the map neat line is georeferenced to the surface of the earth. These maps make excellent starting points for GIS projects, and they often contain many features that can be extracted for use, such as contour lines and building footprints.

DATA EXCHANGE AND GIS: The final prospect for GIS data is the one of exchange. As GIS becomes more widespread, the various map-generating and map-using communities will need to trade data more than ever before. Already, nautical charts and world maps have needed to be standardized, edge-matched, and cross-checked across national and even continental borders. This implies that there is a need to build a formal structure for data exchanges, and the several new standards for data transfer have already had a major impact on this issue. Standard transfer formats mean, for example, that a ship sailing into foreign national waters can download the latest navigation chart for immediate use.

Many sets of standards for data have emerged. Internationally, NATO has produced the DIGEST standard, the International Hydrographic Organization has produced the DX-90 standard, and other nations have established their own data transfer standards. Industries such as television, computer software, and communications have seen standards take on a critical role-and even critically influence technologies, such as videotape formats. Standards will have a great impact on the future of GIS. With formal, explicitly defined formats for features, open exchange will be easy and data will no longer be a constraint to GIS use. As the world becomes more and more a single global market, early elimination of the data transfer barrier will assure the future of GIS for many years to come. In United States, map data have evolved the Spatial Data Transfer Standard (SDTS), now formalized as the FIPS 173 (a FIPS is a Federal Information Processing Standard). The year 2000 Census in the United States was the first full-scale mapping effort to generate all of its digital map data in the SDTS format. Other agencies, such as the USGS, have already converted many files, such as the Digital Line Graphs, into FIPS 173 format and structure.

A critical element of data exchange is simply finding out who has data that already exist about a geographic area. Those who have GIS data may be willing to share not necessarily all the data but at least the metadata that give information about data coverage, accuracy, timeliness, and availability. Standards have now been developed by the U.S. Federal Geographic Data Committee that specify how data can be indexed for effective search. Prototype systems for coding data, and Web and other computer-based tools for searching and browsing for data, have given rise to the concept of a digital map library. Such a library allows searching, and then allows the user to access a public or other Web location that can provide data for downloading. As data become more and available, these metadata systems will become increasingly useful for sorting through the huge quantity of available digital map data.

LOCATION-BASED SERVICES AND GIS: Location-based services (LBSs) are computer-based services that exploit information about where a user is located in geographic space. Location-based services take advantage of GPS, but also may rely on E911, an initiative of the Federal Communications Commission that requires wireless telephone carriers to pinpoint a caller's telephone number to emergency dispatchers. This may use the location of the telephone itself with respect to the nearest cellular transmitters, solved by signal triangulation. E91l is the most widely used location-based service in the United States, although manufacturers of cellular telephones are also incorporating GPS chips into new cellular telephones. The power of LES means that the Internet can also be made location oriented. Many such Web-based services already exist, often using map providers like MapQuest to provide maps and directions along with the geographic search capacity.

Users of LBS so far seem to be either vehicle-based, where the GPS and computer are in the car and used to query geographically ordered information, or mobile. Mobile users are usually either working on a personal digital assistant that contains a cellular phone connection to the Internet and a GPS card (which is often an add-on feature, and comes on a PCMCIA or other card), or they are using the fairly limited interactive communications capabilities of a cellular telephone. Early uses of the systems have included automotive roadside assistance, emergency and collision notification, stolen vehicle tracking, on-demand navigation assistance, traffic alerts, and vehicle diagnostics. Broadly, LBS uses selected subsets of GIS functionality, but delivers them to the user on demand. Most applications are in navigation route finding, and space constrained search. One unresolved issue with LBS is how 'open' the geographic information will be because the privacy issues and possible abuse of information is of great concern and obvious.

Future Hardware

Hardware for GIS has gone through at least four revolutions in the last decade: the workstation, network, microcomputer, and mobility revolutions. Each one of these has already had a profound impact on computer hardware and will influence the future of GIS significantly.

THE WORKSTATION REVOLUTION: The first of these-the workstation revolution-has given GIS an operating platform that has all of the necessary power and storage to work with massive databases. In the space of just a few years, the capability of a 515,000 workstation has gone from megabytes to gigabytes of storage, while increasing the size of RAM beyond 64 megabytes and the processor speed well above and beyond the capabilities of most mainframe computers. Along with the expansion of the workstation has been the spread of Unix, the TCP/IP communications protocol, and graphical user interfaces such as Sun's OpenLook, Motif, and MIT's X-Windows. The more powerful systems of the future and the falling price of workstations seem to make this the preferred GIS work environment for large-scale projects, although Windows, Macintoshes, Linux, and even DOS remain for low-end systems, small projects, and for education.

THE NETWORK REVOLUTION: The network capabilities built into workstations have broadened to include many other types of computers, including microcomputers. Many computers are now connected to the Internet and can use network search tools such as Windows Explorer and Netscape to 'surf' the World Wide Web (WWW). Already, the Internet has become a primary means for data exchange and information search and retrieval. Many GIS packages, including Arc/Info, GRASS, and IDRISI, have support services on the Internet's network conference groups. The national spatial data infrastructure, a linked distributed database of public GIS information with common metadata, is being built upon the capabilities of the Internet and the WWW.

Many commercial GISs have now developed modules that allow entire GISs to be Web-enabled, including ESRI's Internet Map Server, MapInfo's MapXtreme, and Intergraph's GeoMedia WebMap. This means that the GIS can be searched, queried, or analyzed over the Web and the results displayed locally on a client using software tools such as Java and a standard browser. GIS is behind many of the map display tools now proliferating on the Internet, including the Web serving of public information in many communities and cities around the country. Full GIS functionality is rarely delivered over the Internet, and these systems usually feature simplified user interfaces and simplified data searching and map construction. If complete functionality were deliverable, then the GIS user need not 'own' the GIS software, or even the data, and could simply pay for their use over the network when desired. Some Web-based educational systems already use this approach, such as ESRI's virtual campus.

THE MICROCOMPUTER REVOLUTION: The microcomputer has matured and increased in power significantly, making this platform widely distributed, relatively inexpensive (especially when compared with the other components of a GIS), and easily capable of running many GIS packages. Here, the Intel Pentium chip, the CD-ROM drive, and simple graphical user interfaces such as Microsoft Windows, Linux, and others have led the way. While the first and even the second generations of microcomputers were at best only modestly suitable for GIS applications, present-day systems have crossed the size and power threshold and become useful professional and educational GIS platforms. The implication of this revolution has been largely one of broad distribution-GIS can now go almost anywhere a microcomputer can go.

THE MOBILITY REVOLUTION: The fourth major technological revolution represented by microcomputers has been the trend toward mobility. Here, driving forces have been the laptop, portable, subportable, and even palm-top computer; the PCMCIA and USB interface allowing easily transferable data storage and interoperability of devices; and the mobile communications and GPS technology that now accompany them. However, the linkage that allows a GPS unit to compute a position, download it to a portable computer, receive by modem and mobile phone differential corrections to the GPS location, and then write these data directly into a GIS format, and to do all this so simultaneously that the points appear as if by magic in real time on the portable computer's GIS map display, was literally beyond belief only a few years ago.

Some GIS vendors now offer limited versions of their GIS for use on highly portable devices, such as the Palm Pilot and Compaq iPAQ. Among these are MapInfo and ESRI, with the ArcPad software. When these devices are coupled with a GPS card, often available as a plug-in on a PCMCIA card, they become completely mobile GIS systems in their own right.

Added to the continued miniaturization of computer and communications equipment, personal mobility of GIS hardware has reached and gone far smaller than the field portable minimum level. Along with these new capabilities come the terms ubiquitous computing (go anywhere, remain connected to the Internet via cellular telephone) and augmented reality, in which the GIS data view can be superimposed on the 'real' view by direct entry into the human vision field. These are prototypes now, but are apparently already in use in some professions.

The Impact Of The Revolutions

Extending these concepts into the future gives us the following four observations. First, the workstation and all its characteristics will continue to dominate the GIS workplace as the primary tool for advanced applications, but will become immensely more powerful. This will entail more local disk, perhaps workstations capable of terabytes of storage locally, and more distributed and shared data resources, with file servers acting as the data libraries or depositories for GIS projects.

Similarly, as the amount of random access memory (RAM) available approaches the gigabyte range, many processes now performed as input/output or file manipulation operations will be possible to do inside the workstation RAM in real time, making even computationally complex and sophisticated operations very fast, perhaps interactive, and certainly fast enough to allow use of the new techniques of scientific visualization. The dominance of the Unix/Motif/OpenLook/X-Windows environment looks certain, as does a shift toward programming GIS in new systems, languages, and environments. The move toward visual programming tools, object-oriented programming, expert systems, and so on has already started to deliver new and more user-friendly GIS systems; at the same time, the high-end systems are likely to acquire new and even more powerful capabilities. The new computing method most likely to have a major impact on GIS is the move toward parallel processing, which, once in place within the high-end workstations (and already in effect today), will allow real-time processing of imagery in new ways, promising immense speed-up in processing.

The role of the network is another simple extension of today's environment. Already we have prototype systems in place of future systems. NASA's EOSDIS, the National Spatial Data Clearinghouse, and the entire WWW are testimonials to the rapid growth, acceptance, and exploitation of the Internet as the primary future tool for the searching, distribution, and distributed storage of spatial data. Yet the Internet can deliver far more than data and metadata (data about data). It can deliver information, advice, and assistance, often tailored to a specific environment or GIS package.

The Internet can offer formal means for the dissemination of ideas and research, much as today we depend on the printed page in books and journals. It can also remove the GIS analyst almost entirely from the traditional workplace. Also of significance is the fact that the Internet can deliver shareware, meaning that the new user can experiment with a free or inexpensive GIS before making a purchasing decision. Finally, and most important to the academic world, the Internet can deliver both real-time and programmed university education, in the form of multimedia and hypertext 'virtual' classrooms free from the restraints of national boundaries and geographic separation. Again, the democratization of the GIS field offers some exciting prospects for a future information-based economy.

Both the power and the increased flexibility of the microcomputer have been pivotal not in increasing the power of GIS applications, for this has been the domain of the workstation, but in penetrating new fields of GIS application and in the domain of GIS education. New fields to GIS are archaeology, forestry, epidemiology, emergency management, real estate, marketing, and a host of others. In every instance, the first steps in these areas were taken by new users in a microcomputer environment.

Obviously, improvements in microcomputer user-friendliness have been critical; especially the move to Windows-based graphical user interfaces (GUls). The acceptance of GISs, which are necessarily complex and often counterintuitive to the newly initiated, has really dated only from the widespread use of these windows-based GUIs. In addition, the movement away from smaller hard disks to CD-ROMs, PCMCIA cards, and tape backup has helped. Another important step has been the large price decreases in devices for basic graphic input, including small digitizing tablets and scanners, and for output, such as colour printers and pen plotters.

Education has benefited significantly from the low cost of hardware, because the budget for hardware in colleges, universities, and schools is usually small and under constant threat. The trend toward the microcomputer classroom with a networked server running shared software licenses is broad enough that this configuration is now common in many high schools, and even there GIS has entered the curriculum in some places. As geography moves back into the curriculum in high schools, GIS will lead the way, bringing forth a new generation of GIS-literate students for the information economy.

Increased mobility has also generated many new GIS uses. Here, however, it is the coalescence of mobile technologies, communications, navigation, and data processing that has been pivotal. Obviously, the exciting new data capture prospects of GPS have been very important; however, the migration of software and hardware for image processing and remote sensing into the mobile environment offers many exciting prospects.

FUTURE PROSPECTS OF HARDWARE

Finally, some of the trends on the edges of computer science and engineering have real prospects for GIS application. Among these are stereo and head-mounted displays; input and output devices that are worn; parallel and self-maintaining fault-tolerant computers, and above all, mass storage and computing power much greater and faster than that available today.

A vision of a future GIS system might be a pocket-held integrated GIS, GPS, and image-processing computer capable of real-time mapping on a display worn as a pair of stereo sunglasses. Data capture would consist of walking around and looking at objects, and speaking their names and attributes into an expert-system-based interpreter that encodes and structures the data and transmits them immediately to a central network accessible storage location. This implies that a single person, or even an unmanned vehicle or pilot-less aircraft, could move around gathering data while any interested person displays and analyzes the information in real time in his or her office or home.

Another future prospect is that of the data analyst becoming a data explorer, delving into three-dimensional realistic visualizations of the data, seeking out patterns and structure instead of the user of the simple statistical analysis of today. The human mind is capable of some amazing parallel processing of its own and can easily seek out structures that computers miss. Similarly, the same systems could manage the very systems they support, perhaps allowing for integrated modelling and prediction of future 'what-if' scenarios. Regardless of the actual hardware used, there is little doubt that the tools and devices required for GIS work will become commonplace in the very near future: perhaps never cheap enough to come free with a fill-up at the local gas station, but undoubtedly cheap enough that the likelihood of GIS hardware being a limiting factor in the GIS future is minimal.

Future Software

A review of GIS software trends of recent years is in order if we are to speculate in a similar way about where GIS software is going in the future. Several themes suggest themselves.

SOFTWARE TRENDS: The first major trend over the last few years has been in operating systems. In the 1970s, complex mainframe operating systems predominated, and system interaction was limited both by the inflexibility of the user interface and by the nature of the early time sharing of systems. The first minicomputer operating systems were little better, with the exception of Unix, a simple and much abbreviated set of instructions for doing file and systems management that has proven very flexible and long lived. Today, operating systems can 'multitask', working on two problems at once, with ease. Microcomputer operating systems now also have this capability.

Early systems were somewhat poor at user interaction, yet the revolutionary Apple Macintosh system, followed by the various flavors of Windows and X-Windows, led to a significant improvement in user simplicity and comprehensiveness. Standardization was an additional unseen improvement: that is, every application could use a standard and commonly understood set of menus instead of making its own flavor. Most recently, operating systems that run on multiple platforms have flourished, including UNIX. The ability to divorce standard operations such as printing and digitizer communication from the GIS led to some major improvements. Similarly, commonly accepted industry standard formats and languages, such as PostScript, led to another level of standardization, this time for hardware devices such as printers and plotters.

THE USER INTERFACE AND WIMPs: The computer era has seen radical changes in the very nature of both the computer and GIS user interfaces. Early systems used only the screen and the keyboard to communicate to the user. Systems now have these same functions, but also a mouse, pointing devices such as a track ball or light pen, multiple windows on the screen, sound, animation, and many other options. Most significant has been the rise of the ***WIMP*** (windows, icons, menus, and pointers) interface. Windows are multiple simultaneous screens on a single display, usually serving different tasks and fully under user control. When inactive, windows can be closed and kept visible as icons, or icons can be attached to tasks and used to activate them-programs, for example. Menus can take a variety of forms. Many user interfaces place a set of menus along a bar at the top of the screen, controlling more and more specific tasks as one goes from left to right. Menus are often 'nested,' that is, a selection reveals another menu level and even more selections. Menus can 'pop up' from a space or window, or can be 'pulled' from other menus of messages. Pointers are devices for communicating location on the screen and in windows, and they most commonly take the form of a mouse or a track ball.

Central to the GUIs of recent years has been a metaphor. The metaphor most commonly used has been the desktop; that is, the screen of the computer is designed to resemble the top of a desk, and the icons and other elements are allowed to rest on it, awaiting use. Some operating systems have gone beyond the constraints of this suite of interactions, and many operating systems now allow input from voice, touch screen, and even direct input from GPS receivers and other recording devices, such as digital cameras and videocams.

The map itself is a useful metaphor, and a future GIS can easily be imagined in which the map and its elements, such as the scale and the legend, are used to manage and manipulate the data. This is already what a GIS does, but the user-interactive element would be a new addition to the system. Several systems already use icons as elements of a process or transformation model to track sequences of operations. This is clearly a taste of the future of GIS. The GUIs will probably allow the user to specify tasks independent of the data, in the abstract. Possible alternative metaphors are the English language, a symbolic language such as Dana Tomlin's Map Algebra, or pictorial languages. As most GIS operations contain maps, they have been used as a metaphor. It is highly likely that the next generation of GIS will incorporate some or all of these features, making them considerably easier to use.

THE RASTER VERSUS VECTOR DEBATE: Another major software trend has been a massive change in the distinctions between systems based on their data structures. As we have seen in earlier chapters, quite often the process of geocoding, or sometimes a particular GIS process such as map overlay, leaves an 'imprint' on the data that remains as one of the restrictions on data use and flexibility. The last few years have seen almost every GIS package become capable of supporting both raster and vector data structures, and in some cases many others besides. This has become the sort of single super-flexible data structure that many sought to develop in the early days of GIS research. Instead of one structure winning out, GIS developers have realized that each structure has its strengths and weaknesses, in particular for analytical operations. Systems can take advantage of the strengths of a particular structure for a particular operation-map overlay or edge detection, for example. The disadvantage is that the transformation between data structures often entails significant error in and of itself and can lead to some serious problems in GIS analysis. Nevertheless, if done carefully, the raster/vector dichotomy can be eliminated.

In the future, GIS software is likely to have incorporated the strengths of the various structures and should be capable of intelligently converting data between structures without the intervention of the GIS user. This means that some of the principles of what is happening may be ‘hidden’ from the user. Self-configuring GIS software does not seem too far-fetched. In addition, the spatial data transfer standard has allowed data to be encoded along with the necessary information to move easily between structures. A GIS could simply read a standard file header, establish just what is stored in the file, and then reconfigure the data as necessary for whatever the user demands. In time, also, an intelligent GIS could learn about the demands of the GIS’s own user and hold data in suitable structures for the most commonly performed operations and analyses.

OBJECT – ORIENTED GIS: Another major development in the software world has been languages, and now databases, that support ‘objects,’ called object-oriented systems. Geographic features map very closely onto objects. Object-oriented programming systems (OOPSs) allow the definition of standard ‘classes’ that contain all the properties of an object. As a simple example, an object class could be a point containing the latitude and longitude of the point, a feature code for the point such as ‘tube well,’ and any necessary text describing the object. If we wish to create another point feature, this can be done simply by cloning the original with all its class information, a process called inheritance.

In addition, we can encode the fact that points often have data conversion or analysis constraints. For example, the centroid of a set of points is itself a point and can inherit a point’s properties. This approach has allowed the development of entire GIS packages, and is seen as a way of building far more intelligent GIS systems in the future. While the OOPS is not the tool for all GIS operations or systems, it is indeed a powerful way of modelling data and will influence the future of GIS software significantly.

