

INTRODUCTION TO GEOGRAPHIC INFORMATION SYSTEMS (GIS)

1. INTRODUCTION TO GIS

1.1 What is GIS

Our understanding of the world around us has always been limited by our lack of information, as well as our lack of wisdom and knowledge. For things too small to see, we have developed microscopes that can image down to the molecular level. At the other end of the continuum, for things that are (in a very real sense) too large to see, we have geostationary satellites that can take an image of an entire hemisphere. GIS is a means of integrating spatial data acquired at different scales and times, and in different formats.

As a Technical Term, **GIS** stands for **Geography Information System**. The first word of this term, geography may be defined on the basis of its constituent parts: geo and graphy. Geo refers to the Earth, and graphy indicates a process of writing; thus geography (in this literal interpretation) means writing about the Earth.

Like the field of geography, the term **Geographic Information System (GIS)** is hard to define. It represents the integration of many subject areas. Accordingly there is no absolutely agreed upon definition of a GIS (deMers, 1997). A broadly accepted definition of GIS is the one as follows provided by the National Centre of Geographic Information and Analysis (NCGIA), US.

A GIS is a system of hardware, software and procedures to facilitate the management, manipulation, analysis, modeling, representation and display of georeferenced data to solve complex problems regarding planning and management of resources (NCGIA, US 1990)

Another similar definition has been given by USGS (United States Geology Survey): A GIS is a computer system capable of capturing, storing, analyzing, and displaying geographically referenced information; that is, data identified according to location. Practitioners also define a GIS as including the procedures, operating personnel, and spatial data that go into the system.

The conception of GIS can be understood from several viewpoints.

- In terms of technology and application GIS is a tool, a method or a technology to solve complex problems on spatial analysis.
- Scientifically, GIS is an independent subject system developed on the basis of geography, cartography and the computer technology.
- Functionally, GIS can be utilized for spatial data acquisition, saving, expressing, processing, analyzing and outputting, and possesses its own structure to be a complete application system.

Contrast to other systems like a database management system (DBMS), GIS interprets and judges spatial data with its unique processing methods, but not limited only in data management. Therefore, GIS can be used for data development and mining.

A more comprehensive and easy way to define GIS is the one that looks at the disposition, in layers (Figure 1.1), of its data sets. "Group of maps of the same portion of the territory, where a given location has the same coordinates in all the maps included in the system". This way, it is

possible to analyze its thematic and spatial characteristics to obtain a better knowledge of this zone.

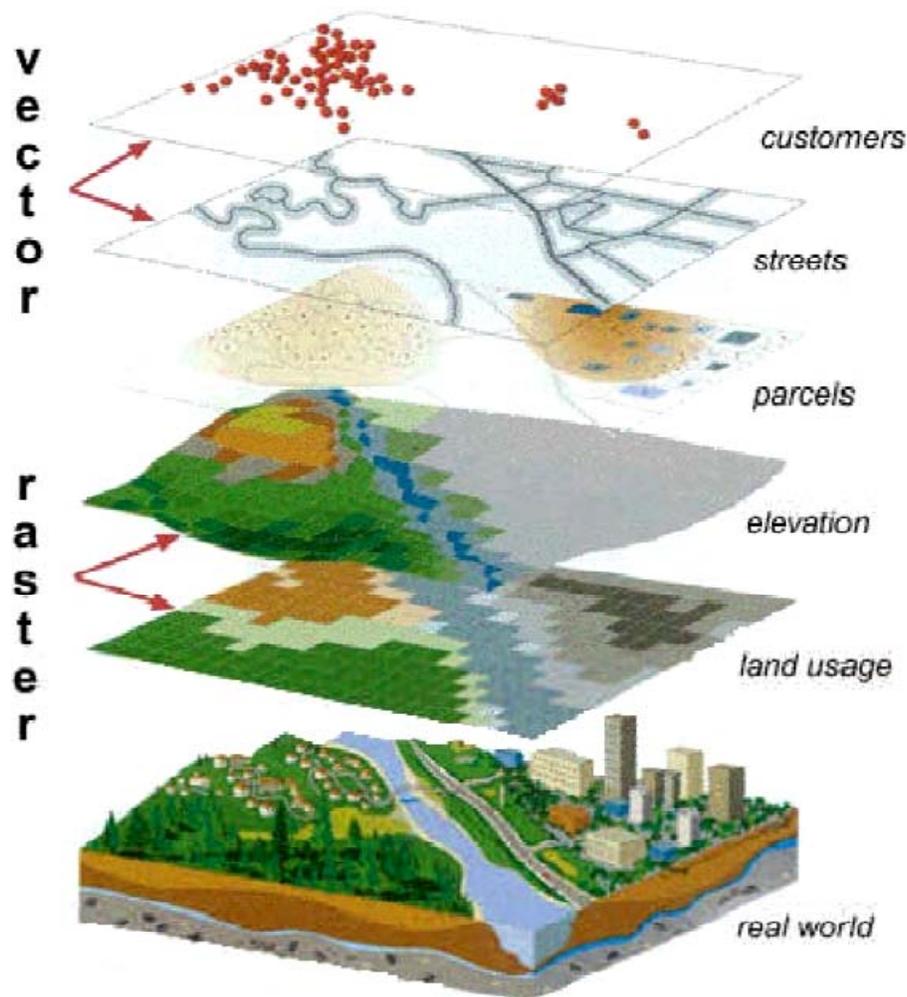


Figure 1.1: The Concept of GIS Layers

1.2 Why GIS

GIS and 4Ms

Basically, urban planners, scientists, resource managers, and others who use geographic information work in several main areas. They observe and **measure** environmental parameters. They develop **maps** which portray characteristics of the earth. They **monitor** changes in our surroundings in space and time. In addition, they **model** alternatives of actions and processes operating in the environment. These, then, are the four Ms: measurement, mapping, monitoring, and modeling. These key activities can be enhanced through the use of information systems technologies, and in particular, through the use of a GIS.