DISTRIBUTED DATABASES: A major transition within the GIS industry has been the movement toward distributed databases. This has happened at two levels, first within a local area network; data and software have migrated from individual hard disks to file servers, computers dedicated solely to disk storage and moving information over the local network to the client workstations or sometimes microcomputers. This is a direct equivalent of the transformations made possible when the availability of printed books was revolutionized by the advent of public libraries. Library users need not worry about getting the latest information, specifics of book ordering from publishers, and so on. They can use the library as an information delivery service. The price to the user is a security system of some kind and the loss of ‘ownership’ of the data or software on the server. The ability of computers to make almost unlimited immediate copies of files without loss from the original source has changed the library model somewhat. Quite clearly, though, a distributed data system can lead to a large-scale reduction in storage duplication. Second, connection to the Internet has made it possible to have distributed databases on a massive scale, across national boundaries and even across major hardware and software barriers. Thus it is possible to let the organizations maintain a library of data and to download the data sets of interest only when they are needed. This arrangement is ideal but leans heavily on the ability to locate and transfer data on demand. Various network search tools such as WAIS, Netscape, Archie, Gopher, and Mosaic have made this metadata accessibility possible, leading to some major breakthroughs in Internet-wide distributed databases. Threats to this situation would be privatization of the Internet, implementation of a pay-per-use system for data retrieval, or taking public data out of this broadly accessible distribution system.

The Internet supplies far more to GIS users than data. It delivers software, research papers, advice, shared knowledge, and the routine contact necessary for efficient operation of a GIS. Increasingly, GIS companies and shareware services are using the Internet as the primary means by which support is delivered. A GIS user can send e-mail questions to an expert anywhere in the world. Use of the File Transfer Protocol (FTP) in 'anonymous' mode allows downloading of software fixes (called bridges and patches), and even some tailor-made debugging and testing. Remote log-ins allow an expert to get onto a sick computer and cure software ills without leaving the office. In the future, this sort of service may grow to become the major means of GIS software user support.

As GIS systems have grown, so has that part of the GIS industry that acts as a supplier of data. Many companies work to update, enhance, or correct all sorts of existing data and many also generate new data from scratch. These services have acted to provide data in a broad variety of common GIS formats and offer subscription services for regular data updates, after a new release. New GIS projects especially often require digitizing and scanning even before basic operations can begin. The data services conduct turnkey operations, handing over to the GIS staff a complete data set for use. As the costs of data supply fall and the distribution mechanisms such as CD-ROM become more widespread, the cost of GIS data is likely to plummet. This GIS data price war should result in very low cost data in the future, at least for basic cartographic data. This is exactly the model that has been followed for paper maps. The data services will turn increasingly to custom services and data enhancement as a means to survive and prosper.

GIS USER NEEDS: Another issue of interest to the future of GIS is how the industry will continue to develop. Obviously, GIS users have broadened into two types: the large organization wide projects with huge databases and often specific missions; and the small, usually one-person operations run by a jack-of-all-trades. Although GIS can serve both sets of users, the specifics of hardware, software, and the computing environment mean that different GISs suit each world.

At the organizational level, labour can be divided. But for an individual level, all tasks are the responsibility of one individual. Small users will probably not be able to add significant amounts of new data, with the exception of field data collection with GPS. They will be more reliant on public-domain data, and the data will probably be less up to date and at a coarser scale. It is at this level that the GIS use is closest to the domain expertise. Getting the GIS as far into the field as possible is often a key to the success of a system. Field operatives can use the GIS quickly to make ordinary but informed decisions about the use of resources on a day-to-day basis; that is, where the payoff is greatest. Sophisticated analytical operations may not even be necessary at this level, and using the GIS as a graphic inventory and map production system is more than sufficient for success. Large systems, by contrast, can maintain up-to-date and detailed information, and can use it in its full GIS context, performing the roles of inventory, analysis, decision making, and management. Here, also, better information means better use of resources. Clearly, the GIS industry must continue to exploit both types of environment. Often, this means taking large systems and packaging them small, or taking lessons learned by advanced users and translating them for the general user.

Finally, the GIS users themselves have become a sort of self-help facility. Most major software packages or regional-interest organizations using GIS have user groups, often with special conferences, workshops, newsletters, and Internet discussion groups. This is an excellent grass-roots level for GIS to flourish, one that GIS vendors have discovered. As GIS packages become more complex but also more user friendly, these user groups will converge on some common principles for GIS use. These principles should be, and are, shared with all users. Often, a good idea in one software environment can lead to productive duplication in another.

GIS SOFTWARE RESEARCH: Some of the future expectations for GIS software are the results of research now under way, and as such are also somewhat predictable. For example, for some time, scholars in GIS have been interested in the impact on GIS of supporting geographic and attribute data from many time periods. Obviously, the digital map in a GIS is 'time stamped' at the time the data were created. In the real world, however, data become out of date and must be revised, or new data sets are released to replace the old. Some data have very short duration such as – weather forecasts, and revision and update quickly become a major part of the GIS maintenance. In most cases, GIS data are simply given an additional attribute of the date the data were created, even though often the date of the data and the date of entry into the GIS are not always the same. The implications on the design of the GIS to facilitate use, automatic update, for instance, or automatic selection of the most up-to-date version of every feature are now being integrated into the GIS's functions.

Another trend that today fills research journals is use of the more recent object-oriented programming systems and database managers as the tools with which to construct GISs. The advantage of object-oriented systems is that the features within the GIS can be described in advance, categorized by types, and that actual data represent an 'instance' of one of these types or 'classes' of object. This advance knowledge of types allows operations and algorithms to be stored with the objects. The objects become a 'hidden layer' for which the user need not perform many of the operations one performs as routine in a regular GIS. For example, an object-oriented GIS can know in advance the steps necessary for, and outcomes of, map overlay along with any data conversions necessary for its performance. Disadvantages of object-oriented systems are that they are often memory and computationally intensive and that their sophistication is unnecessary for most of the basic GIS operations.

Some GIS research has focused on the user interface with GIS. Most GIS systems have evolved from command-line and macro control to an interactive menu system. As the GUI improves, GISs can improve, too. One suggested improvement has been to incorporate natural language interfaces, in which the user communicates in English with the system. Others have suggested that GISs incorporate the 'fuzzy' characteristics of English as well. Such a system could be asked to show a buffer containing features that are 'near' something rather than within 5 kilometers of it.

More advanced user interfaces could be icon driven, as in the Geographer's Desktop, and could use a symbolic manipulation language such as the Idrisi, ERDAS Imagine, ER-Mapper, Khoros and Stella modelling systems, in which the user plans out operations by drawing a highly stylized flow diagram and then makes the process operational to carry out the tasks. Even more sophisticated interfaces are obviously possible, and we have yet to even start work on effective use of interfaces for multimedia, interactive, and animated GIS systems.

GIS INTEROPERABILITY: Another area of concentrated GIS research is that of interoperability. An effort is currently under way to standardize and publish a set of specifications for GIS functions and capabilities, allowing a standard language and a higher degree of mobility among systems. This effort, termed OPEN/GIS, is an attempt to repeat the success that an open description of the user interface had for GUIs, an effort known as Open/Systems, which gave us OpenLook and Motif. Such a specification, when openly published, allows vendors to develop products along a common line and toward common goals while maintaining the individuality of their own software package. The payback from this effort will be that GIS software will run in a manner that is totally unaltered from the user's perspective, on virtually any computer and under any operating system.

The last, and a major trend as far as interoperability is concerned, is the arrival of the standards for spatial data, the spatial data transfer standard (FIPS 173). This standard means that data that comply with the standard will be able to move directly into a GIS with all the stored characteristics, topology, attributes, and graphics fully intact. For the first time, identical data sets can migrate between GIS software packages without losing the resolution, accuracy, or descriptive poignancy necessary for rigorous GIS analysis. This effort is already close to realization, and most GIS vendors have declared their intent to support the standard in the very near future.

As GIS becomes mainstream, the GIS package will become yet another basic requirement of using a computer, and the software will become so ubiquitous that it will be available either bundled on a computer on purchase, or shrink-wrapped at the local computer store. When GIS reaches this stage and when even the advanced, let alone the basic, GIS operations become standard operating parts of decision making on a daily basis, GIS will be a part of every person's life, known or unknown. A GIS will affect how we live, travel, communicate, manage our finances, work.

Future Issues and Problems

Assuming that GIS is now only a few years away from this degree of permeation into the economy, we also need to look at the issues and problems we are likely to face with the future of GIS. How well we as a user community react to the challenges of the issues will play a major role in the future of GIS.

PRIVACY: An issue that raises itself again and again as GIS databases become more and more widespread is that of personal privacy. We very often take our right to privacy for granted, yet all the time, by the use of telephones, credit cards, mail order, and the like, we are constantly revealing to other people what can be personal property. Facts we consider of the greatest privacy – our personal income, information about the family, our health record, and employment history-are all tucked away in somebody's database. GIS offers the integration of these data through their common geography. Although it is to the public benefit, for example, to build a link between environmental pollution and health, the more local and individual the link, the more the issue of personal privacy arises. Whole sectors of the economy now rely on linking data from individuals, such as magazine subscriptions and purchases by mail, with demographic and other information by district, such as census tract or zip code. A personal credit history can be amazingly revealing about an individual, and data are often bought and sold as a side benefit of computerized ordering and mailing systems. Just assembling every item of information about an individual, once an extremely difficult task, is now considerably easier.

As GIS becomes used in lawsuits, voting district delineation, and, as always, in mapping of property, the legal profession will come increasingly to use GIS as a tool, and then by extension to challenge the means by which data are collected and transformed, analyses are conducted, and conclusions are drawn. This will force GIS analysts to become somewhat more explicit in their methods and more accountable in their operations. GIS software, for example, should keep a log of the functions used, commands given, menu choices selected, and somehow attach this 'data lineage log' to the data sets themselves. It is well known that regular statistics can be used to support many viewpoints, and even maps can be manipulated to show different points of view. GIS offers the mapping and analysis processes full accountability, and this must be stressed in the future if GIS is not to become yet another courtroom gimmick, like computer graphics, as far as the law is concerned.

DATA OWNERSHIP: There are two philosophies about GIS data ownership. At the one extreme, the federal government produces and distributes digital data in common formats at the marginal cost of distribution, the 'cost of fulfilling user requests'. This means that the cost of producing the data should not enter into the pricing of the data. The logic here is that, because the federal government has already created the data at the public's expense, it cannot charge a second time for data to the same people when they request copies for their own use. The computer networks have made the dissemination cost for the user effectively zero, so that data are usually available for setting up and using a GIS free or at least for only a very modest price. At the opposite extreme lie the groups who believe that GIS data are a commodity, a product to be protected by copyright and patent and sold only at a profit. The argument for this view is that when the market demands a data set, the profit motive will generate the data, and the profit will draw in competitive data producers, who will eventually drive down the cost. There is a great deal of motive to produce a data set that may sell many times, but little motive to map a corner of the country with little demand and poor existing digital maps. Extended to the international context, neither will there be a motive to map for GIS the poorest and most needy nations.

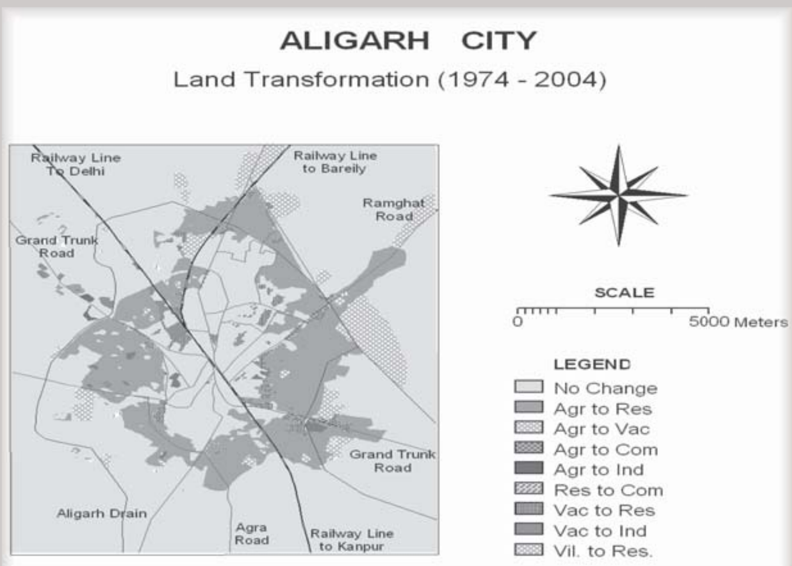
SCIENTIFIC VISUALIZATION: A critical issue for the future of GIS is the degree to which the systems become integrated with those new parts of computer graphics and cartography most suitable for GIS applications. The entire field of scientific visualization is an example. Scientific visualization seeks to use the processing power of the human mind, coupled with the imaging and display capabilities of sophisticated computer graphics systems, to seek out empirical patterns and relationships visible in data but beyond the powers of detection using standard statistical and descriptive methods. Key to the issue of visualization is the ability to model very large and complex data sets and to seek the inherent interrelationships by visual processing alone or with the assistance of standard empirical and modelling methods. Obviously, GIS is the provider of such data sets. GIS data are complex, and the use of maps to begin with already implies that a visual processing mechanism is being used. GIS should move toward full integration with the tools and techniques of scientific visualization and has much to gain by doing so. This would greatly enhance the analysis and modelling component of GIS use, and in a way that is inherently compatible with a GIS and the tools in the GIS toolbox. Many GIS data are also inherently three dimensional, such as atmospheric and ocean concentrations of chemicals, topography, or abstract statistical distributions such as crime rates and population densities over space. New software allows the user of a GIS not only to map and analyze three-dimensional distributions, but also to model and display them in new ways. Among the cartographic methods now familiar to GIS and to automated cartographic system users are simulated hill-shading, illuminated contour, gridded perspective and realistic perspective views, and stepped statistical surfaces. Even simple maps, such as weather maps, now use sophisticated hypsometric colouring with interwoven hill-shading. In addition, new types of display, such as stereo screens with shutters and head-mounted displays, along with the new types of three-dimensional input devices, gloves, track balls, and three-dimensional digitizers, have expanded the suite of interaction means for the GIS user remarkably. Many people who deal with image registration and digitizing work with anaglyphic (red and green) stereo and use soft-copy or computer screen photogrammetry to take measurements. Animation has added another dimension to display and is now commonplace. What was once highly innovative, is now commonplace during the weather forecast of the evening television news. Usually, weather satellite data is animated and the perspective changed to simulate a flyby. The possibilities of animated and interactive cartography, the sort we now see as interactive kiosk-type displays at hotels, airports, and supermarkets, are remarkable, and may strongly influence the future of GIS, especially as the computing power and tools necessary for animation become cheaper and more widespread. Animation has a particular role to play in showing time sequences in GIS applications. Just as it is hard to see exactly what happened during a particular play in a sports contest without slow-motion viewing of film or videotape, so GIS users can compress long time sequences or view short time sequences to reveal geographic patterns that were not visible in other ways.

NEW FOCUS: As GIS moves into the future, changes are inevitable, for GIS is a science and a technology based on change. Nevertheless, there are broad movements within science toward topics or challenges that are national or international areas of new emphasis. A few trends are already obvious; fortunately, GIS has a role to play in each of them. First, science has become increasingly focused on issues of global importance. The earth as a whole system is now a valid way at which to approach issues of global climate change such as global warming and the ozone hole; global circulation, such as the patterns and flows within the earth's oceans and atmosphere; and the global scale of the impact of people on the environment. The new global nature of the world economy, the increasingly strong efforts to solve the world's problems with global legislative bodies, such as the World Bank and the United Nations, and the corning into being of methods and tools for approaching these problems with hard data have allied toward a new global science. GIS has an immense amount to offer this global science. Global distributions need mapping, global mapping needs map projections, and the understanding of flows and circulations are based on an understanding of spatial processes. Even global data collection efforts for GIS are now under way, and organizations use GIS to attack global problems such as crop-yield estimation and famine prediction.

Moreover, GIS has also been at the forefront of a new approach to science. More and more the traditional boundaries between disciplines in the sciences and the social sciences have disappeared, although there are many who fail to recognize it and even resist this trend. Most major research is now conducted by teams, with representatives from a host of different but interrelated sciences working together on a problem. GIS is a natural tool for this sort of work environment because it is able to integrate data from a variety of contexts and sources and seek out interrelationships based on geography, the mapping of distributions, and visualization.

Chapter 12

GIS Project Design and Management



The theoretical and technical knowledge provides a solid foundation which helps in executing GIS projects. Here we examine a framework for the development of some GIS applications. The emphasis is on the practical aspects of designing and managing a GIS application. Design techniques help to identify the nature and scope of a problem, define the system to be built, quantify the amount and type of data necessary and indicate the data model needed and the analysis required.

Management techniques help a project to be delivered on time and ensure quality work. Good project design and management are essential to produce a useful and effective GIS application. The project design and management approach outlined here are suitable examples for small-scale GIS projects-the type of project which may be required by a GIS course or as part of a research project execution. The approach does not embrace any specific design methodology or management philosophy, but it is an integration of many ideas. Various elements of the approach, when scaled up, could provide a methodology for the implementation of larger projects. Any design and management approach adopted should be adapted to meet the needs of the application, the available technology, the users of the system and the organizational culture in which the GIS must reside.

Here we start by considering how the character of the problem for which a GIS solution is being sought can be identified. Two methods are introduced: the rich picture and root definition. A method for constructing a GIS data model is then discussed. A distinction is made between the conceptual data model and the physical implementation of this model in the computer. Cartographic modelling is then considered, as an approach for structuring the GIS analysis required by an application. A review of some project management approaches and techniques and the tools available for the implementation of a GIS project is also discussed. Further, implementation problems and project evaluation are considered.

Problem Identification

Before developing a GIS application the problem that the GIS will address must be identified. There are two techniques that can be used to assist problem identification: creating a rich picture (a schematic view of the problem being addressed), or developing a root definition (a statement of an individual's or group's perspective on the problem). Both these techniques are drawn from the soft systems approach to system design.

The Rich Picture: A rich picture is a schematic view of the problem a project will address. It presents the main components of the problem, as well as any interactions that exist. The rich picture for the urban sprawl GIS study adopts the conventions of the authors, in particular Reeve (1996) and Avison and Wood-Harper (1991). These include the use of:

- *Crossed Swords* – A crossed swords symbol expresses conflict. It is used to indicate the differences between the urban residents and the fringe area residents. There is conflict since the motives of the two groups for system development are different. The urban residents wishes to find land that best suits their needs in the fringe area (where only land is available), whereas the fringe area residents may consider this land for economic activity or open space for healthy environment.

- *EYES* – Eyes are used to represent external observers. Property developers interested in identifying new areas for housing development may be external observers.
- *SPEECH BUBBLES* – Personal or group opinions are indicated in speech bubbles. The different priorities urban land buyers see for the system may be included in the rich picture in this way.

Box 21 : *The soft systems approach*

> The original soft systems ideas were developed by Checkland (1981) and have been added to more recently by other researchers (Wood-Harper et al., 1995). The soft systems approach to problem identification provides a method for addressing unstructured problems (Skidmore and Wroe, 1988). This is useful in a GIS context because many GIS problems are unstructured and often difficult to define. To formulate a problem users should appreciate the context, or world view, from which the problem is being considered. This is the key to the soft systems approach. From the soft systems perspective it is not models of real-world activities which are created, but models of people's perception of an activity. How people feel about and view the activity are included. Therefore, soft systems models are abstract logical models that help with our unerstanding and structuring of a problem.

This way, the rich picture records thoughts on paper and helps to organize ideas. For a small scale project, the rich picture may be drawn by one individual. A rich picture drawn by a project team will represent a consensus view of a problem reached by all the project participants. A single composite rich picture can be achieved by asking all members of the team to draw their own rich pictures. These are then discussed and combined to create a single picture that reflects the views of all parties. Skidmore and Wroe (1988) suggest that rich pictures are particularly useful when considering the design of computer systems within organizations because:

- they focus attention on important issues;
- they help individuals to visualize and discuss the roles they have in the organization;
- they establish exactly which aspects of the information flows within the organization are going to be covered by the system;
- they allow individuals to express worries, conflicts and responsibilities.

The development of a rich picture should not be rushed, particularly if it is trying to reflect an unstructured problem. A poorly defined rich picture may translate into a poor GIS application. An additional check to ensure that the problem is well understood is to develop a root definition.

THE ROOT DEFINITION: Like rich picture, the term root definition also comes from the soft systems approach. The root definition is a view of a problem from a specific perspective. Different users have different views of a problem. In the urban sprawl GIS, the views of groups involved in the design process might be quite different and lead to a degree of conflict. For example, land buyers may see the GIS as 'a system to help identify and rank possible lands', whereas the estate agents may see it as 'a system to help identifying high rent lands which are available for sales'. These two statements are the root definitions of these particular groups. The system developer must get these two groups to agree on a common root definition, for example, 'a system that identifies land for sale which meet the requirements of individual land buyers'.

Establishing a common root definition for a problem will help others to evaluate and understand why a GIS has been constructed in a particular way. Likewise, understanding that others may view a problem from a different perspective will ensure a GIS application is designed to address a range of needs. If a single root definition can be agreed upon then there is a greater chance the GIS will meet the requirements of all concerned. Once rich picture and root definition exist, the main aims and objectives for a project can be identified and a GIS data model can be created.

If it proves difficult to draw a rich picture or formulate a root definition then the problem being addressed may be unstructured. Unstructured problems are the most difficult to address with GIS. However, the rich picture method can still be used; typically it will start with only a few elements of the problem clearly defined. Additional elements are added after talking to potential users of the GIS, consulting the literature and discussing the project with others working in a similar field.

As the rich picture is developed and the root definition formulated the resources available to the project must also be considered. In some cases the only resource will be one individual's time and commitment. In larger projects there may be access to several members of staff and a budget. It is important to consider, given the resources available, whether it is possible to address the whole problem that is unfolding, or whether it will be necessary to break the problem down into smaller parts. Breaking the problem down into more manageable pieces may allow quicker results, which may be important where the GIS activities are taking place in a large organization. Small but useful results, produced relatively quickly, will gain recognition and respect for a project. This may result in further support and resources being allocated to the project. Therefore, many system designers use pilot projects to produce results quickly. These results can be disseminated widely throughout the organization to encourage support for the GIS.

Designing a Data Model: The rich picture and root definitions that define a problem must be turned into a GIS data model. Here the term data model is used as a collective term for the process of identifying all the design elements used in the construction of a GIS (Peuquet, 1984 and Frank and Mark, 1991). Worboys (1995) offers a useful solution to this confusion by distinguishing between conceptual and physical data models. The conceptual data model is a high-level view that is independent of the computer system. This is the user's view of a problem and its elements. This is close to the way that Peuquet (1984) and Frank and Mark (1991) use the term. The physical data model, on the other hand, describes the organization of data in the computer.

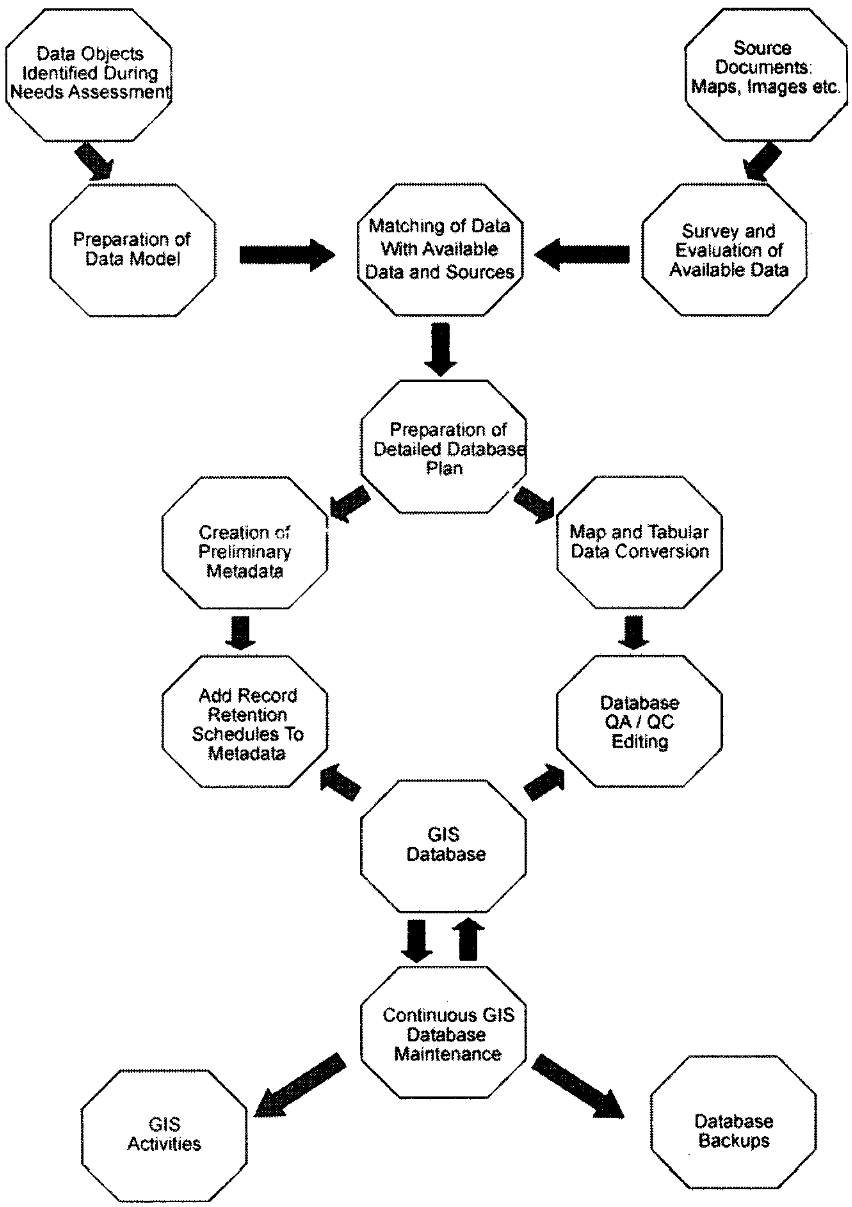


Figure 12.1: Conceptual design of GIS. It includes formal modelling of GIS database and its planning activity.

Therefore, from the project design and management perspective, it is useful to think of the GIS data model as consisting of two parts: a conceptual model and a physical model. The conceptual data model adds spatial detail to the rich picture by including elements of spatial form and spatial process. The physical data model is concerned with how to represent the conceptual model within the computer. Details about the spatial data model (raster or vector), the appropriate data structure and the analysis scheme are included in the physical data model. The data modelling stage is frequently neglected in the design and development of GIS projects, often with disastrous consequences. Insufficient attention to data modelling may lead to the failure of the GIS to meet the expectations of users.

CONCEPTUAL AND PHYSICAL DATA MODEL: One way to create a conceptual data model is to borrow heavily from the ideas of hard systems analysis. Hard systems analysis advocates the clear identification of the elements of the data model: the entities, their states and their relationships to each other. One method of presenting this is using a flowchart. In systems analysis, flowcharts use a range of symbols to communicate different aspects of the model. GIS terminology can be avoided when constructing the conceptual data model. This is a good idea as the resulting flowchart will then explain what it is the GIS application will do in a way that is clear to all interested parties. It will give those with little GIS experience the opportunity to provide feedback on the approach. Moreover, if the conceptual model is jargon-free it can be given to GIS programmers with different software backgrounds. This may be an advantage in large-scale GIS projects where an organization wishes to compare how well different software products can address a task.

The GIS Paradigm

Data Management Principles: The logical structuring of databases that contain geographic information and other related data.

Technology: The effective combination of hardware and software components that enables the automation of numerous geographic data handling functions.

Organizational Setting: A management environment that provides resources and enables changes to be made for incorporating GIS utilization within the organization.

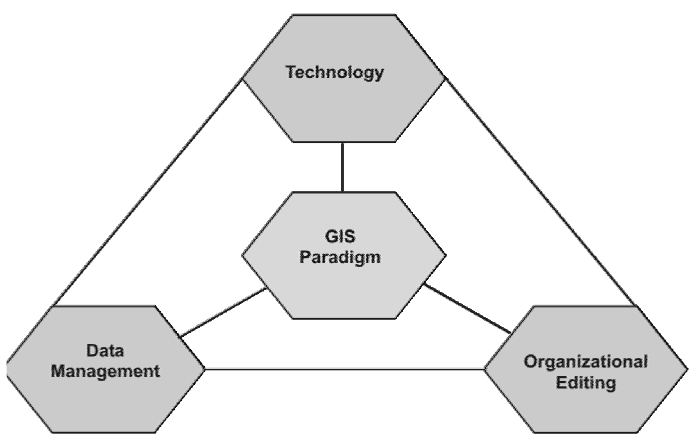


Figure 12.2: GIS paradigm.

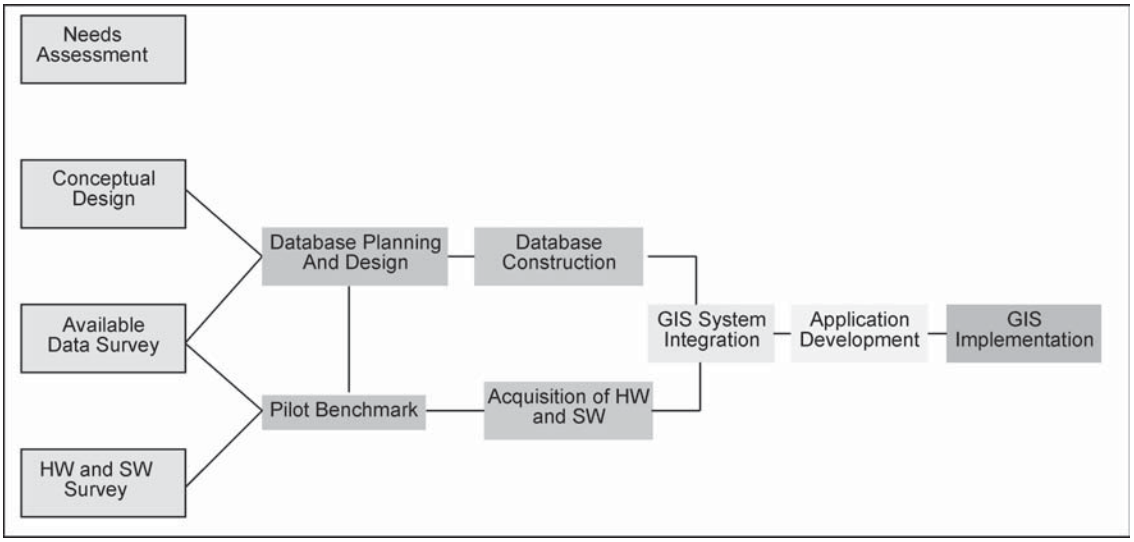


Figure 12.3: GIS development cycle.

Box 2.2: Hard systems analysis and GIS

Hard systems analysis advocates trying to understand reality by rebuilding part of it. The link to GIS is clear, as GIS data models attempt to reconstruct parts of reality for specific purposes. During the 1970s and early 1980s the hard systems approach was the dominant methodology used for the design of computer systems. It is possible that the early developers of GIS software used a hard systems approach to design.

There are four phases in hard systems analysis. These are outlined below (after Huggett, 1980). There are three important terminology of the hard systems approach. These are entities, states and relationships. The *entities, or elements* of a system, are either physical objects or concepts. In GIS terms entities are points, lines, areas, surfaces and networks. Entities also possess properties known in hard systems terms as states. The states associated with an entity give its character. In GIS terms states are attributes. In addition, relationships exist between entities. In GIS this relationship could be the topological links between features.

The four stages in hard systems analysis, in a GIS context, are:

- **The lexical phase-**The objectives of the lexical phase are:
  - to define the problem;
  - to define the boundaries of the problem;
  - to choose the entities that define the components of the problem;
  - to establish the states of these entities.

Contd....

In GIS this involves:

- identifying the nature of the application;
- selecting the study area;
- defining the real-world features of interest;
- identifying associated attributes.

- **The parsing phase-**In the parsing phase the relationship between entities and groups of entities are defined. The entities and knowledge about their states are used to create a computer model.
- **The modelling phase**-In this phase the GIS is used to address the problems identified during the lexical phase. The way in which entities and their states will interact and respond under differing situations is expressed. This may involve linking GIS software to other software.
- **The analysis phase**-This phase is the validation of the modelling phase. Testing occurs to find out how closely the GIS model (of both form and process) fits what is observed in reality.

Bell and Wood-Harper (1992) provide a useful checklist for the development of a conceptual model:

i. **Develop a rich picture and root definition:** Everyone associated with the problem should agree upon these. They are used to focus the aims and direction for the project.

ii. **Create a list of actions the system must be able to perform:** In the urban sprawl example these actions may include permitting users to select neighbourhood characteristics such as proximity to roads, railway stations and shops, land rent and allowing users to weight these characteristics in terms of their relative importance. These actions are known as activities.

iii. **Identify a list of system inputs and outputs:** In GIS terms system inputs are data sources and outputs are products such as maps. In the urban sprawl example, the data sources would include land use data, development authority plans, street networks, existing land rents and the location of properties for sale. Outputs might be a list, or map, of properties meeting land buyers criteria.

iv. **Group activities, inputs and outputs into a logical, chronological order:** Arrows symbolizing some form of action are used to join activities together. For example, in the urban sprawl GIS the combination of data from different sources could be effectively represented in this way.

The physical data model requires additional detail that describes how to model the spatial entities, their associated attributes and the relationships between entities in the computer. Therefore the emphasis here is on developing a model of the relationships between entities. This is frequently referred to as an analysis scheme. There are a number of different techniques for designing an analysis scheme that can be used; here we describe an approach known as cartographic modelling.

Algebraic Equation for a Simple Overlay Operation

A + B = C

Where A = Map showing river,
B = Map showing canals,
C = Map showing sources of irrigation

Graphical Representation

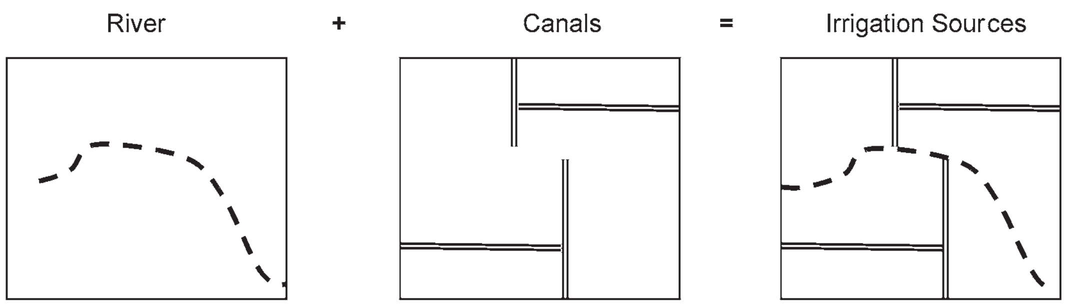


Figure 12.4: A simple map algebra equation.

CARTOGRAPHIC MODELLING: Tomlin (1991) states, cartographic modelling are derived from a collection of old ideas that have been organized, augmented and expressed in terms amenable to digital processing. However, it is the work described by Tomlin (1983) as 'Map Algebra' and Berry (1987) as 'Mapematics' that established cartographic modelling as an accepted methodology for the processing of spatial information. Cartographic modelling, at its simplest, is a generic way of expressing and organizing the methods by which spatial variables, and spatial operations, are selected and used to develop a GIS data model. Tomlin (1991) considers that the fundamental conventions of cartographic modelling are not those of any particular GIS. They are generalized conventions intended to relate to as many systems as possible. A number of GIS software products uses the concepts of cartographic modelling in their approach to spatial analysis.

The concepts that underpin cartographic modelling borrow heavily from mathematics. Cartographic modelling is a geographic data processing methodology that views maps (or any spatial data layer) as variables in algebraic equations. In algebra, real values are represented by symbols, such as x, y and z. In map algebra these symbols may represent numeric attributes of map elements (for example, pH values associated with a given soil type) or even whole maps. Numbers assigned to symbols in an equation interact to generate new numbers using mathematical operators such as add, subtract, multiply and divide. In the same way, in map algebra, maps are transformed or combined into new maps by the use of specific spatial operations.

There are four stages in the development of a cartographic model:

i. Identify the map layers or spatial data sets required.
ii. Use natural language to explain the process of moving from the data available to a solution.
iii. Draw a flowchart to represent graphically the process in step 2. In the context of map algebra this flowchart represents a series of equations one must solve in order to produce the answer to the spatial query.
iv. Annotate this flowchart with the commands necessary to perform these operations within the GIS one is using.