1.2.1 Data Integration

70% or more budgets for a GIS establishment are generally used for establishment of the data base in the GIS. And it is often said that a database is the heart of a GIS. Therefore, the achievement of a GIS establishment is greatly depends on how well all kinds of spatial and non-spatial data would be dealt with in it.

A good GIS can not only capture data from individual data source via various kinds of devices, but also can convert all kinds of spatial data formats to makes sure all data being linked properly so that an integrated information system can be established (Figure 1.2). This data integration is difficult for almost any other means. Thus, a GIS can use combinations of mapped variables to build and analyze new variables.

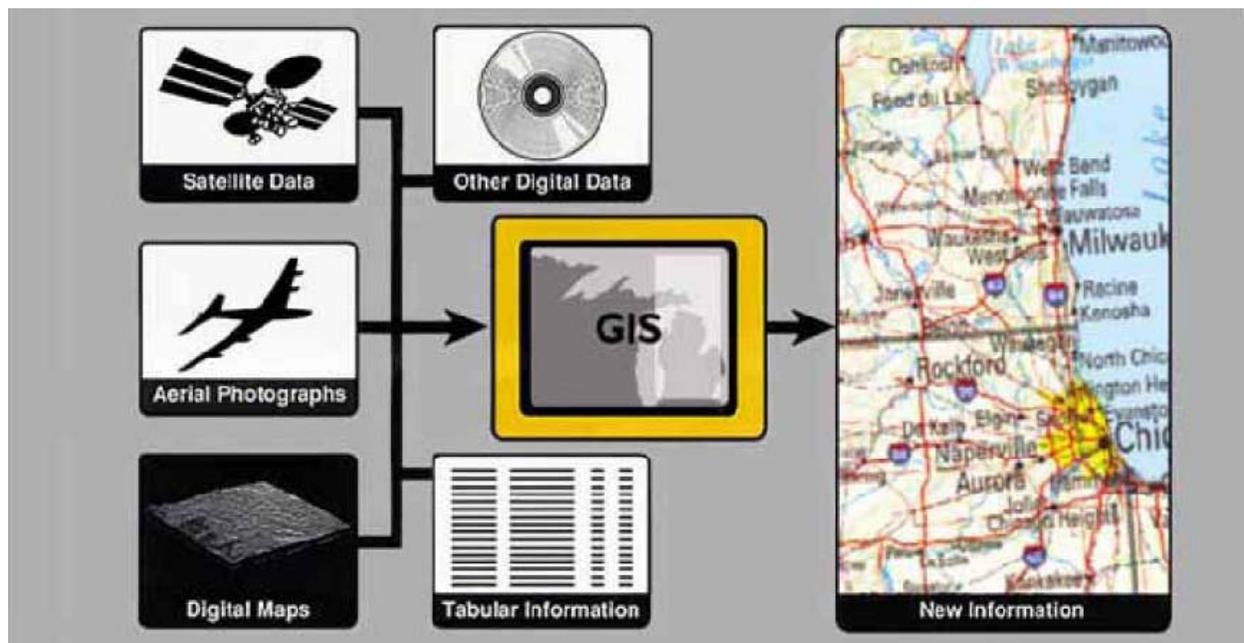


Figure 1.2: Data integration is the linking of information in different forms through a GIS

For example, using GIS technology, it is possible to combine agricultural records with hydrography data to determine which streams will carry certain levels of fertilizer runoff. Agricultural records can indicate how much pesticide has been applied to a parcel of land. By locating these parcels and intersecting them with streams, the GIS can be used to predict the amount of nutrient runoff in each stream. Then as streams converge, the total loads can be calculated downstream where the stream enters a lake.

1.2.2 Information Retrieval

Information retrieval is the most essential function of GIS and can be realized using many methods as by query, by menu, by application interface, by spatial filter, by attribute filter and so on. The retrieval function makes you very easy to answer question as:

- ✦ What is at some location?

- What location is some object?
- What locations satisfy some requirements?
- What is the spatial relationship between objects?
- Where is it?
- What other information can we know about it?