Box 23 : Examples of natural language keywords

Keyword – **Spread**
Operation – to create a corridor from a linear data set or a zone of influence around a point.
Description – calculate the distance of all geographical positions in the data set from a given point or line.
Example – to create a buffer zones along the main roads and analyze the urban land use.

Keyword – **Overlay**
Operation – to find the intersection of two different sets of area entities covering the same geographical area.
Description – lay two different sets of area entities over each other to produce a new complex set of areas.
Example – to create a map to analyze the influence of accessibility on urban land transformation along the roads.

Keyword – **Extract**
Operation – to extract a new data set from an existing data set.
Description – select specified values or class from one overlay to make a new map.
Example – to create a new data layer showing agricultural land which is transformed in buffer zones along the main roads.

Example of simple equation for GIS analysis:

Equation 1 – EXTRACT from a land use map

A – B = C here, A = land use, B = Kharif cropped area, C = area under rice.

Equation 2 – SPREAD from irrigation potential map

(D – E) + F = G here, D = canal, E = areas which are not canals, F = potential canal irrigated area, G = actual canal irrigated area.

Equation 3 – EXTRACT good water quality tube wells from all the tube wells

H – I = J here, H = all the tube wells in the area, I = tube wells with hard water, J = good water quality tube wells.

Equation 4 – OVERLAY identifying good water quality tube wells which can irrigate area under rice

J + K = L here J = good water quality tube wells, area under rice which is not irrigated by canal, L = tube wells which can irrigate the rice area and that can not be irrigated from canal.

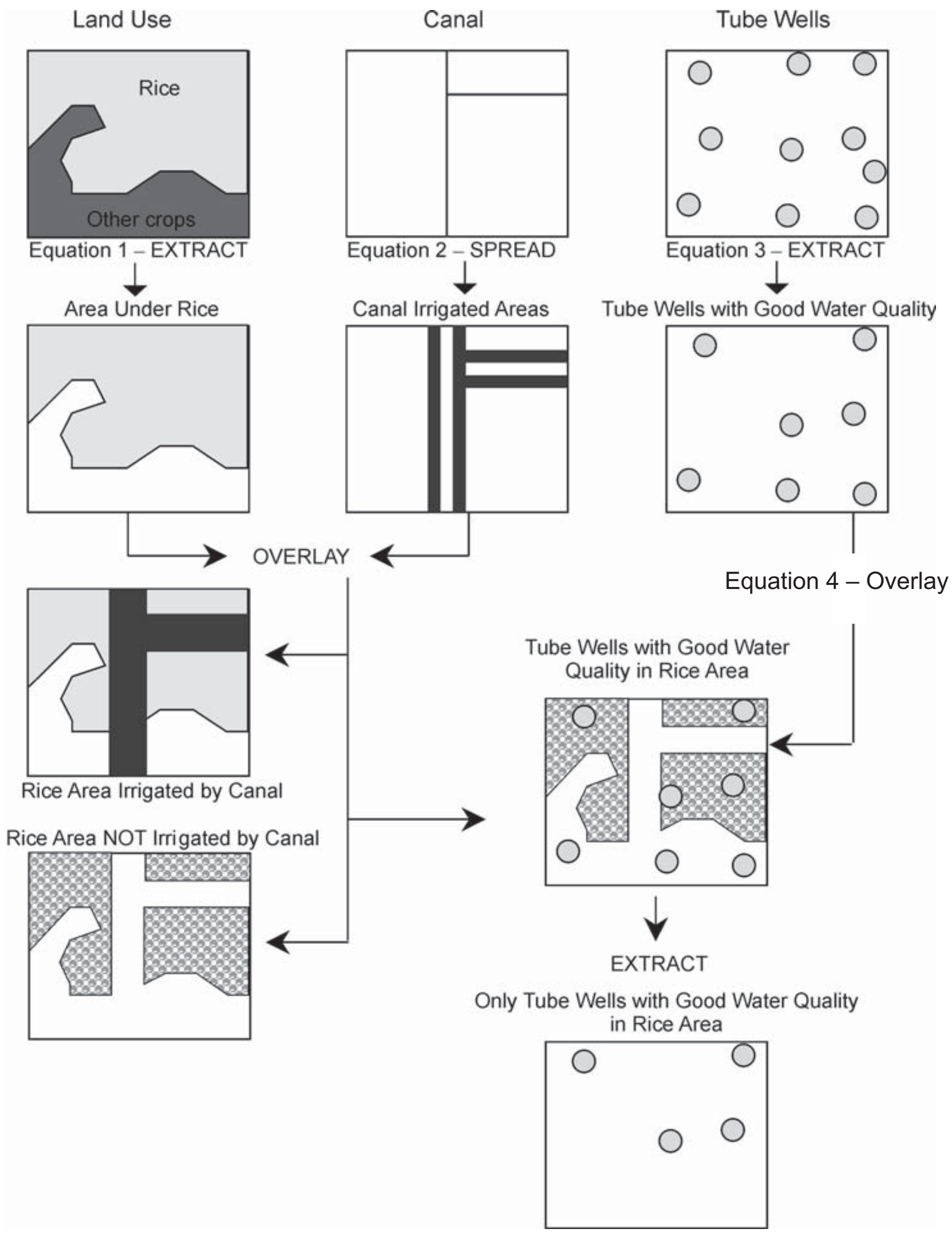


Figure 12.5: An example of GIS analysis for identifying good water quality tube wells which can irrigate area under rice.

This structure allows the designer to tackle a complex spatial problem by breaking it down into its components. Simple statements, sections in a flowchart or solvable equations can then express these. However, it may not be possible or sensible to do all the analysis in the GIS. In certain cases, it may be necessary to couple the GIS with other applications to obtain results. If care has been taken over the construction of the rich picture and root definition and their subsequent translation into a conceptual and physical data model, then, at least in theory, the computer implementation of a GIS data model should be relatively straightforward. It is likely that building an application will start while detail is stil being added to the GIS data model. There is nothing wrong with this approach as long as the enthusiasm for implementation does not take over, leaving missing details forgotten. Trial and error with systems development is an accepted approach, but often individuals working alone may spend many hours developing a solution only to find that they have not documented their work and are therefore unable to explain to others how the result was reached. How one implement a physical data model will depend upon the nature of the problem and the organizational setting in which the individual works. However, good project management will help ensure that goals are met.

Project Management

Here we consider techniques for identifying the character and extent of a spatial problem and techniques for helping with the design of GIS data models. Once the data model is constructed the GIS must be implemented and in many cases integrated into the wider information strategy of an organization. To help this process good project management is an essential prerequisite for success. There are many different approaches to managing information technology (IT) projects. Two approaches commonly used by GIS designers are the system life cycle and prototyping.

Systems Life Cycle Approach : The systems life cycle (SLC) approach advocates a linear approach to managing the development and implementation of an IT system. It is also referred to as the 'waterfall model' (Skidmore and Wroe, 1988). The waterfall analogy is used because the outputs from the first stage of the process inform the second phase, and the outputs from the second phase affect the third phase, and so on. There are many variations on the general approach:

i. *Feasibility Study:* This would involve asking the real estate agents and land buyers questions about whether they would make use of the system being proposed for development and what the costs and benefits of developing a GIS would be. If the feasibility study is positive then the project moves to the second phase.
ii. *System Investigation and System Analysis:* The GIS designer would try to establish the current way in which land buyers and real estate agents interact to identify land for sale in appropriate neighbourhoods. This would include identifying the data and analysis requirements as well as the preferred output types. A soft systems approach could be used to help with this phase.
iii. *System Design:* The GIS data model is constructed using information collected in the previous phase. In the urban sprawl example, cartographic modelling techniques might be used to help structure the analysis requirements of the GIS.
iv. *Implementation, Review and Maintenance:* Now the urban sprawl GIS is built and provided to users. This may be the first opportunity for users to comment on, or interact with, the system since their involvement in the feasibility study. Users experiences

inevitably require changes to the system. These may include the addition of new data layers, new analysis techniques or new ways of visualizing the output.

The main advantage of the systems life cycle approach is that it provides a very structured framework for the management of a GIS project. This can be extremely important when good time management is an essential aspect of the project. In addition, it is often easier to budget for the resources required by a systems life cycle approach because the requirements of the system are established at an early stage in the project. Despite its popularity as a project management tool for IT projects there are a number of problems with the systems life cycle approach:

- Designers who use the systems life cycle approach often fail to address the context of the business for which the system is being developed. The approach encourages the designer to focus on only apart of the information problem of an organization.
- The timescale and linear nature of the systems life cycle process do not allow for change in the scope and character of the problem. By the time the system is implemented it may be out of date.
- The systems life cycle approach does not put the user at the centre of the system design. It emphasizes the identification of flows of information rather than understanding why they are required. This creates problems because it only allows a system that reflects the current way of doing things. This may be a problem for GIS design as a new system may radically change the way information is managed.
- The systems life cycle approach is often considered to favour hierarchical and centralized systems of information provision. It offers a very technocentric view of system development.

THE PROTOTYPING APPROACH: The prototyping approach to IT project management developed as a response to the criticisms of the systems life cycle approach, particularly in response to the lack of consideration of users. The user first defines the basic requirements of the system. This could be achieved by using the rich picture and root definition techniques. The system designer takes these basic ideas to construct a prototype system to meet the needs identified by the user. In GIS projects such systems are often described as demonstrators. The users who identified the original requirements for the system then experiment with the demonstration system to see if it is what they expected. Other potential users of the final system may be brought in at this stage to see if the system is of wider value. The system designer uses their recommendations to improve the system.

The prototyping approach has a number of advantages over the systems life cycle method:

- Users have a more direct and regular involvement in the design of the system.
- It is easier to adapt the system in the face of changing circumstances which were not identified at the outset of the project.
- The system can be abandoned altogether after the first prototype if it fails to meet the needs of users. This reduces the cost of developing full systems.

- If money and time are available a number of prototypes can be built until the user is satisfied.

The drawbacks of prototyping approach are:

- Prototyping can be difficult to manage. There may be large numbers of users with large numbers of ideas and opinions.
- The resource implications may change following the development of the first prototype.
- Knowing when to stop development can also be a problem. However, some GIS designers argue that this is a positive aspect of the approach since few, if any, GIS systems are ever finished.

Box 24: Project management techniques

SWOT Analysis (Strength, Weakness, Opportunities and Threats) – This technique is used to establish the SWOT associated with the development of the GIS. It is used as part of the feasibility study in systems life cycle approach.

Rich Picture and Root Definition – These techniques comes from 'soft system' methodology, they are used to help system designers determine the scope of a problem.

Demonstration Systems – These are demonstration GIS applications, designed to help users evaluate systems efficiency. It is used more commonly in prototyping approach is followed for project management.

Interviews and Data Inventories – These techniques are used for problem definition, establishing current information and analysis requirements. Data audits are more structured and valuable in GIS because it evaluates the availability of spatial data.

Organization Charts, System Flowcharts and Decision Trees – These three techniques are all variations on the flowcharting theme. The organization chart maps out the flows of information with in the organization. The system flowchart describes how the system will model these information flows. The decision tree shows the problem from a decision making perspective and focuses on showing how different decisions cause information to be used in different ways within the organization. The technique used will depend on the experience of the system designer and the character of the problem.

Data Flow Diagram and Dictionaries – These techniques are drawn from hard systems analysis and represent a more structured approach to system design. They can be of immense value in GlS for tracking what happens to a data layer through the analysis process. This is extremely valuable in monitoring data quality and providing lineage information.

Cartographic Models and Entity Relationship Diagrams – These techniques are of most valuable in structuring the analysis schemes used in GIS. They help in planning the functional requirements of the analysis.

The systems life cycle and prototyping approaches are just two of many that can be adopted for the management of a GIS project. There are also many variations on the basic approaches outlined above. It is also possible to pick and mix aspects of the two approaches to develop a management style that suits the development environment. In addition, there are a wide range of project management techniques and tools which can be used to help with various phases in the systems life cycle and prototyping approaches.

IMPLEMENTATION PROBLEMS: There will always be problems for GIS design and development if proper planning is not done for developing the project. Three of the most common problems are:

i. data in the wrong format for the GIS software;
ii. a lack of GIS knowledge imposing technical and conceptual constraints on a project;
iii. users of the GIS frequently changing their mind about what they want the GIS to do.

In many GIS projects the data required are unavailable in a format compatible with the GIS software or analysis needs. If this is the case there are two options: to look elsewhere for a supplier, or to convert the data into the desired format. However, in the case of the latter, errors may creep in as data are changed from one format to another. Alternatively, the conceptual data model could be revisited to assess the importance of the data, and evaluate alternative data options.

It is inevitable, at some point, that applications will be limited by users technical or conceptual knowledge about which spatial data model, data structure or analytical operation is most appropriate for the task required. Much can be learnt from other applications and other users. Colleagues working in similar areas, other organizations, or the Internet can all be sources of help. For many organizations, the solution is to employ an independent expert to undertake application development or specific analysis.

The dynamic nature of the GIS design process is such that the information needs of users are often in a constant state of flux. By the time a GIS data model is implemented, the needs of the users and the scope of the problem may have moved away from the original defined by the rich picture. This is a major issue in the development of GIS applications for larger organizations, where applications development may take considerable time, and the awareness of key players about GIS may increase in the meantime. The solution is to gain frequent feedback from the individuals who will be the end-users of the GIS. They should reconfirm that the scope of the project has not changed, or allows changes to be brought into the design process. Even in a small-scale GIS project, as the knowledge of the GIS analyst grows there may be changes in the aims and scope of a project.

PROJECT EVALUATION: After a GIS application has been constructed, some problems may be just about to start. It is important that the output produced by the system is usable, valid and meets the goals set at the beginning of the project. Validation of results is often difficult to achieve, particularly if results are in the form of predictions. However, if one is working with an organization, testing the GIS and validating output will be a crucial part of the design process. In many cases, this may well result in adjustments to the rich picture and the GIS data model. In extreme cases the GIS may have to be abandoned, and the project restarted.

This feedback process can be very costly and often explains why many organizations adopt the prototyping approach to project management. Prototyping should prevent a system from being inappropriate, as frequent testing and evaluation should be taking place.

There are three tests that can be used to check whether a GIS application meets the goals set for it at the start of the design process. First, all the parties involved in the design and development of the GIS can be asked if they are using the application for the purpose for which it was designed. If they are not, or have even reverted to using the old methods, it is a sure sign that something has gone wrong somewhere. The goals originally used to guide them in helping to identify the scope of the problem may have changed as time and work priorities have altered. Alternatively, users may find the application difficult to use, or without a key feature. In such cases, further training, or adaptations to software, may be all that is required to ensure that the GIS are effectively used.

Second, the GIS output can be checked against reality. For example, this type of test would be appropriate for the flood prediction model as the location of flood could be predicted before the start of a season and then compared against actual flood at the end of the season.

Third, the adaptations and changes that had to be made when moving from the rich picture through the GIS data model to the GIS implementation can be evaluated. Whether these were due to knowledge deficiencies, the problem of definition or to system adaptations because the software or data would not permit implementation of the model as planned can be assessed. If the system adaptations have dominated the development of the application then it may be that a technical solution has been provided that has little resemblance to the reality of the initial problem.

Project Design – An Example (Urban Sprawl GIS)

GIS project design is identical to any plan needed to solve a problem or fill a need. Here the author gives an example of one of his work, where urban – land use, land transformation and sprawl was examined.

1. **Objective**
    a. **Identification of the problem:** Examine the land use data of different time period and find the changes in land use over time. What are the classes which are increasing or which are decreasing? Does it suggest any trend?
    b. **What are the final products:** Hard copy maps, digital data base, statistics, reports, decision support system etc.
    c. **Who is the audience:** Urban development authority, urban administrators, land revenue officials, politicians, environmentalists, researchers etc.
    d. **Who else can use this data:** The state level and national level policy makers and planners as the study would reflect a general trend for growing urban centre in the country and the findings of the city may help in assessing the loss to

agricultural land, intensification of urban lands leading to congestion and stress on basic amenities etc.

2. **Database Design**

   a. ***Identification of required and optional data layers:*** What do we need to map urban land use – the different time period land use data.

   b. ***Identification of required and optional attributes:*** What type and level of classification will be employed and what other data we need to support our modelling of urban expansion.

   c. ***Definition of attributes and codes:*** What is the classification system for urban land use and how do we code that into the database. In case of additional data to support the mapping and analysis is used, what would be the nature of the existing attributes and do we need to modify them.

   d. ***Registration of map layers to a standard base map:*** What are the scales of existing database, and are those scales appropriate to our mapping scale? Are the maps in the same projection and coordinate system? What is the accuracy of the digitizing process that input these data?

   e. ***Geographic data encoding schemes (point, line, polygon):*** How we will represent urban land use, land ownerships, land values, ground truth locations etc.

   f. ***Allocation of storage space:*** Given that we will generate a digital data set consisting of many attributes, how large will be the dataset and how we manage, store and preserve these data.

3. **Database Automation**

   a. ***Input data***

   b. ***Topology creation***

   c. ***Input attributes***

   d. ***Building user interface (specific to the needs of users)***

4. **Database Management**

   a. ***Creation of coordinate system:*** Putting all the database to a common coordinate system and projection.

   b. ***Data tiling:*** Joining adjacent areas into database or splitting large database into tiles and develop a management strategy for tiles. Tiles are physical subset of a larger geographic area which contains identical themes. Tiles are essential to reduce data load on computers allowing faster computation.