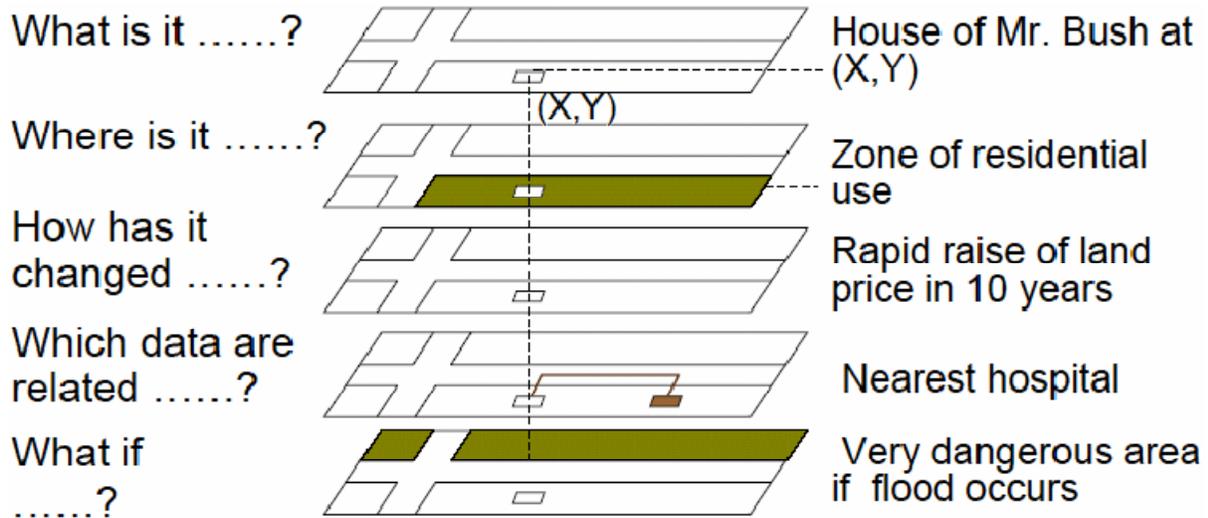


Figure 1.3: Data retrieval function of GIS.

- Location: What is at...?
- Condition: Where is it?
- Trends: What has happened since...?
- Patterns: What spatial patterns exist?
- Modeling: What if...?

1.2.3 Spatial Analyses

“What distinguishes a GIS from other types of information systems are its spatial analysis functions. These functions use the spatial and non-spatial attribute data in the GIS database to answer questions about the real world”.

— Aronoff, 1989, pp 189.

With a GIS, you can link information (attributes) to location data, such as wells to addresses, water supply system and water consumption amount to parcels or villages. You can then layer that information to get a better understanding of how it all works together. You choose what layers to combine based on what questions you need to answer.

A GIS provides a great deal of method for you 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-referenced data, as well a set of operations for working with the data. In a sense, a GIS may be thought of as a higher-order map.

The same as data retrieval function, spatial analyses is one of the most important functions of a GIS and includes many methods for the analysis.

One of the basic analysis methods for polygon data analysis is overlay analysis. And this method (processing) is furthermore divided into some sub-methods as:

Union: to combine features of an input polygon1 with the overlay polygon2 to produce a new output polygon3 that contains the attributes and full extent of both polygon1 and polygong2.

Intersect: to cut an input polygon1 with the feature from an overlay polygon2 to produce a new output polygon3 with features that attribute data from both polygong1 and polygong2.

Clip: use a clip polygon1 like a cookie cutter on your input polygong2 to produce a new polygong3 with the attributes being kept.

.....

Using overlay analysis, you can generate some kinds of new map as shown in figure 1.4 (left) from a topography map and a soil map to generate a topography-soil map, or conduct a soil distribution statistics based on administrative area as shown in figure 1.4 (right).

Other types of spatial data overlay are overlay of point data and line data or area data, and line data and area data. All these overlaying methods can help you analyze the relationship between overlaid data.

In many cases, it is necessary to calculate length of lines or area of polygon. To perform these calculations is also essential function of GIS.

It is impossible to collect data over every square meter of the Earth's surface. Therefore, samples must be taken at discrete locations. A GIS can be used to depict two- and three-dimensional characteristics of the Earth's surface, subsurface, and atmosphere from points where samples have been collected.

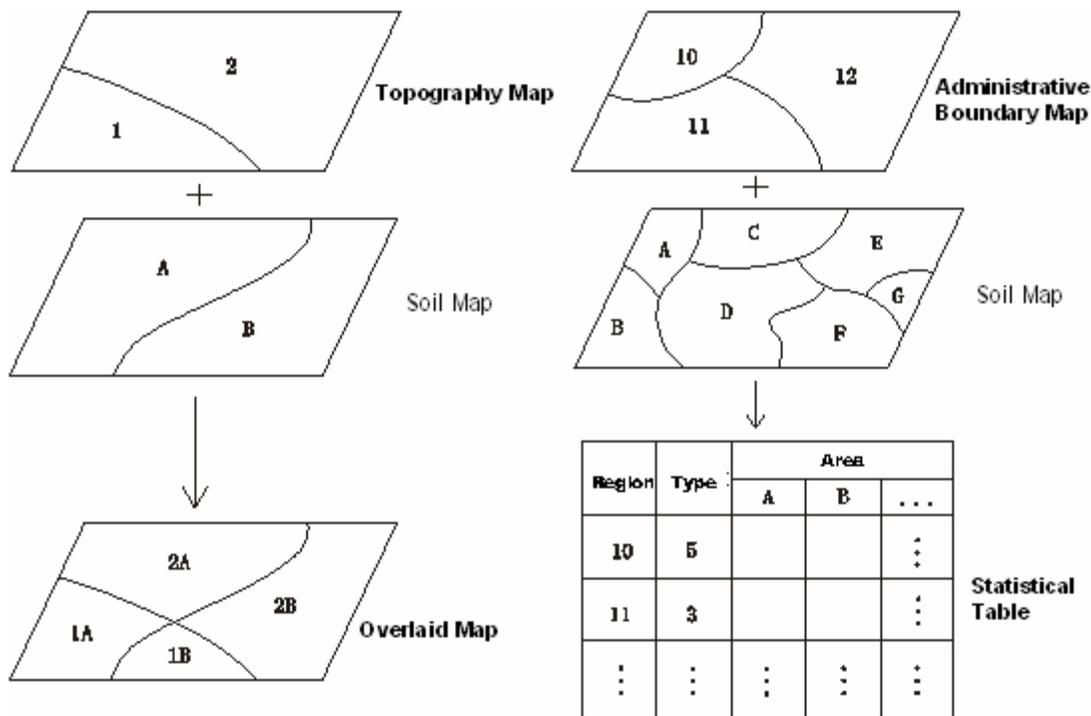


Figure 1.4: Using overlay process to create new map and conduct new statistics.

1.2.4 Map Creation

A GIS is most often associated with maps. A map, however, is only one way you can work with geographic data in a GIS, and only one type of product generated by a GIS. This is important, because it means that a GIS can provide a great deal more problem-solving capabilities than using a simple mapping program or adding data to an online mapping tool.

The function of an information system is to improve one's ability to understand the world around us and make decisions. An information system is that chain of operations that takes us from planning the observation and collection of data, to storage and analysis of the data, to the use of the derived information in some decision-making process (Calkins and Tomlinson, 1977). This brings us to an important concept: a map is a kind of information system. A map is a collection of stored, analyzed data, and information derived from this collection is used in making decisions. To be useful, a map must be able to convey information in a clear, unambiguous fashion, to its intended users.

Mapping locations: GIS can be used to map locations. GIS allows the creation of maps through automated mapping, data capture, and surveying analysis tools.

Mapping quantities: People map quantities, like where the most and least are, to find places that meet their criteria and take action, or to see the relationships between places. This gives an additional level of information beyond simply mapping the locations of features.

Mapping densities: While you can see concentrations by simply mapping the locations of features, in areas with many features it may be difficult to see which areas have a higher concentration than others. A density map lets you measure the number of features using a uniform aerial unit, such as acres or square miles, so you can clearly see the distribution.

Mapping and monitoring change: GIS can be used to map the change in an area to anticipate future conditions, decide on a course of action, or to evaluate the results of an action or policy.

A critical component of a GIS is its ability to produce graphics on the screen or on paper to convey the results of analyses to the people who make decisions about resources. Wall maps, Internet-ready maps, interactive maps, and other graphics can be generated, allowing the decision makers to visualize and thereby understand the results of analyses or simulations of potential events (figure 1.5).

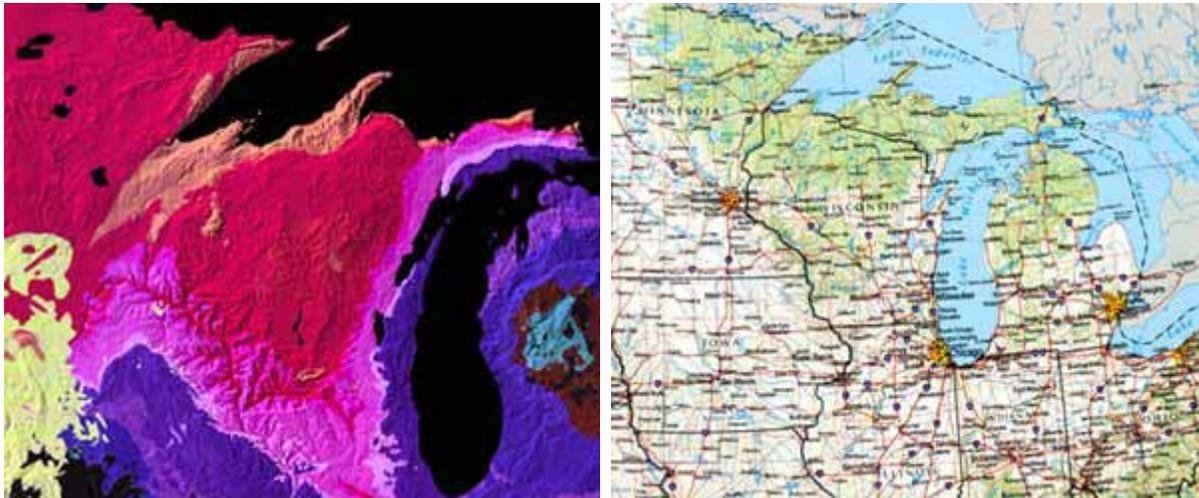


Figure 1.5: Examples of finished maps that can be generated using a GIS, showing landforms and geology (left) and human-built and physical features (right)

1.2.5 Products Generation

Product generation is the phase where final outputs from the GIS are created. These output products might include statistical reports, maps, and graphics of various kinds. Some of these products are **soft copy** images: these are transient images on television-like computer displays. Others, which are durable since they are printed on paper and film, are called **hard copy**. Increasingly, output products include computer-compatible materials: for example, CDs in standard formats for storage in an archive or for transmission to another system. The capability of taking the output of an analytic process, and placing it back into the geographic database for future analysis, is extremely important.

GIS should be capable for distinguishing product generation than any other system. The products of a GIS might neither be just a text, nor just a graphics, it usually be some kinds of maps involves coordinate systems, map projections, symbolization.....

1.2.6 Spatial Navigation

Navigation using GIS technology make it possible for you to study the whole area you are concerned with, and provides you a very convenient method by its roaming function to concentrate your study into the special area where you are more interested in.

Improvement of GIS also makes the navigation system to be developed in our life. Figure 1.6 shows a car navigation system, which has become prevalent in Japan.