5. **Data Analysis**

   a. ***Overlay***

   b. ***Buffer***

- *c.* ***Merge***
- *d.* ***Recode***
- *e.* ***Network***
- *f.* ***Terrain modelling***
- *g.* ***Spatial adjacency***
- *h.* ***Data transformations***

6. Presentation of Results

- *a.* ***Preparation of hard copy maps***
- *b.* ***Preparation of digital database***
- *c.* ***Preparation of summaries and statistics***
- *d.* ***Final report and recommendations***

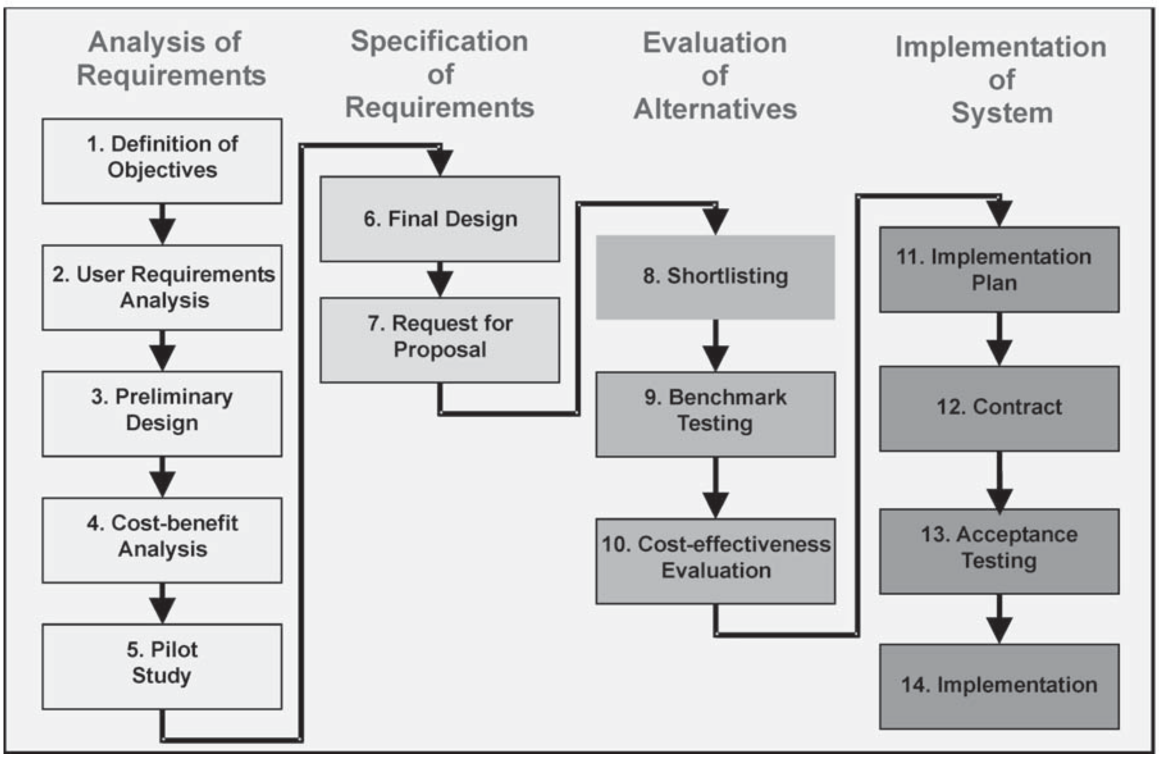


Figure 12.6: GIS acquisition process.

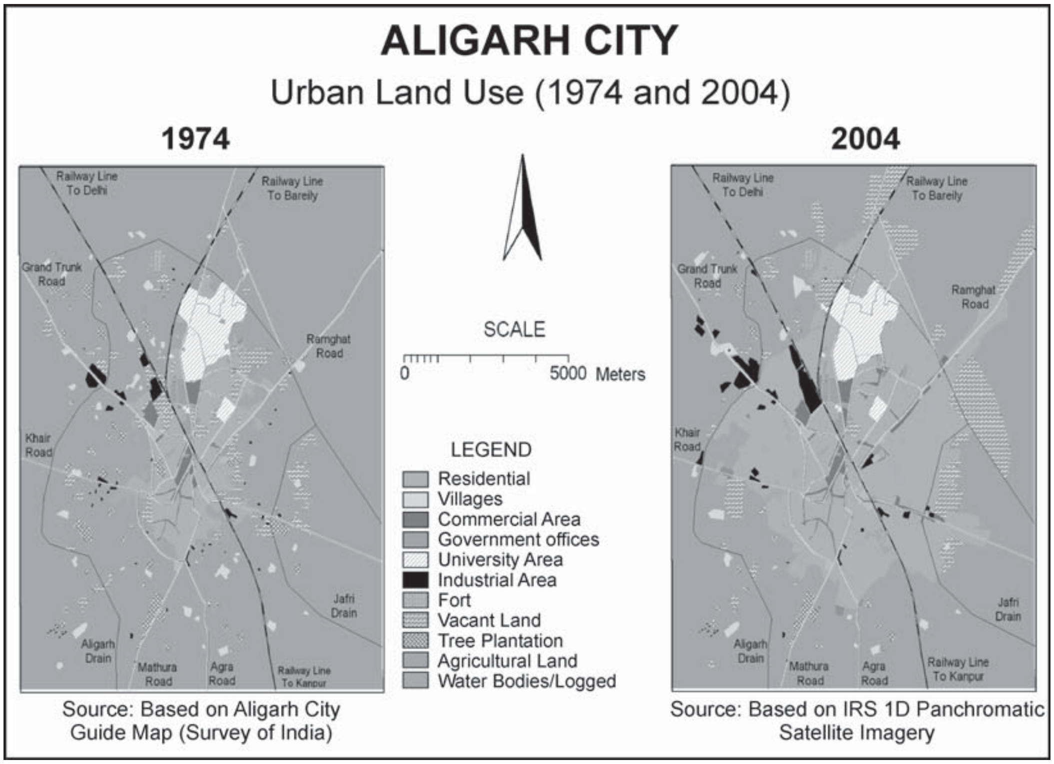


Figure 12.7: The urban sprawl GIS study, we prepared two urban land use map of Aligarh city, the map was prepared by scanning and on screen digitizing the Survey of India, City Guide Map and IRS 1D Panchromatic Satellite Imagery.

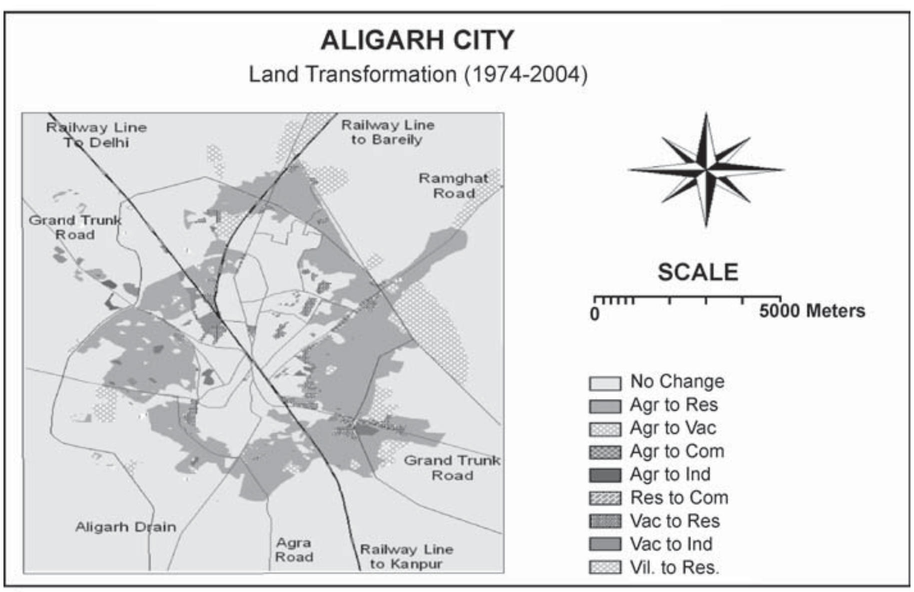


Figure 12.8: Land transformation map was prepared by overlaying the two different period land use map. The areas with no change represent areas where the same land use class exists, while other areas have been transformed from their earlier class.

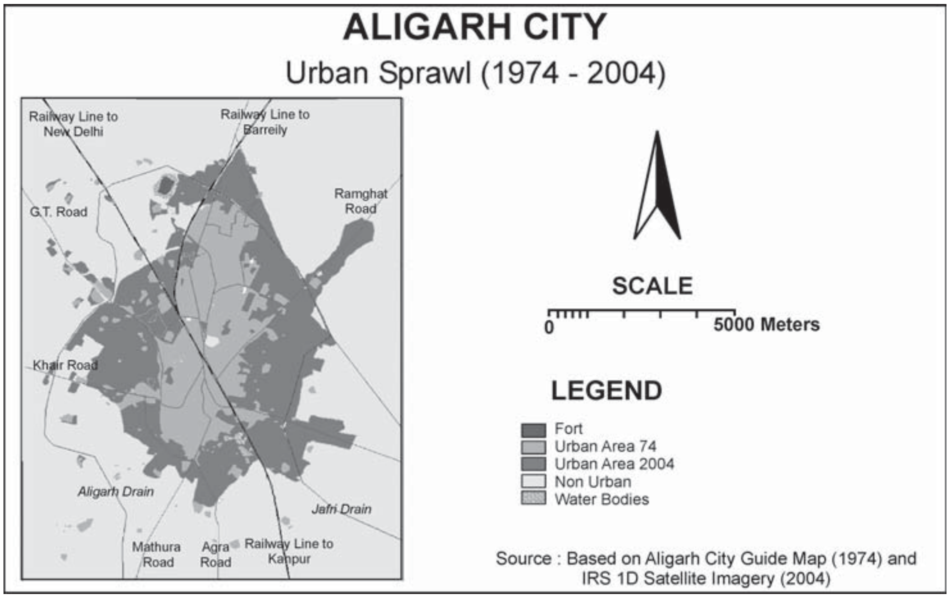


Figure 12.9: Urban sprawl map was prepared by first merging all the urban land use class, such as residential, commercial, institutional, vacant and university area and assigning them a class – 'urban area'. This reclassification of map was done for both time period data. Finally the two maps were overlaid over each other to find urban expansions.

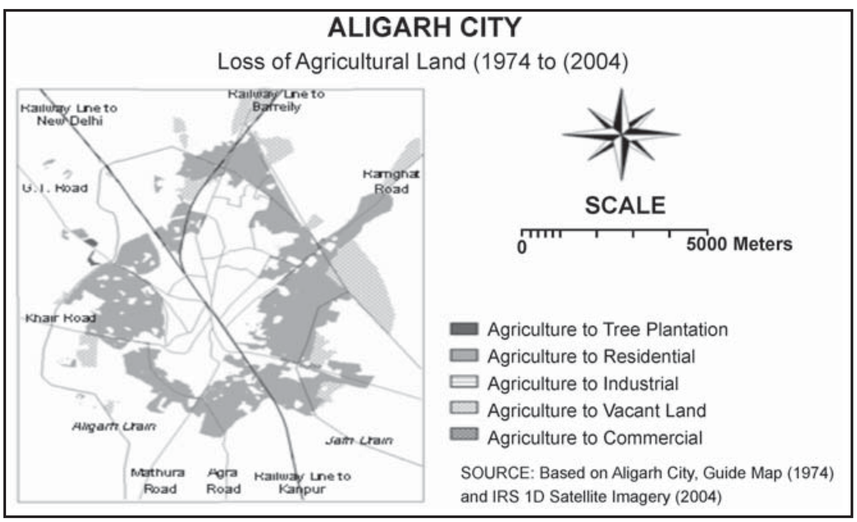


Figure 12.10: Loss of agricultural land was estimated by masking all the land transformation class except where agricultural land was transformed.

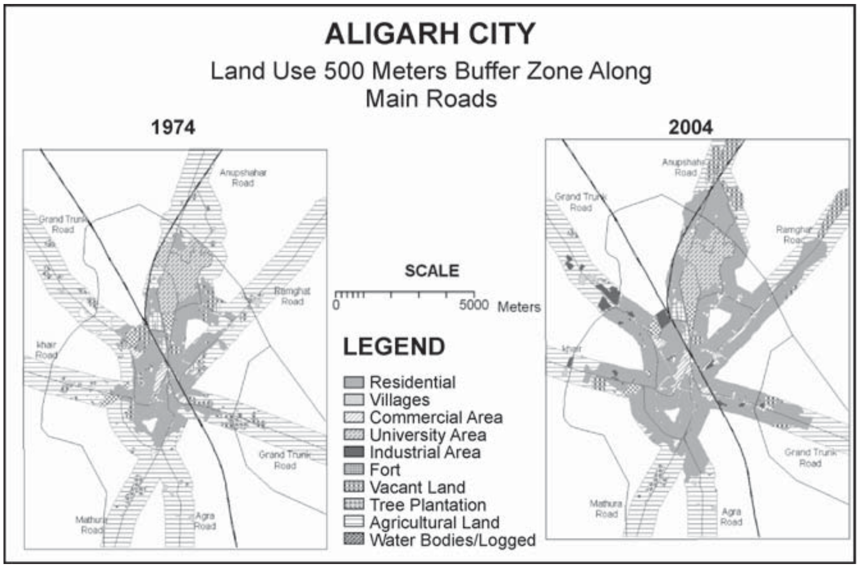


Figure 12.11: Land use buffer along the main roads of Aligarh city for 500 meters on either side of roads was prepared. First, a buffer along the roads was made and then it was overlaid on land use map.

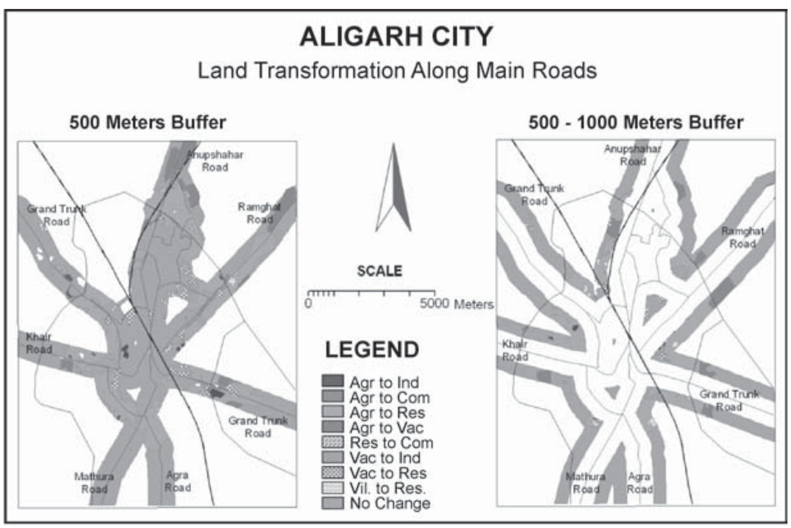


Figure 12.12: Land transformation along the main roads of Aligarh city for two separate buffer zones on either side of roads was prepared. First, two buffers along the roads (one 500 meters from the roads and another 500 to 1000 meters, excluding the first 500 meters) were made and then it was overlaid on land transformation map.

This page intentionally left blank

Glossary

This glossary contains the general definition of most non-standard words used in the text. It is not intended as a comprehensive GIS glossary. For a comprehensive glossary, refer to GIS Terminology by Fazal, S. and Rahman (2007).

AAT (Arc Attribute Table): A table containing attributes for a line coverage such as streets or streams.

Absolute Location: A location in geographic space given with respect to a known origin and standard measurement system, such as a coordinate system.

Active Data: Data that can be reconfigured and recomputed in place. Spreadsheet term for data for attributes or records created by formulas within a spreadsheet.

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

Adjacency: The topological property of sharing a common boundary or being in immediate proximity.

Affine Transformation: Any set of translation, rotation, and scaling operations in the two spatial directions of the plane. Affine transformations allow maps with different scales, orientations, and origins to be co-registered.

AM/FM: Automated Mapping and Facilities Management. The management of mapping and facilities management using integrated computer software.

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

Arc: A string of x,y coordinate pairs (vertices) that begin at one location and end at another. Connecting the arc's vertices creates a line.

ARC/NODE: Early name for the vector GIS data structure.

ARRAY: A physical data structure for grids. Arrays are part of most computer programming languages, and can be used for storing and manipulating raster data.

ARTIFICIAL INTELLIGENCE (AI): Field of study concerned with producing computer programs capable of learning and processing their own 'thoughts'.

ASCII: The American Standard Code for Information Interchange. A standard that maps commonly used characters such as the alphabet onto one byte long sequences of bits.

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

BASE LAYER OR MAP: A GIS data layer of reference information, such as topography, road network, or streams, to which all other layers are referenced geometrically.

BASIC SPATIAL UNIT (BSU): The smallest spatial entity to which data are encoded.

BIT: The smallest storable unit within a computer's memory with only an on and an off state, codable with one binary digit.

BOOLEAN OVERLAY: A type of map overlay based on Boolean algebra.

BUFFER: A zone of a specified distance around coverage features. Both constant and variable width buffers can be generated for a set of coverage features based on each features attribute values.

BYTE: Eight consecutive bits.

CAD: Computer Aided Design. An automated system for the design, drafting and display of graphically oriented information.

CALFORM: An early computer mapping package for thematic mapping.

CAM: (Computer-Assisted Mapping): A map projection and outline plotting program for mainframe computers dating from the 1960s.

CARTESIAN COORDINATE SYSTEM: A two-dimensional coordinate system in which x measures horizontal distance and y measures vertical distance. An x,y coordinate defines every point on the plane.

CARTOGRAPHIC SPAGHETTI: A loose data structure for vector data, with only order as an identifying property to the features.

CENTROID: A point location at the center of a feature used to represent that feature.

CGIS (CANADIAN GEOGRAPHIC INFORMATION SYSTEM): An early national land inventory system in Canada that evolved into a full GIS.

CLIP: The spatial extraction of those features from one coverage that reside entirely within the boundary defined by features in another coverage. Clipping works much like a cookie cutter.

CLUMP: To aggregate spatially; to join together features with similar characteristics into a single feature.

COGO: Abbreviation for the term Coordinate Geometry. Land surveyors use COGO functions to enter survey data, to calculate precise locations and boundaries, to define curves, and so on.

COMPRESSION: Any technique that reduces the physical file size of data in a spatial or other data format.

CONCEPTUAL DATA MODEL: A model, usually expressed in verbal or graphical form, that attempts to describe in words or pictures quantitative and qualitative interactions between real-world features.

CONCORDANCE-DISCORDANCE ANALYSIS: Method of MCE based on lengthy pair-wise comparison of outranking and dominance relationships between each choice alternative in the choice set.

CONNECTIVITY: The topological property of sharing a common link, such as a line connecting two points in a network.

CONTINUITY: The geographic property of features or measurements that gives measurements at all locations in space. Topography and air pressure are examples.

CONTROL POINTS: A set of points on the ground whose horizontal and vertical location is known. Control points are used as the basis for detailed surveys.

CONVERGE: The eventual agreement of measurements on a single value.

COOKIE-CUT: A spatial operation to exclude area outside a specific zone of interest. For example, a state outline map can be used to cut out pixels from a satellite image.

COORDINATE: An x,y location in a Cartesian coordinate system or an x,y,z coordinate in a three dimensional system. Coordinates represent locations on the Earth's surface relative to other locations.