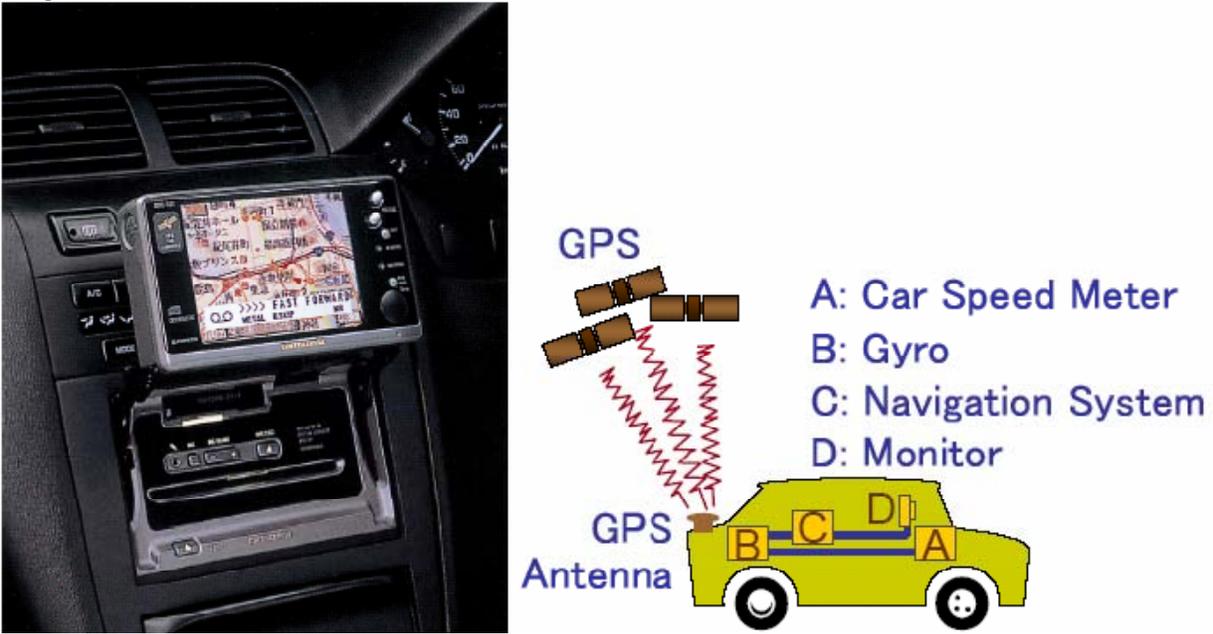


Figure 1.6: Car navigation system

1.2.7 Spatial Data Quality Control

The quality of spatial data refers to the completion, logicity, accuracy, availableness and metadata. Data quality control means a serious of scientific methods to ensure and improve the data quality in the process of data acquisition, management, edition and output.

1.2.8 Spatial Visualization

Through a process known as visualization, a GIS can be used to produce images— not just maps, but drawings, animations, and other cartographic products. These images allow researchers to view their subjects in ways that they never could before. The images often are helpful in conveying the technical concepts of a GIS to nonscientists.

Maps have traditionally been used to explore the Earth. GIS technology has enhanced the efficiency and analytical power of traditional cartography. As the scientific community recognizes the environmental consequences of human activity, GIS technology is becoming an essential tool in the effort to understand the process of global change. Map and satellite information sources can be combined in models that simulate the interactions of complex natural systems.

1.3 The Evolution of GIS

1.3.1 The Pioneering-Phase of GIS (1960s)

In 1969, Ian McHarg's *Design with Nature* was published. This work formalized the concept of land suitability/capability analysis (SCA). SCA is a technique in which data concerning land use in a locale being studied is entered into an analog or digital GIS. SCA programs are used to

combine and compare data types via a deterministic model, in order to produce a general plan map.

A system called STORET was developed by the Public Health Service for storage of spatial information about water quality (Green, 1964). Another, called MIADS, was developed by the U.S. Forest Service for the analysis of recreation alternatives and hydrology (Amidon, 1964). The Census Bureau was also heavily involved in geocoding and automated spatial data processing at this time.

In the university community at this time, Harvard's Laboratory for Computer Graphics developed and made available a series of automated mapping and analysis programs. The University of Washington at Seattle also made important contributions, particularly in the areas of transportation analysis and urban planning and renewal (Gaits, 1969). Urban planning applications blossomed with the development of these kinds of tools; by 1968 thirty-five urban and regional planning agencies in the United States were using automated systems (Systems Development Corp., 1968).

1.3.2 The Developing Period of GIS • 1970s •

The first system in the modern era to be generally acknowledged as a GIS was the Canada GIS or CGIS (Peuquet, 1977). Roger Tomlinson (1982), involved in the design and development of the system, states that CGIS was designed specifically for the Agricultural Rehabilitation and Development Agency Program within the Canadian government.

In 1977, a report issued by the United States Department of the Interior's Fish and Wildlife Service compares the selected operational capabilities of 54 GIS (USFWS, 1977). This survey, which is representative of several others conducted in the late 1970s, provides information on the hardware environment, programming language, documentation, and characteristics of the systems. This survey lists many GIS developed by federal and state agencies, as well as universities. However, it contained information on only a few commercial GIS. Even today, few commercial firms offer fully integrated GIS.

In addition to the beginnings of commercial GIS development, the 1970s also saw significant developments in image processing and remote sensing systems, which often had some GIS functions. Such firms as the Environmental Systems Research Institute in California began operation.

Image processing systems with some GIS elements were developed at the Jet Propulsion Laboratory and at the Purdue University laboratory for Applications of Remote Sensing. These latter systems incorporated GIS capabilities as the remote sensing community quickly realized that ancillary GIS data could play an important role in improving the accuracy of the interpretation of remotely sensed data.

1.3.3 The Leap Period of GIS Technology • 1980s •

The development of GIS, in terms of both the underlying concepts and the technology, has drawn on the talent and experience of many researchers and investigators. It has grown out of concerns about the state of the physical and cultural environment, and it has been advanced by efforts in both the public and private sectors.

Many early systems were developed to solve relatively narrow, specific kinds of problems. The past twenty years have seen an explosion in the technological base for these systems, particularly in the areas of data processing and remote sensing systems.

1.3.4 Leap of GIS Application (from 1990s)

With the leaping development of the computer industry, the Internet has become prevalent all over the world. The same time, GIS has been installed in many agencies as a part of their working system. Specifically, GIS has brought changes for the decision making department of the governments to vary their working style, configuration and even their plans. With the acquaintances by more and more people, the interest of GIS has been getting surge in many countries. The US has decided a “Digital World” strategy with the GIS including in it.

For example, in America, with the “Digital Globe Project” has been decided as a national strategy, and GIS in it. GIS will definitely become one of the most essential service systems of the modern society.

1.4 GIS Components

A typical GIS composed by the following five components (Figure 1.7):

- Hardware: Computer system: its internal and external device.
- Software: GIS program and other programs
- Data and database
- Analysis:
- People

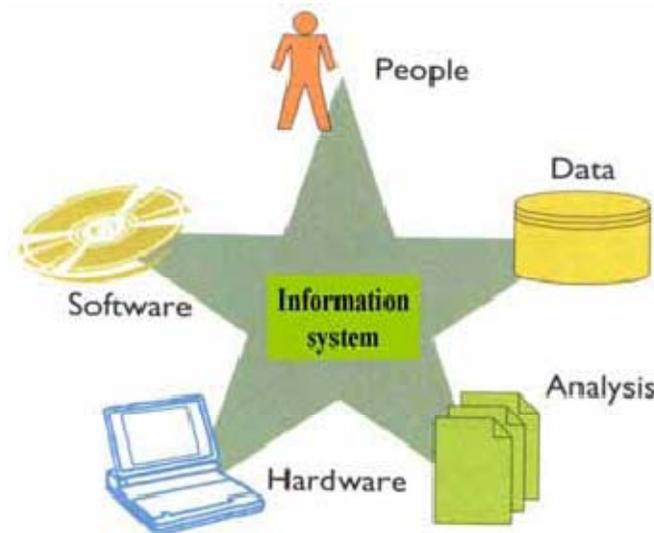


Figure 1.7: Components of GIS

1.4.1 Hardware

The hardware of GIS is made up of a configuration of core and peripheral equipment that is used for the acquisition, storage, analysis and display of geographic information (Figure 1.8).

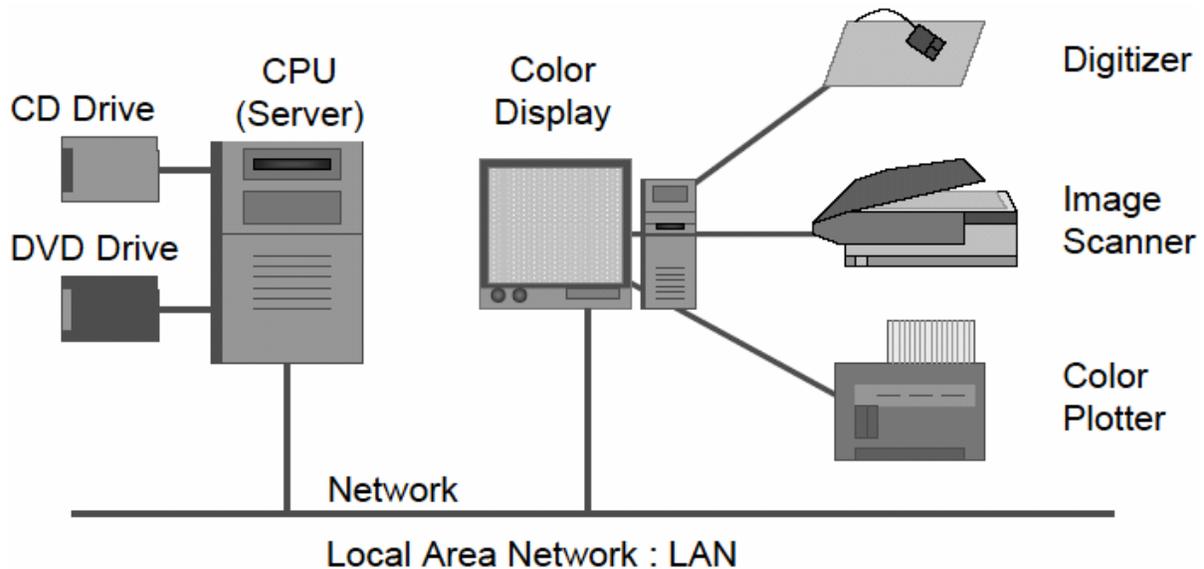


Figure 1.8: Hardware components of GIS

Core (Computer):

➤ Processing Unit (CPU)

The heart of GIS hardware architecture is the central processing unit (CPU) of the computer. The CPU consists of hard disk drive (HDD) for storing data and programs. The CPU also performs all the data processing and analysis tasks and also controls the input-output connectivity with data acquisition, storage and display systems. Depending on the data processing power of the CPU, computers are classified as supercomputers, main frame, minicomputers, work stations, and microcomputers or personal computers.

➤ Color display

Peripheral (External) device mainly includes input, output and storage device.

Input device:

➤ Scanners

➤ Digitizers

A digitizer and scanner are used to convert maps and documents in to digital form so that they can be used by computer programs in GIS environment.