COORDINATE SYSTEM: A system with all the necessary components to locate a position in two- or three-dimensional space: that is, an origin, a type of unit distance, and axes.

CORINE: The Coordinated Information on the European Environment Programme, initiated in 1985 by the European Union to create a database that would encourage the collection and co-ordination of consistent information to aid European Community policy.

COVERAGE: A digital version of a map forming the basic unit of vector data storage in ARC/INFO. A coverage stores map features as primary features (such as arcs, nodes, polygons, and label points) and secondary features (such as tics, map extent, links, and annotation). Associated feature attribute tables describe and store attributes of the map features. A coverage usually represents a single theme, or layer, such as soils, roads, or land use.

DATA BIAS: The systematic variation of data from reality.

DATA CONVERSION: The translation of data from one format to another. ARC/INFO supports data conversion from many different geographic data formats in addition to routines for converting paper maps. Those data formats include DLG, TIGER, DXF, and DEM.

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

DATA EXCHANGE: The exchange of data between similar GIS packages but groups with a common interest.

DATA EXTREMES: The highest and lowest values of an attribute, found by selecting the first and last records after sorting.

DATA FORMAT: A specification of a physical data structure for a feature or record.

DATA MINING: Revisiting existing data to explore for new relationships using new and more powerful tools for analysis and display.

DATA MODEL: A logical means of organization of data for use in an information system.

DATA RETRIEVAL: The ability of a database management system to get back from computer memory records that were previously stored there.

DATA STRUCTURE: The logical and physical means by which a map feature or an attribute is digitally encoded.

DATA TRANSFER: The exchange of data between non-communicating computer systems and different GIS software packages.

DATA: A set of measurements or other values, such as text for at least one attribute and at least one record.

DATABASE: A logical collection of interrelated information, managed and stored as a unit. A GIS database includes data about the spatial location and shape of geographic features recorded as points, lines, and polygons as well as their attributes.

DATABASE MANAGER: A computer program or set of programs that allows a user to define the structure and organization of a database, to enter and maintain records in the database, to perform sorting, data reorganization, and searching, and to generate useful products such as reports and graphs.

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

DATUM: A base reference level for the third dimension of elevation for the earth's surface. A datum can depend on the ellipsoid, the earth model, and the definition of sea level.

DBMS: Data Base Management System Software that manages manipulates and retrieves data in a database.

DECISION SUPPORT SYSTEM (DSS): A system, usually computerized, dedicated to supporting decisions regarding a specific problem or set of problems.

DEFAULT: The value of a parameter or a selection provided for the user by the GIS without user modification.

DELAUNAY TRIANGULATION: An optimal partitioning of the space around a set of irregular points into non-overlapping triangles and their edges.

DEM (DIGITAL ELEVATION MODEL): A raster data format for digital topography, containing an array of terrain elevation measurements.

DESIGN LOOP: The iterative process in which a GIS map is created, examined for design, improved, and then replotted from the modified map definition until it is satisfied that a good design has been reached.

DESKTOP MAPPING: The ability to generate easily a variety of map types, symbolization methods, and displays by manipulating the cartographic elements directly.

DGPS: Differential Global Positioning System. A positioning procedure that uses two receivers, a rover at an unknown location and a base station at a known, fixed location. The base station computes corrections based on the differences between its actual and observed ranges to the satellites being tracked.

DIFFERENCE OF MEANS: A statistical test to determine whether or not two samples differ from each other statistically.

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

DIMENSIONALITY: The property of geographic features by which they are capable of being broken down into elements made up of points, lines, and areas. This corresponds to features being zero-, one-, and two-dimensional. A well is a point, a stream is a line, and a forest is an area, for example.

DISSOLVE: Eliminating a boundary formed by the edge or boundary of a feature that becomes unnecessary after data have been captured: for example, the edges of sheet maps.

DISTANCE DECAY: A function that represents the way that some entity or its influence decays with distance from its geographical location. Douglas-Peucker algorithm: A geometric algorithm used to thin out the number of points needed to represent the overall shape of a line feature.

DISTORTION: The space distortion of a map projection, consisting of warping of direction, area, and scale across the extent of the map.

DISTRIBUTED NETWORK: A network-connected set of locations, each storing one element of a system. A distributed GIS may have the GIS software running on a workstation but use data dispersed at many computer storage locations over a local or wide area network.

DLG: Digital Line Graph files from the U.S. Geological Survey.

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

DUEKER'S DEFINITION (OF GIS): 'A special case of information systems where the database consists of observations on spatially distributed features, activities or events, which are definable in space as points, lines, or areas. A geographic information system manipulates data about these points, lines, and areas to retrieve data for ad hoc queries and analyses.'

DXF: Data Exchange Format. A format for storing vector data in ASCII or binary files; used by AutoCad or other CAD software and convertible to ARC/INFO coverages.

DYNAMIC MODEL: A model in which time is the key variable while all other input variables remain constant. Outputs from the model vary as time progresses.

DYNAMIC SEGMENTATION: GIS function that breaks a line into points at locations that have significance, and that can have their own attributes. For example, the representing a highway can have a new node added every 10 kilo-meters as a distance marker that can hold attributes about the traffic flow at that place.

EDGE MATCHING: The GIS or digital map equivalent of matching paper maps along their edges. Features that continue over the edge must be "zipped" together and the edge dissolved. To edge-match, maps must be on the same projection, datum, ellipsoid, and scales and show features captured at the same equivalent scale.

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

END NODE: The last point in an arc that connects to another arc.

EPSILON MODELLING: A method of estimating the effects of positional error in GIS overlay operations. Epsilon modelling is based on the use of buffer zones to account for digitizing error around point, line and area features.

ERROR PROPAGATION: The generation of errors in a GIS database at various stages of the data stream and during subsequent analyses.

EXPORT: The capability of a GIS to write data out into an external file and into a non-native format for use outside the GIS, or in another GIS.

FAT LINE: Raster representation of a line that is more than one pixel wide.

FEATURE CLASS: The type of feature represented in a coverage. Coverage feature classes include arcs, nodes, label points, polygons, tics, annotation, links, boundaries, routes, and sections.

FEATURE: A single entity that composes part of a landscape.

FILE HEADER: The first part of a file that contains metadata rather than data.

FILE: Data logically stored together at one location on the storage mechanism of a computer.

FIPS 173: The federal information processing standard maintained by the USGS and the National Institute of Standards and Technology, which specified a organization and mechanism for the transfer of GIS data between dissimilar computer systems. FIPS 173 specifies terminology, features types, and accuracy specifications, as well as a formal file transfer method.

FIX: A solution to a software problem or bug. Usually, a section of a computer program or a file to be overwritten to correct the problem, called a patch.

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

FLOW MAP: A linear network map that shows, usually by proportionally varying the width of the lines in the network, the amount of traffic or flow within the network.

FORMAT: The specific organization of a digital record.

FORTRAN: An early computer programming language, initially for converting mathematical formulas into computer instructions.

FORWARD/REVERSE LEFT: Moving along an arc, the identifier for the arc connected in the direction/opposite direction of the arc to the immediate left.

FOURTH DIMENSION: A common way of referring to time; the first three dimensions determine location in space, the fourth dimension determines creation, duration, and destruction in time.

FULLY CONNECTED: A set of arcs in which forward and reverse linkages have identically matching begin and end nodes.

FUNCTIONAL CAPABILITY: One of the distinctive processes that a GIS is able to perform as a separate operation or as part of another operation.

FUZZY TOLERANCE: Linear distance within which points should be snapped together.

GANTT CHARTS: Graphical time charts used to assist project management.

GBF (GEOGRAPHIC BASE FILE): A database of DIME records.

GENERALIZATION: The process of moving from one map scale to a smaller (less detailed) scale, changing the form of features by simplification, and so on.

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

GEOGRAPHIC INFORMATION SCIENCE: Research on the generic issues that surround the use of GIS technology, impede its implementation, or emerge from an understanding of its capabilities.

GEOGRAPHIC SEARCH: A find operation in a GIS that uses spatial properties as its basis.

GEOMETRIC TEST: A test to establish the spatial relationship between features. For example, a point feature can be given a point-in-polygon test to find if it is "contained" by an area.

GNU: Free Software Foundation organization that distributes software over the Internet.

GOODNESS OF FIT: The statistical resemblance of real data to a model, expressed as strength or degree of fit of the model.

GPS (GLOBAL POSITIONING SYSTEM): An operational U.S. Air Force-funded system of satellites in orbits that allow their use by a receiver to decode time signals and convert the signals from several satellites to a position on the earth's surface.

GUI (GRAPHICAL USER INTERFACE): The set of visual and mechanical tools (such as window, icons, menus, and toolbars, plus a pointing device such as a mouse) through which a user interacts with a computer.

HARD SYSTEMS ANALYSIS: A set of theory and methods for modelling the complexity of the real world.

HIERARCHICAL DATA MODEL: An attribute data model based on sets of fully enclosed subsets and many layers.

HYPERTEXT: Textual information in which direct links can be made between related text through "hot links," where pointing to a highlighted term moves the user to the text context for that term in the same or a different document.

HYPOTHESIS: A supposition about data expressed in a manner to make it subject to statistical test.

IDEAL POINT ANALYSIS: An MCE algorithm based on the evaluation of choice alternatives against a hypothetical ideal solution.

IDENTIFY: To find a spatial feature by pointing to it interactively on the map with a pointing device such as a mouse.

IDENTITY OVERLAY: Polygon-on-polygon overlay corresponding to the Boolean OR and AND overlays, The output map will contain all those polygons from the first map layer and those which fall within these from the second map layer.

IMAGE MAP: A map that in two dimensions shares many of the characteristics of a map, that is, cartographic geometry, some symbols, a scale and projection, and so on, but is a continuous image taken from an air photo, a satellite image, or a scanned paper map used as a backdrop in a GIS becomes an image map.

IMPORT: The capability of a GIS to bring data in an external file and in a non-native format for use within the GIS.

INDEPENDENT VARIABLE: A variable on the right-hand side of the equation in a model, whose value can range independently of the other constants and variables.

INFORMATION SYSTEM: A system designed to allow the user to be delivered the answer to a query from a database.

INTEROPERABILITY: The extent to which users, software, and data can move between computer environments without change or retraining. In a fully interoperable GIS, the user interface will look and feel the same in two different environments (say, a microcomputer and a Unix workstation), and the same set of functions will have same effect on the same data.

INTERSECT: The topological integration of two spatial data sets that preserves features that fall within the spatial extent common to both input data sets.

INTERVAL: Data measured on a relative scale but with numerical values based on an arbitrary origin. Examples are elevations based on mean sea level, or coordinates.

IN-VEHICLE NAVIGATION SYSTEM: A navigation aid allowing the driver of a car, pilot of a plane, or navigator of a boat direct assistance during operation. Combinations of GPS, on-board digital maps, GIS functions such as routing, and voice information are common in these systems. Those using the sensed motion of the vehicle are called inertial.

ITEM: In an attribute table, a field of information commonly displayed as a column. A single attribute from a record in an INFO data file.

KEY ATTRIBUTE: A unique identifier for related records that can serve as a common thread throughout the files in a relational database.

KILLER APP: A computer program or "application" that by providing a superior method for accomplishing a task in a new way becomes indispensable to computer users. Examples are word processors and spreadsheets.

LABEL POINT: A point digitized within a polygon and assigned its label or identifier for use in topological reconstruction of the polygon.

LABEL: Any text cartographic element that adds information to the symbol for a feature, such as the height number label on a contour line.

LASER LINE FOLLOWER: An automatic digitizer that uses a laser beam to follow and digitize lines on a map.

LATITUDE-LONGITUDE: A spherical reference system used to measure locations on surface. Latitude measures angles in the north south direction and longitude measures angles in the east west direction.

LAYER: A logical set of thematic data described and stored in a map library. Layers organize a map library by subject matter, *e.g.,* soils, roads, wells, and extend over the entire geographic area defined by the spatial index of the map library.

LAYER-BASED APPROACH: An approach to organizing spatial data into thematic map layers, wherein each map layer contains information about a particular subject and is stored as a separate file (or series of files) for ease of management and use.

LEVEL OF MEASUREMENT: The degree of subjectivity associated with a measurement. Measurements can be nominal, ordinal, interval, or ratio.

LINE-IN-POLYGON: A spatial operation in which arcs in one coverage are overlaid with polygons in another to determine which arcs, or portions of arcs, are contained within the polygons. Polygon attributes are associated with corresponding arcs in the resulting line coverage.

LINK: The part or structure of a database that physically connects geographic information with attribute information for the same features. Such a link is a defining component of a GIS.

LINKED DISPLAY: Method of dynamically linking map and non-map output such as charts and data plots such that changes in one are reflected by changes in the other. Such displays are used to aid exploratory data analysis.

LOCATION: A position on the earth's surface or in geographic space definable by coordinates or some other referencing system, such as a street address or space indexing system.

LOGICAL SELECTION: The process of selecting a subset of features from a coverage using logical selection criteria that operate on the attributes of coverage features (*e.g.,* area greater than 16,000 square feet). Only those features whose attributes meet the selection criteria are selected. Also known as feature selection by attribute.

LOGICAL STRUCTURE: The conceptual design used to encrypt data into a physical structure.

MANY-TO-ONE-RELATE: A relate in which many records in one table are related to a single record in another table. A goal in relational database design is to use one to many relates to reduce data storage and redundancy.

MAP ALGEBRA: ***Tomlin's terminology*** for the arithmetic of map combination for coregistered layers with rasters of identical size and resolution.

MAP OVERLAY: Placing multiple thematic maps in precise registration, with the same scale, projections, and extent, so that a compound view is possible.

MASK: A map layer intended to eliminate or exclude areas not needed for mapping and analysis.

MEAN CENTER: For a set of points, that point whose coordinates are the means of those for the set.

MEDIAN: The attribute value for the middle record in a data set sorted by that attribute.

METADATA: Data about data, usually for search and reference purposes. Index-type information pertaining to the entire data set rather than the objects within the data set. Metadata usually includes the date, source, map projection, scale, resolution, accuracy, and reliability of the information, as well as data about the format and structure of the data set.

MISSING DATA: Elements where no data is available for a feature or a record.

MIXED PIXEL: A pixel containing multiple attributes for a single ground extent of a grid cell. Common along the edges of features or where features are ill defined.

MODEL: A theoretical distribution for a relationship between attributes. A spatial model is an expected geographic distribution determined by a given form such as an equation.

MODELLING: The stage in science when a phenomenon under test is sufficiently understood that an abstract system can be built to simulate the real system.

MOSAICING: The GIS or digital map equivalent of matching multiple paper maps along their edges. Features that continue over the edge must be "zipped" together and the edge dissolved. A new geographic extent for the map usually has to be cut or clipped out of the mosaic. To permit mosaicing, maps must be on the same projection, datum, ellipsoid, and scale, and show features captured at the same equivalent scale.

NETWORK MAP: A map that shows as its theme primarily connections within a network, such as roads, subway lines, pipelines, or airport connections.

NODE SNAP: Instructing the GIS software to make multiple nodes or points in a single node so that the features connected to the nodes match precisely, say, at a boundary.

NODE: The end of an arc. At first, any significant point in a map data structure. Later, only those points with topological significance, such as the ends of lines.

NOMINAL: A level of measurement at which only subjective information is available about a feature. For a point, for example, the name of the place.

NORMALIZE: To remove an effect biasing a statistic, for example, the influence of the size of the sample.

OBJECT-ORIENTED: Computer programming languages and databases that support "objects." Objects are standard "classes" that contain all the properties of an object. As a simple example, an object class could be a point and will contain the latitude and longitude of the point, a feature code for the point, such as "radar beacon", and any necessary text to describe the object.

OCTREE: Three-dimensional modification of the quad tree data structure.

ONE-TO-MANY-RELATE: A type of relate connecting a unique value in one file to many records (that have the same value) in another file.

OPEN/GIS: An active effort to assure interoperability among GIS software packages by specifying a standard set of functions and a common user interface.

OPTIMIZATION MODEL: A model that is constructed to maximize or minimize some aspect of its output.

ORDINAL: A level of measurement at which only relative information is available about a feature, such as a ranking. For a highway, for example, the line is coded to show a Jeep trail, a dirt road, a paved road, a state highway, or an interstate highway, in ascending rank.

OVERLAY WEIGHTING: Any system for map overlay in which the separate thematic map layers are assigned unequal importance.

PAT: (Point Attribute Table, Polygon Attribute Table) A coverage can have either a point attribute table or a polygon attribute table, but not both.

PATCH: A fix to a program or data set involving a sequence of data that are to be overwritten onto an older version.

PCMCIA: A credit-card-like device interface for microcomputers and other devices, such as GPS receivers, that meets the standards of the Personal Computer Memory Card International Association. PCMCIA cards can act as memory, connectors to disk drives, and links to other types of devices, perform many other functions, and are interoperable across computers.

POINT-IN-POLYGON: A spatial operation in which points from one coverage are overlaid with a polygonal coverage to determine which points fall within the polygon boundaries. Points assume the attributes of the polygons within which they fall.

POLYGON: A multisided figure that represents area on a map. A feature defined by the arcs that make up its boundary. Every polygon contains one label point within its boundary. Polygons have attributes that describe the geographic feature they represent.

POLYGON INTERIOR: The space contained by a ring, considered part of a polygon.

POLYGON LEFT: Moving along an arc, the identifier for the polygon adjacent to the left.

POLYGON OVERLAY: A process that merges spatially coincident polygons from two coverages, and their attributes, to create a third coverage, that contains new polygons and describes new relationships.

POLYGON RIGHT: Moving along an arc, the identifier for the polygon adjacent to the right.

QUAD TREE: A way of compressing raster data based on eliminating redundancy for attributes within quadrants of a grid.

QUADRANGLE (QUAD): Typically refers to a map sheet published by the U.S. Geological Survey, a 7.5 minute quadrangle series or the 15 minute quadrangle series. Also known as a topographic or topo map.

QUERY LANGUAGE: The part of a DBMS that allows the user to submit queries to a database.

QUERY: A question, especially if asked of a database by the user via a database management system or GIS.

RASTER: Data displayed as discrete picture elements (pixels).

RECORD: A set of values for all attributes in a database. Equivalent to a row of a data table.

REFERENCE MAP: A highly generalized map type designed to show general spatial properties of features. Examples are world maps, road maps, atlas maps, and sketch maps. Sometimes used in navigation, often with a limited set of symbols and few data. A cartographic base reference map is often the base layer or framework in a GIS.