Output devices:

➤ Printers

➤ Plotters

A plotter or a printer or any other kind of output device is use to present the result of data processing.

➤ Display monitors

Storage devices:

➤ Primary - memory

➤ Secondary - magnetic and optical disks

➤ Tertiary - near-on-line or off-line (tape libraries)

GIS at present are mostly implemented in client/server model of computing. A server is the computer on which data and software are stored. A client is the computer by which the users access the server. The application programs can be executed either on the server or the client

computer. A client can access multiple servers, and a server can provide services to a number of clients at the same time.

A network modem is also one of the hardware components of GIS used in inter-computer communication (internet) using telephone lines or special data lines with optic fibers.

1.4.2 Software

Software mainly consists of OS (Operation System), DB (database) and GIS software (Figure 1.9). GIS software is the core part of them. More than 300 sorts of GIS software have been developed and some of them are free ware which you can download from internet.



Figure 1.9: Software components of GIS

The basic functions of a GIS includes at least 4 main parts as shown in Figure 1.10, data input, data edition, spatial analysis and data output.

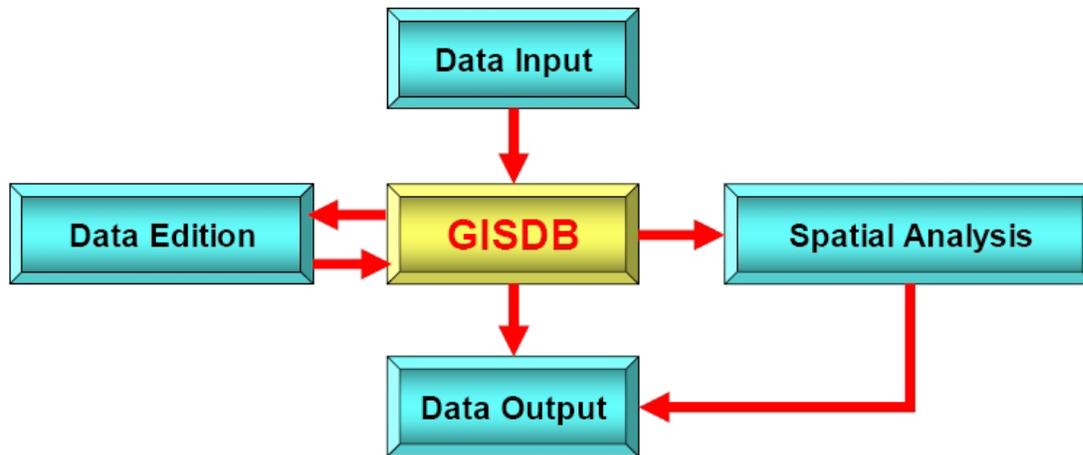


Figure 1.10: Main modules of GIS.

1.4.3 Data or Database (DB)

Data is the most important part in a GIS. The follows are basic items to be taken into consideration when building a GIS:

- What kinds of data have to be prepared?
- What kinds of data are existing?
- Is the accuracy of the existing data high enough?
- Are the formats of the existing data matching your GIS software?
- How can you collect the deficient but necessary data?
- What process is needed for the data processing before inputting?

1.4.4 Human Source

The responsibility and ability of the GIS staff and/or project manager is the key factor than any other components in a GIS.

1.5 Application of GIS

The application component of GIS can be explained from tree perspectives: areas of applications, nature of applications, and approaches of implementation. Table 1.1 summarizes the major areas of GIS application today. It is interesting to note how quickly these applications have grown in a relatively short history of development. As noted previously, when GIS was first developed it had a relatively narrow focus on land and resource management. Today, GIS are used in all sectors of the economy and for applications pertaining to both Earth's natural environment and human activities.

Table 1.1: Major Application Areas of GIS

Academic	<ul style="list-style-type: none"> ✓ Research in humanities, science and engineering ✓ Primary and secondary schools-School district delineation, facilities management, bus routing ✓ Spatial digital libraries
Business	<ul style="list-style-type: none"> ✓ Banking and insurance ✓ Real estate-development project planning and management, sales and renting services, building management ✓ Retail and market analysis ✓ Delivery of goods and services
Government	<ul style="list-style-type: none"> ✓ Federal government- national topographic mapping, resource and environmental management, weather services, public land management, population census, election and voting ✓ State/provincial government- surveying and mapping, land and resource management, highway planning and management ✓ Local/municipal government-social and community development, land registration and property assessment, water and wastewater services ✓ Public safety and law enforcement-crime analysis, development of human resources, community policing, emergency planning and management ✓ Health care ✓ International development and humanitarian relief
Industry	<ul style="list-style-type: none"> ✓ Engineering-surveying and mapping, site and landscape development, pavement management, pavement management ✓ Transportation- route selection for goods delivery, public transit, vehicle tracking ✓ Utilities and communications- electricity and gas distribution, pipelines, telecommunications networks ✓ Forestry-forest resource inventory, harvest planning, wildlife management, and conservation ✓ Mining and mineral exploration ✓ Systems consulting and integration
Military	<ul style="list-style-type: none"> ✓ Training ✓ Command and control ✓ Intelligence gathering

As the areas of GIS application have become more diversified, the nature of GIS applications has also undergone significant changes over the years. Such changes have been particularly drastic since the mid 1990s when GIS started to be implemented within computer networks. As a result of the ability to access multiple computers simultaneously through a LAN, GIS became an integrating tool for all types of geographically referenced data housed at different locations. This prompted the development of enterprise GIS applications that aim to address the business information needs of the entire organization, rather than those of specific work groups or departments.