RELATE: An operation that establishes a temporary connection between corresponding records in two tables using an item common to both. A relate gives access to additional feature attributes that are not stored in a single table.

RELATIONAL DBMS: A database management system based on the relational data model.

RELATIONAL MODEL: A data model based on multiple flat files for records, with dissimilar attribute structures, connected by a common key attribute.

RELATIVE LOCATION: A position described solely with reference to another location.

RETRIEVAL: The ability of a database management system or GIS to get back from computer memory records that were stored there previously.

RGB: The system of specifying colours by their red, green, and blue saturations.

R-SQUARED: A common term for the coefficient of determination.

RUN-LENGTH ENCODING: A way of compressing raster data based on eliminating redundancy for attributes along rows of a grid.

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

SIEVE MAPPING: The consecutive overlay of various maps to find a set of feasible areas that satisfy a given set of criteria.

SIFT: To eliminate features that are smaller than a minimum feature size.

SLIVER: Very small and narrow polygon caused by data capture or overlay error that does not exist on the map.

SNAP: Forcing two or more points within a given radius of each other to be the same point, often by averaging their coordinate.

SOFT SYSTEMS ANALYSIS: A general-purpose methodology for investigating unstructured management problems.

SORT: To place the records within an attribute in sequence according to their value.

SPATIAL ANALYSIS: The process of modelling, examining, and interpreting model results. Spatial analysis is the process of extracting or creating new information about a set of geographic features. Spatial analysis is useful for evaluating suitability and capability, for estimating and predicting, and for interpreting and understanding. In GIS there are four traditional types of spatial analysis: spatial overlay and contiguity analysis, surface analysis, linear analysis, and raster analysis.

SPATIAL DATA TRANSFER STANDARD (SDTS): The formal standard specifying the organization and mechanism for the transfer of GIS data between dissimilar computer systems. SDTS specifies terminology, feature types, and accuracy specifications as well as a formal file transfer method for any generic geographic data. Subsets for the standard for specific types of data, vector, and raster, for example, are called profiles.

SPATIAL DATA: Data that can be linked to locations in geographic space, usually via features on a map.

SPATIAL DISTRIBUTION: The locations of features or measurements observed in geographic space.

SPATIAL INTERACTION MODELS: Models that are used to help understand and predict the location of activities and the movement of materials, people and information.

SPATIAL MODELLING: Analytical procedures applied with GIS. There are three categories of spatial modelling functions that can be applied to geographic data within a GIS: geometric models, such as calculating the distance between features, generating buffers, calculating areas and perimeters, and so on; coincidence modelling, such as polygon overlay; and adjacency modelling such as redistricting and allocation.

SQL (Structured Query Language)*:* A syntax for defining and manipulating data from a relational database. Developed by IBM in the 1970s, it has become an industry standard for query languages in most relational database management systems.

STOCHASTIC MODEL: A model that recognizes that there could be a range of possible outcomes for a given set of inputs, and expresses the likelihood of each one happening as a probability.

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

SUBSETTING: Extracting a part of a data set.

SURFACE DRAPE: The draping of an image on top of a 3D view of a terrain model for the purpose of landscape rendering or visualization.

SURFACE SIGNIFICANT POINTS: Points in a TIN model that cannot be closely interpolated from the height values of neighbouring points.

SYMAP: An early multipurpose computer mapping package.

TEMPORARY MAP: A map designed for use as an intermediate product in the GIS process and not usually subjected to the normal map design sequence.

TIGHT COUPLING: A method of linking models to GIS in which the link between the GIS and the model is hidden from the user by an application interface and GIS and model share the same database.

TIN (TRIANGULATED IRREGULAR NETWORK): A series of triangles constructed using elevation data points taken from coverages. These triangles are used for surface representation and display.

TOLERANCE: The distance within which features are assumed to be erroneously located different versions of the same thing.

TOPOLOGICALLY CLEAN: The status of a digital vector map when all arcs that should be connected are connected at nodes with identical coordinates and the polygons formed by connected arcs have no duplicate, disconnected, or missing arcs.

TOPOLOGY: The numerical description of the relationships between geographic features, as encoded by adjacency, linkage, inclusion, or proximity. Thus a point can be inside a region, a line can connect to others, and a region can have neighbors.

TRANSFORMATION: The process that converts coordinates from one coordinate system to another through translation, rotation, and scaling.

TRIANGULATION: A method of surveying in the location of an object may be calculated from the known locations of two other objects. Creating a triangle from the three items, the angles and sides of the triangle can be measured and the location of the unknown object is calculated algebraically.

UPDATE: Any replacement of all or part of a data set with new or corrected data.

UTM (UNIVERSAL TRANSVERSE MERCATOR): A standardized coordinate system based on the metric system and a division of the earth into sixty 6-degree-wide zones. Each zone is projected onto a transverse Mercator projection, and the coordinate origins are located systematically. Both civilian and military versions exist.

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

VALUE: The content of an attribute for a single record within a database. Values can be text, numerical, or codes.

VARIANCE: The total amount of disagreement between numbers. Variance is the sum of all values with their means subtracted and then squared, divided by the number of values less one.

VECTOR: A map data structure using the point or node and the connecting segment as the basic building block for representing geographic features.

VERIFICATION: A procedure for checking the values of attributes for all records in a database against their correct values.

References

Ackoff, R.L., (1971). Towards a System of Systems Concepts. Management Science 17 (11): 661–71.

Armstrong, M.P. and P.J. Densham, (1990). Database organization alternatives for spatial Decision Support Systems, International Journal of Geographical Information Systems, Vol. 3(1).

Aronoff, S. (1989). Geographic Information Systems: A Management Perspective. WDL Publications, Ottawa.

Arthur, C. (1992). Ambulance Computer System was too Complicated. New Scientist 136 (1847): 7.

Atkinson, M., Bancilhon, F., DeWitt, D., Dittrich, K., Maier, D., and Zdonik, S. (1989). The Object-oriented Database System Manifesto. In Proceedings DOOD 89, Kyoto, December.

Avison, D.E. and Fitzgerald, G., (1988). Information Systems Development: Methodologies, Techniques and Tools. Blackwell Scientific, Oxford.

Avison, D.E. and Wood-Harper, A.T., (1991). Information Systems Development Research: An Exploration of Ideas in Practice. Computer Journal 34 (2): 98–112.

Barnard, M.E., (1992). The Global Positioning System. IEE Review (March) 99–102.

Barr, R. (1995). Addressing the Issues. GIS Europe 4 (9): 18–19.

Barr, R. (1995). Mind the Generation Gap. GIS Europe 4 (7): 18–19.

Barr, R. (1996). Data, Information and Knowledge in GIS. GIS Europe 5 (3):14–15.

Batty, M. and Xie, Y., (1994). Modelling Inside GIS: Part 1. Model Structures, Exploratory Spatial Data Analysis and Aggregation. International Journal of Geographical Information Systems 8 (4): 291–307.

Batty, M. and Xie, Y., (1994). Modelling Inside GIS: Part 2. Selecting and Calibrating Urban Models Using ARCINFO. International Journal of Geographical Information Systems 8 (5): 451–70.

Batty, P. (1990). Exploiting Relational Database Technology in GIS. Mapping Awareness 4 (6): 25–32.

Beale, H., (1987). The Assessment of Potentially Suitable Repository Sites. In: The Management and Disposal of Intermediate and Low Level Radioactive Waste. Mechanical Engineering Publications, London, pp. 11–18.

Beard, M.V. and N.R. Chrisman, (1988). Zipping: A Locational Approach to Edgematching, The American Cartographer 15:163–72.

Bemhardsen, T., (1992). Geographic Information Systems. VIAK IT, Norway.

Bemhardsen, T., (1999). Choosing a GIS. In: Longley PA, Goodchild M. F., Maguire D. J. and Rhind D.W. (eds) Geographical Information Systems. Wiley, New York, 589–600.

Berry, B.J.L. and A.M. Baker, (1968). Geographic sampling. In B.J.L. Berry and D.F. Marble, editors, Spatial Analysis. Prentice Hall, Englewood Cliffs NJ, 91–100.

Berry, J. K., (1993). Beyond Mapping: Concepts, Algorithms and Issues in GIS. GIS World Inc., Colorado.

Berry, J.K., (1987). Fundamental Operations in Computer– assisted Map Analysis, International Journal of Geographical Information Systems 1:119–136.

Birkin, M., Clarke, G., Clarke, M. and Wilson, A., (1996). Intelligent GIS: Location Decisions and Strategic Planning. GeoInformation International, Cambridge, UK.

Birkin, M., Clarke, G., Clarke, M. and Wilson, A., (1996). Intelligent GIS: Location Decisions and Strategic Planning. GeoInformation International, Cambridge, UK.

Blakeley, J.A. (1991). DARPA Open Object-oriented Database Preliminary Module Specification: Object Query Module. Texas Instruments, Inc., Version 3, Nov. 25.

Blakemore, M. J., (1984). Generalization and Error in Spatial Databases. Cartographica 21; 131–9.

Bonham-Carter, G. F., (1995). Geographic Information Systems for Geoscientists: Modelling with GIS. Pergamon Press, New York.

Booch, G. (1994). Object-oriented Analysis and Design with Applications, 2nd ed. Benjamin/ Cummings Publishing Company.

Boolding, K. E., (1956). On Systems Theory and Analysis in Urban Planning. Papers in Planning 11. UWIST, Cardiff.

Brassel, K. E., (1977). A Survey of Cartographic Display Software, International Yearbook of Cartography, vol.17, pp. 60–76.

Brown, C., (1989). Implementing a GIS: Common Elements of Successful Sites. Presented at Conference of URISA, Boston.

Brusegard, D. and Menger, G. (1989) Real data and real problems: dealing with large spatial databases. In; Goodchild M. F., Gopal, S. (eds) The Accuracy of Spatial Databases. Taylor and Francis, London, pp. 177–86.

Burrough, P. A. and McDonnell, R., (1998). Principles of Geographical Information Systems. Oxford University Press, Oxford.

Burrough, P. A., (1986). Principles of Geographical Information Systems for Land Resources Assessment. Clarendon Press, Oxford.

Buttenfield, B. P. and Mackaness, W. A., (1991). Visualization. In: Maguire D.J., Goodchild M.F., Rhind D.W. (eds) Geographical Information Systems: Principles and Applications Longman, London, Vol. 1, pp. 427-43.

Bylinsky, Gene, (1989). Managing with Electronic Maps, Fortune April 24, 237–254.

Campbell, John (1984). Introductory Cartography. Prentice Hall, New Jersey.

Caron, L. and J. Merchant, (1986). Geographic Information Systems for Non-urban Local-Level Jurisdictions: Existing Alternatives, p. 110. Cartographica 21 (4): 66–113.

Carver, S. and Brunsdon, C. (1994). Vector to Raster Conversion error and Feature Complexity: An Empirical Study Using Simulated Data. International Journal of Geographical Information Systems 8 (3): 261–72.

Carver, S. J., (ed.) (1998). Innovations in GIS 5. Taylor and Francis, London.

Cassettari, S., (1993). Introduction to Integrated Geo-information Management. Chapman & Hall, London.

Checkland, P. B., (1981). Systems Thinking Systems Practice. John Wiley, Chichester.

Chen, P.P. (1976). The Entity-Relationship Model: Toward a Unified View of Data. ACM TODS, 1:1.

Chrisman, N. R., (1997). Exploring Geographic Information Systems. Wiley, New York.

Chrisman, N.R., 1978. “Efficient Digitizing through the Combination of Appropriate Hardware and Software for Error Detection and Editing,” International Journal of Geographical Information Systems 1:265–77.

Chrisman, N.R., 1990. “Deficiencies of Sheets and Tiles: Building Sheetless Databases,” International Journal of Geographical Information Systems 4:157–67. ESRI, 1990. Understanding GIS: The ARC/INFO Way, ESRI, Redlands, CA.

Churchman, C. W., (1968). The Systems Approach Dell, New York.

Clarke, A. L. (1991). GIS Specification, Evaluation and Implementation. In: Maguire D. J., Goodchild M.F., Rhind D.W. (eds). Geographical Information Systems: Principles and Applications. Longman, London, Vol. 1, pp. 477–88.

Clarke, K. C., (2003). Getting Started With Geographic Information Systems, Prentice Hall Publication.

Clarke, K.C. (1990). Analytical and Computer Cartography. Prentice-Hall, Englewood Cliffs, New Jersey.

Codd, E.F. (1970). A Relational Model of Data for Large Shared Data Banks. Comm. ACM 13:6.

Codd, E.F. (1971). ALPHA: A Data Base Sublanguage Founded on the Relational Calculus. In Proc. 1971 ACM SIGFIDET Workshop (San Diego, Nov. pp. 11–12).

Codd, E.F. (1988). Fatal Flaws in SQL (Both the IBM and the ANSI Versions). Datamation, August and September.

Codd, E.F. (1990). The Relational Model for Database Management, Version 2. Addison-Wesley Publishing Company, New York.

Coppock, J. T. and Rhind, D. W., (1991). The History of GIS. In: Maguire D. J., Goodchild M.F., Rhind D.W. (eds). Geographical Information Systems: Principles and Applications. Longman, London, Vol. 1, pp. 21–43.

Cowen, D.J., (1988). GIS versus CAD versus DBMS: What are the Differences? Photogrammetric Engineering and Remote Sensing 54:1551–5.

Cusack, E. (1991). Inheritance in Object-oriented Z. In Pierre America (ed.), Lecture Notes in Computer Science, ECOOP 91, European Conference on Object-oriented Programming, Springer-Verlag.

Dangermond, Jack and Morehouse, Scott. (1987). Trends in Hardware for Geographic Information Systems, Proceedings AUTOCARTO 8, ASPRS/ACSM, Falls Church, VA.

Dangermond, Jack, (1987). Trends in Geographic Information Systems Software, Proceedings IGIS: The Research Agenda, NASA, Washington, DC.

Date, C.J. (1987). Where SQL Falls Short. (abridged). Datamation, May 1. Unabridged version, What is Wrong with SQL, available from The Relational Institute, San Jose.

Date, C.J. (1995). An Introduction to Database Systems (6th ed.) Addison-Wesley, Reading, MA.

Davis, D. (1999). GIS for Everyone. ESRI Press, Redlands.

Day, D. L. (1981). Geographic Information Systems: All that Glitters is not Gold. Proceedings, Autocarto IV, vol. 1, pp. 541–545.

Densham, P.J. and G. Rushton, (1988). Decision Support Systems for Locational Planning, in R. Golledge and H. Timmermans, editors, Behavioural Modelling in Geography and Planning. Croom-Helm, London, pp. 56–90.

Dickinson, H.J., and H.W. Calkins, (1988), The Economic Evaluation of Implementing a GIS, International Journal of Geographical Information Systems 2:307–327.

Dobson, J.E., (1983). Automated Geography, Professional Geographer 35:135–43.

Douglas, D. H. and Peucker, T. K. (1973). Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature. Canadian Cartographer 10 (4): 110–22.

Douglas, D. H. and Peucker, T. K. (1973). Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature. Canadian Cartographer 10 (4): 110–22.

Dueker, K. J., (1987). Geographic Information Systems and Computer-Aided Mapping, American Planning Association Journal, Summer 1987:383–390.

Egenhofer, M. and A.U. Frank, (1987). Object-oriented Databases: Database Requirements for GIS, Proceedings International Geographic Information Systems Symposium: The Research Agenda 2:189–211.

Egenhofer, M. J. and Richards, J. R. (1993). Exploratory Access to Geographic Data based on the Map-overlay Metaphor, Journal of Visual Languages and Computing, vol. 4, no.2, pp. 105–125.

Elmasri, R. and Navathe, S.B. (1994). Fundamentals of Database Systems. Addison-Wesley, Menlo Park, CA.

Epstein, E., and T.D. Duchesneau, (1984). The Use and Value of a Geodetic Reference System, University of Maine, Orono, Maine.

Everett, J.E. and D.S. Simonett, (1976). Principles, Concepts and Philosophical Problems in Remote Sensing, in J. Lintz and D.S. Simonett, Editors, Remote Sensing of Environment, Addison-Wesley, Reading, MA, pp. 85–127.

Everst, G.C. (1986). Data Management: Objectives, System Functions and Administration, McGraw-Hill, New York.

Exler, R.D., (1990). Geographic Information Systems Standards: An Industry Perspective, GIS World Vol 3(2):44–47.

Fazal, S. (2000). Urban Expansion and Loss of Agricultural Land. A GIS Based Study of Saharanpur City, India. Environment and Urbanization, Vol.12. No.2, IIED, London, pp. 133– 149.

Fazal, S. (2001). Application of Remote Sensing and GIS Techniques in Urban Sprawl and Land Use Mapping. A Case Study of A Growing Urban Centre in India. Asian Profile, Vol. 29. No.1. Canada. pp. 45– 62.

Fazal, S. (2001). Land Reorganization Study Along Major Roads. Land Use Policy, Elsevier Science Publication, Vol. 18, No.2. United Kingdom. pp. 191–199.

Fazal, S. (2001). The Need For Preserving Our Farm Lands. A Case Study from a Predominantly Agrarian Economy. Landscape and Urban Planning, Vol. 55, No.1, Elsevier Science Publication, United Kingdom. pp. l –13.

Fazal, S. (2004). The Role of Accessibility in Land Transformation: A GIS Based Study of a Growing Urban Centre in a Developing Economy. Asian Profile, Vol. 32. No.1. Canada pp. 61 – 76.

Fazal, S., Rahman, A., (2007). GIS Terminology, New Age International Publishers, New Delhi.

FICCDC, (1988). The Proposed Standard for Digital Cartographic Data, The American Cartographer, Vol. 15(1).

Fisher, P.F., and R. Lindenberg, (1989). On Distinctions Among Cartography, Remote Sensing, and Geographic Information Systems, Photogrammetric Engineering and Remote Sensing 55(10):1431–1434.

Fleet, H., (1986). SAGIS: A Full-function Public-domain GIS for Micro and Minicomputers, p. 301.

French, S., (1986). Decision Theory: An Introduction to the Mathematics of Rationality, Halsted, New York.

Gahegan, M.N. and S. Roberts, (1988). An Intelligent, Object-oriented Geographical Information System, International Journal of Geographical Information Systems Vol. 2: 101-10.

Ghosh, A. and G. Rushton, (1987). Spatial Analysis and Location- Allocation Models, Van Nostrand, Reinhold, New York.

GIS Sourcebook, (1989). GIS World, Fort Collins, CO.

GIS World, (1989). Spatial Data Exchange Formats, The GIS Sourcebook, GIS World, Fort Collins, CO, pp. 122–23.

Golden, B.L. and L. Bodin, (1986). Microcomputer-based Vehicle Routing and Scheduling software, Computers and Operations Research 13:277–85.

Goodchild, M. F., (1991). The Technological Setting of GIS. In: Maguire D.J., Goodchild M. F., Rhind D.W. (eds). Geographical Information Systems: Principles and Applications.

Goodchild, M. F. (1988). Stepping Over the Line the Technological Constraints and the New Cartography. American Cartographer 15 (3):277–89.

Goodchild, M. F. (1990). Spatial Information Science. Proceedings of 4th International Symposium on Spatial Data Handling. International Geographical Union, Columbus Ohio. Vol. 1, pp. 3–12.

Goodchild, M. F. (1991). Geographic Information Systems. Progress in Human Geography 15 (2): 194–200.

Goodchild, M. F. (1995). GIS and Geographic Research. In: Pickles J (ed.). Ground Truth: The Social Implications of Geographic Information Systems. Guilford Press, New York, pp. 31–50.

Goodchild, M. F., (1992). Geographical Information Science. International Journal of Geographical Information Systems 6 (1): 31–45.

Goodchild, M. F., Yang, S., (1989). A Hierarchical Spatial Data Structure for Global Geographic Information Systems. NCGIA Technical Paper 89–5. National Center for Geographic Information and Analysis (NCGIA), University of California, Santa Barbara.

Goodchild, M.F. (1992). Geographic Data Modelling. Computers and Geosciences. Vol. 18, No. 4, pp. 401–408.

Goodchild, M.F. and B.R. Rizzo, (1987). Performance Evaluation and Work-load Estimation for Geographic Information Systems, International Journal of Geographical Information Systems 1:67–76.

Goodchild, M.F. and J.A. Donnan, (1987). Optimum Location of Liquid Waste Disposal Facilities: Formation fluid in the Petrolia, Ontario Oilfield, in M. Chatterji, Editor, Hazardous Materials Disposal: Siting and Management, Gower, Aldershot, UK, pp. 263–73.

Goodchild, M.F., (1984). Geocoding and Geosampling, in G.L. Gaile and C.J. Willmott, eds., Spatial Statistics and Models, Reidel Publishing Company, Dordrecht, Holland, pp. 33–53.

Goodchild, M.F., (1987). Application of a GIS Benchmarking and Workload Estimation Model, Papers and Proceedings of Applied Geography Conferences 10:1–6.

Goodchild, M.F., (1988). A Spatial Analytical Perspective on GIS, International Journal of Geographical Information Systems 1:327–34.

Goodchild, M.F., (1988). Towards An Enumeration and Classification of GIS Functions, Proceedings, IGIS ‘87. NASA, Washington DC 2:67–77.

Goodchild, M.F., (1989). Modeling Error in Objects and Fields, in M.F. Goodchild and S. Gopal, Editors, Accuracy of Spatial Databases, Taylor and Francis, New York, 107–13.

Goodwin, T. (1995). GIS-Getting Started. In: GIS in Business: Discovering the Missing Piece in your Business Strategy. GeoInformation International, Banbury, UK, pp. 144–8.

Green, D. R., Rix, D., Corbin, C., (1997). The AGI Source Book for Geographic Information Systems 1997. Association for Geographic Information, London.

Griffith, D. A., Amrhein, C. G., (1991). Statistical Analysis for Geographers. Prentice-Hall, Englewood Cliffs, New Jersey.

Guptill, S., (1988). A Process for Evaluating GIS, USGS Open File Report 88–105. Smith, D.R., 1982. “Selecting a Turn-key Geographic Information System Using Decision Analysis,” Computers, Environment and Urban Systems 7:335–45.

Haggett, P., Chorley, R. J. (1967). Models, Paradigms and the New Geography. In: Chorley R. J., Hagget P. (eds). Models in Geography. Methuen, London, pp. 19–42.

Healey, R. G. (1991). Database Management Systems. In: Maguire D. J., Goodchild M.F., Rhind D.W. (eds). Geographical Information Systems: Principles and Applications.

Herring, J.R., (1989). A Fully Integrated Geographic Information System, Proceedings, AutoCarto 9:828–37.

Herskovits, A., (1987). Spatial Prepositions in English. Cambridge University Press.

Heywood, D.I. (1997). Beyond Chorley: Current Geographic Information Issues. AGI, London.

Heywood, D.I., Blaschke, T., Carlisle, B. (1997). Integrating Geographic Information Technology and Multimedia; the Educational Opportunities. In: Hodgson S, Rumor M, Harts

J.J. (eds) Proceedings of the Joint European Conference and Exhibition on Geographical Information Systems, Vienna, pp.1281–90.

Heywood, D.I., Oliver, J., Tomlinson, S. (1995). Building an Exploratory Multi-criteria Modelling Environment for Spatial Decision Support. In; Fisher p (ed.) Innovations in GIS 2. Taylor and Francis, London, pp. 127–36.

Heywood, I., (1990). Monitoring for Change: a Canadian Perspective on the Environmental Role for GIS. Mapping Awareness 4 (9); 24–6.

Heywood, I., Carver, S., (1994). Decision Support or Idea Generation: the Role for GIS in Policy Formulation. Proceedings Symposium fur Angewante Geographische Informationsverarbeitung (AGIT '94), Salzburg, Austria, July, pp. 259–66.

Heywood, I., Cornelius, S., Carver, S., (2002). Introduction to Geographic Information Systems, Pearson Publication.

Hopkins, L., (1984). Evaluation of Methods for Exploring Ill- Defined Problems, Environment and Planning B 11:339–48.

House, W.C. (ed.), (1983). Decision Support Systems, Petrocelli, New York.

Huggett, R., (1980). Systems Analysis in Geography. Clarendon Press, Oxford.

Jackson, M., Woodsford, P. A., (1991). GIS Data Capture Hardware and Software, In Maguire, D. J. Goodchild, M. F., Rhind, D. W. (eds) Geographical Information Systems: Principles and Applications. Longman, London pp. 239–249.

Johnson, D.S., and U. Basoglu, The Use of Artificial Intelligence in the Automated Placement of Cartographic Names, pp. 225–230.

Kang-tsung Chang (2002). Introduction to Geographic Information Systems, Tata McGraw Hill Publication.

Keates, J. S., (1982). Understanding Maps. Longman, London.

Keating, T. W. Phillips and K. Ingram, (1987). An Integrated Topologic Database Design for Geographic Information Systems, Photogrammetric Engineering and Remote Sensing Vol. 53.

Keeler, B. K., Smith, J. L., Gregoire, T. G. (1988). Simulating Manual Digitizing Error with Statistical Models. Proceedings of GlS/LIS '88. ACSM, ASPRS, AAG and URISA, San Antonio, Texas, pp. 475–83.

Kellaway, G. P. (1949). Map Projections. Methuen, London.

Kemp, K. K. and Goodchild, M. F. (1991). Developing a Curriculum in Geographic Information Systems: The National Center for Geographic Information and Analysis Core Curriculum Project. Journal of Geography in Higher Education, vol. 15, no.2, pp. 121–132.

Kemp, Z. (ed.) (1997). Innovations in GIS 4. Taylor and Francis, London.

King, J.L. and K.L. Kraemer, (1985). The Dynamics of Computing, Columbia University Press, New York.

Korte, G. B. (1993). The GIS Book, 5th edn. OnWord Press, USA.

Krol, E. (1992). The Whole Internet: Users Guide and Catalog. Sebastopol, CA: O'Reilly and Associates.

Kuipers, B., (1978). "Modeling Spatial Knowledge," Cognitive Science 2:129–53.

Lai, Poh–Chin, (1988). Resource Use in Manual Digitizing. A Case Study of the Patuxent Basin Geographical Information System Database, International Journal of Geographical Information Systems 2(4):329–46.

Lang, L. (1998). Managing Natural Resources with GIS. ESRI Press, Redlands.

Langran, G. (1992). Time in Geographical Information Systems. Taylor and Francis, London.

Laurini, R., Thompson, D. (1992). Fundamentals of Spatial Information Systems. Academic Press, London.

Lillesand, T.M. and Kiefer, R.W., 2004, "Remote Sensing and Image Interpretation", 5th Edition, John Wiley & Sons.

Longley, P.A., Goodchild, M. F., Maguire, D. J., Rhind, D. W. (2001). Geographical Information Systems and Science. Wiley, Chichester.

Longley, P.A., Goodchild, M. F., Maguire, D. J., Rhind, D. W. (eds) (1999). Geographical Information Systems: Principles, Techniques, Management and Applications. John Wiley, New York.

Lucas, H.C., (1975). Why Information Systems Fail. Columbia University Press.

Maguire, D. J. (1989). Computers in Geography. Longman, London.

Maguire, D. J. (1991). An Overview and Definition of GIS. In: Maguire D.J., Goodchild M. F., Rhind D.W. (eds) Geographical Information Systems: Principles and Applications.

Maguire, D. J. Goodchild, M. F., Rhind, D. W. (eds) (1991). Geographical Information Systems: Principles and Applications. Longman, London.

Maguire, D.J., (1989). Computers in Geography, John Wiley and Sons, Inc., New York.

Maguire, D.J., M.F. Worboys and H.M. Hearnshaw, (1990). An Introduction to Object-oriented GIS, Mapping Awareness 4(2):36–39.

Maling, D.H., (1973). Coordinate Systems and Map Projections, George Phillip and Sons Limited, London.

Marble, D. (1997). Rebuilding the Top of the Pyramid: Providing Technical GIS Education to Support Both GIS Development and Geographic Research. Presentation to GIS in Higher Education 97 Conference, 30 Oct-2 Nov, NCGIA and Towson University, Virginia, USA.

Marble, D. F. (ed.) (1980). Computer Software for Spatial Data Handling. Ottawa: International Geographical Union, Commission on Geographical Data Sensing and Processing.

Marble, D.F. and L. Sen, (1986). The Development of Standardized Benchmarks for Spatial Database Systems, in D.F. Marble, Editor, Proceedings of the Second International Symposium on Spatial Data Handling, Seattle, 488–496.

Marble, D.F. et al., (1983). Geographic Information Systems and Remote Sensing, Manual of Remote Sensing. ASPRS/ACSM, Falls Church, VA, 1:923–58.

Martin, D. (1991). Geographical Information Systems and Their Socio-economic Applications. Routledge, London.

Martin, D. (1996). Geographical Information Systems and their Socio-economic Applications 2nd Ed. Routledge, London.

Massam, B.H., (1980). Spatial Search, Pergamon, London.

Mather, P. M. (1991). Computer Applications in Geography. Wiley, New York.

Mather, P. M. (ed.) (1993). Geographical Information Handling- Research and Applications. Wiley, Chichester.

McDonnell, R., Kemp, K., (1998). Informational GIS Dictionary, Geoinformation International, London.

McHarg, I.L., 1969. Design With Nature, Doubleday, New York.

Meyer, B. (1997). Object-oriented Software Construction, 2nd ed. Prentice Hall.

Monmonier, M. (1991). How to Lie with Maps. Chicago: University of Chicago Press.

Mounsey, H. and R.F. Tomlinson, (1988). Building Databases for Global Science, Taylor and Francis, New York.

Muehrcke, Phillip C. and Muehrcke, Juliana O. (1992). Map Use: Reading, Analysis and Interpretation. JP Publications, Madison, USA.

Nyerges, T.L., (1989). Schema Integration Analysis for the Development of GIS Databases, International Journal of Geographical Information Systems 3:152–83.

Olle, T.W. (1978). The Codasyl Approach to Data Base Management. John Wiley & Sons, New York.

Openshaw, S. (1989). Learning to live with Errors in Spatial Databases. In: Goodchild M.F., Gopal S. (eds). The Accuracy of Spatial Databases. Taylor and Francis, London, pp. 263–76.

Openshaw, S. (1990). Spatial Referencing for the User in the 1990s. Mapping Awareness 4 (2): 24.

Parent, P. and R. Church, (1987). Evolution of Geographical Information Systems as Decision Making Tools, Proceedings, GIS '87, pp. 63–71, ASPRS/ACSM, Falls Church, VA.

Parker, D. (ed.) (1996). Innovations in GIS 3. Taylor and Francis, London.

Peterson, M. P. (1995). Interactive and Animated Cartography. Upper Saddle River, NJ: Prentice Hall.

Peuquet, D. J. (1984). A Conceptual Framework and Comparison of Spatial Data Models.

Peuquet, D. J. (1999). Time in GIS and Geographical Databases. In: Longley PA, Goodchild M.F., Maguire D.J., Rhind D.W. (eds) Geographical Information Systems: Principles and Technical Issues. Wiley, New York, Vol. 1, pp. 91–103.

Peuquet, D. J., Marble, D. F. (eds) Introductory Readings in Geographical Information Systems. Taylor and Francis, London, pp. 209–14.

Peuquet, D. J., (1981). An Examination of Techniques for Reformatting Digital Cartographic Data, Part II: The Vector-to-raster Process, Cartographica 18:21–33.

Peuquet, D.J. (1991). Methods for Structuring Digital Cartographic Data in a Personal Computer Environment. In Geographic Information Systems: The Microcomputer and Modern Cartography by Taylor, D.R.F. (ed.), Pergamon Press, Oxford.

Peuquet, D.J., (1984). A Conceptual Framework and Comparison of Spatial Data Models, Cartographica 21(4):66–113.

Pressman, R.S. (1997). Software Engineering: A Practitioner's Approach (4th ed.) McGraw-Hill, New York.

Rao, B.R. (1994). Object-oriented Databases: Technology, Applications, and Products. McGraw-Hill, Inc., New York.

Reeve, C. and J. Smith, (1986). The Varied Viewpoints of GIS Justification: A Survey of Vendors and Users, p. 396.

Reeve, D. E., (1996). Module 4: Attribute Data. UNIGIS Postgraduate Diploma in GIS by Distance Learning. Course Material. Manchester Metropolitan University.

Rhind, D., (1989). Why GIS?, Arc News, Summer 1989, Vol 11(3).

Robinson, A.H., R.D. Sale, J.L. Morrison and P.C. Muechrcke, (1984). Elements of Cartography, 5th edition, Wiley, New York.

Rumbaugh, J., Blaha M., Premerlani, W., Eddy, F., and Lorensen, W. (1991). Object-oriented Modeling and Design. Prentice Hall, New Jersey.

Siddiqui, M. A., (2006). Introduction to Geographic Information Systems, Sharda Pustak Bhawan.

Simmel, S.S. and Godard, I. (1991). The KALA Basket: A Semantic Primitive Unifying Object Transactions, Access Control, Versions and Configurations. In Proceedings of OOPSLA '91: Conference on Object-oriented Programming Systems, Languages and Applications, Nov., pp. 230–246.

Skidmore, S., Wroe, B., (1988). Introducing Systems Analysis, NCC Publication, Manchester, UK.

Snyder, J.P., (1987). Map Projections – A Working Manual, US Geological Survey Professional Paper 1395, US Government Printing Office, Washington.

Somers, R., (1990). Where Do You Place the GIS?, GIS World Vol 3(2):38–39.

Sprague, R.H., (1980). A Framework for the Development of Decision Support Systems, Management Information Sciences Quarterly 4:1–26.

Sprague, R.H., and Carlson, E.D., (1982). Building Effective Decision Support Systems, Prentice-Hall, Englewood Cliffs NJ.

Star, J.L. and J.E. Estes, (1990). Geographic Information Systems: An Introduction, Prentice Hall, Englewood.

Star, J.L. and J.E. Estes, (1990). Geographic Information Systems: An Introduction, Prentice Hall.

Strahler, A.N. and A.H. Strahler, (1987). Modern Physical Geography, 3rd edition, Wiley, New York.

Stroustrup, B. (1997). The C++ Programming Language. Addison-Wesley Publishing Company.

Talmy, L., (1983). How Language Structures Space, in H. Pick and L. Acredolo, Editors, Spatial Orientation: Theory, Research and Application, Plenum Press, New York.

Tobler, W. R. (1976). Spatial Interaction Patterns. Journal of Environmental Systems 6: 271–301.

Tobler, W.R., (1959). Automation and Cartography, Geographical Review 49:526–34. An early perspective and prophesy.

Tom, H., (1990). Geographic Information Systems Standards: A Federal Perspective, GIS World Vol 3(2):47–52.

Tomlin, C. D. (1990). Geographic Information Systems and Cartographic Modelling. Prentice-Hall, Englewood Cliffs, New Jersey.

Tomlin, C. D. (1991). Cartographic Modelling. In:Maguire D.J., Goodchild M.F., Rhind D.W. (eds) Geographical Information Systems: Principles and Applications. Longman, London, Vol. 1, pp. 361–74.

Tomlin, C.D., (1990). Geographic Information Systems and Cartographic Modeling, Prentice-Hall, Englewood.

Tomlinson, R. F. (1988). The Impact of the Transition from Analogue to Digital Cartographic Representation. American Cartographer 15 (3): 249–63.

Tomlinson, R. F. and Boy1e, A. R. (1981). The State of Development of Systems for Handling Natural Resources Inventory Data, Cartographica. Vo1. 18, no.4, pp. 65–95.

Tomlinson, R.F., D.F. Marble and H.W. Calkins, (1976). Computer Handling of Spatial Data. UNESCO Press, Paris.

Ullman, J.D. (1988). Principles of Database and Knowledge-Base Systems, Volumes I and II. Computer Science Press, Inc.

Unwin, D. (1981). Introductory Spatial Analysis. Methuen, London.

Upton, G.J.G. and B. Fingleton, (1985). Spatial Data Analysis by Example, Vol I: Point Pattern and Quantitative Data, Wiley, New York.

Walker, P.A. and D.M. Moore, (1988). SIMPLE: An Inductive Modelling and Mapping System for Spatially Oriented Data, International Journal of Geographical Information Systems 2:347–63.

Wise, S. (1995). Scanning Thematic Maps for Input to Geographic Information Systems, Computers and Geosciences 21 (1): 7–29.

Wise, S., (2002). GIS Basics, Taylor and Francis, London.

Wordboys, M. F. (1995). GIS: A Computing Perspective, Taylor and Francis, London.

Worboys, M.F., H.M. Hearnshaw and D.J. Maguire, (1990). Object-oriented Data Modeling for Spatial Databases. International Journal of Geographical Information Systems.