The advent of the Internet has fundamentally revolutionized the nature of GIS applications. The Internet began as a computer network for military communications and for the exchange of scientific information in 1969. It has now become an international computer network of networks logically consisting of millions of academic, military, government, and commercial computers in a cooperative collaboration. By using different protocols of the Internet such as the World Wide Web (WWW) and File Transfer Protocol (FTP), GIS have now a virtual global system that offers all kinds of geographic information services via a world wide system of computer networks. At present, GIS serves not only as the means for geographic data management and spatial decision support, it also provides the mechanism for geographic information resource sharing and the communication of spatial information and knowledge as well.

1.6 Approaches to the Study of GIS

People now study GIS for different purposes. Many students study GIS simply as an academic pursuit; others study GIS to prepare for work as a specialist in a rapidly growing industry. There are also practicing professionals in various fields who study GIS in order to learn a new set of software tools that is increasingly used in their workplaces. The basic approaches to the study of GIS are summarized as follows.

1.6.1 Studying GIS as a Special Field of Academic Study

Before the 1980s, GIS was a set of relatively obscure computer applications that were found only in a small no of government agencies and universities. Today, it is hard to find a department of geography where no GIS course is offered. GIS is also widely taught in programs in Earth and Environmental sciences, computer sciences, business administration, natural resource management, urban and regional planning, geomatics, and civil engineering.

1.6.2 Studying GIS as a Branch of Information Technology

The rapid growth of GIS as a special sector in the computer industry and the proliferation of the use of GIS in public and business sectors have created a strong demand for people with technical skills in GIS. This in tern has led to the establishment of technically oriented GIS programs in community colleges and institutes of vocational training. The focus of these GIS programs is the development of technical skills in application programming, database creation and administration, systems implementation, and user support. In other words, they aim to train people who build GIS and make sure that these systems work, i.e., GIS specialists.

A typical technical GIS programs includes core courses in GIS principles and methods, as well as supporting courses in cartography, computer programming, database management, statistics and mathematics. The training on the use of GIS software packages is an essential part of these programs. It is interesting to note that many students attending these programs are university graduates in various disciplines who see the great opportunities and challenges as information technology (IT) workers in the information economy. Although these programs are very

technical in nature, it is important that students have a solid understanding of GIS concepts and principles. This will allow them to master GIS technical skills more easily. At the same time, it will also enable them to advance to technical managerial positions, where a good understanding of GIS concepts is an important as proficiency in GIS skills.

1.6.3 Studying GIS as a Spatial Data Institution and its Societal Implications

GIS has been developed as a technical tool for geographic information processing and analysis. Conventionally, many people perceived and used GIS simply as a special branch of information technology. As the use of GIS proliferates, its impacts on human society become very interesting and worthwhile research topics.

The interests in the role of GIS in society brought these questions to the core of GIS principles. GIS is no longer simply seen as a branch of information technology, but also as a set of institutionalized systems and practices for data management that must work within particular economic, political, cultural and legal structures. For the first time in the history of GIS, people have come to the realization that GIS will be able to become an important technology only if it can be totally integrated in to the fabric of the information society and serves its objectives well. The study of GIS, therefore, is not limited to the study of technology. It also has a very strong flavor of the humanities and social sciences as well.

2. DIGITAL REPRESENTATION OF GEOGRAPHIC DATA

The backbone of GIS is good data. Inaccurate data can result in inaccurate results and maps, skewing the results of your analysis and ultimately resulting in poor decisions.

2.1 Features of Geographic Data

Consider a simple object in space, a water well. From the point of view of a GIS, the primitive but essential piece of information to record about this water well is its location on the Earth, a data value pair such as longitude and latitude, thus storing the simplest kind of **spatial data**.

Two aspects of geographic entities have to be recorded in a GIS: absolute localization based in a coordinates system and topological relationship referred to other observations. Example: The Department of GeES & DEMS is located at the particular coordinate X, Y, or, The Department is located between Sociology and Anthropology departments. A GIS is able to manage both while computer assisted cartography packages only manage the absolute one.

For a water well, not only its spatial data, there may be a wide range of additional information which is required for many applications. This might include the depth of the well, the volume of water produced over a given period of time, dates of pump tests, and temporal sequences of measurements of dissolved and particulate matter in the water from the well. This second set of **non-spatial** or **attribute data**, which is logically connected to the spatial data, must not be forgotten.

In many GIS, there are tools to both store and manipulate the non-spatial data along with the spatial data. In some applications, the volume of non-spatial data may actually be larger than the volume of the spatial data, and the logical connections between the spatial and non-spatial information may be very important.

A water well is generally displayed in a map as a point, but a map usually contains lines for spatial entities as river and road, and polygon as village, province and countries. Therefore, spatial data can be divided into three categories, point, line and polygon.

Points: Features having specific location, without any extension in any direction can be represented as points. The points are represented by pair of x, y coordinates and a label (name). Example: - Location of oil wells, location of rain gauge stations, electric poles, etc.

Lines: Linear features on the map or earth surface can be represented by lines/polygons in GIS database. Lines consists of series of x, y coordinates with starting and ending points and a label. Line features will have length attributes. Roads, drainage, railways are some examples of linear features.

Polygons: Features with extended area can be represented by polygons. Polygons are closed features designed by a set of linked lines. Polygons will have starting and same ending coordinate (x, y) and a label. Polygons will have area and perimeter attributes.

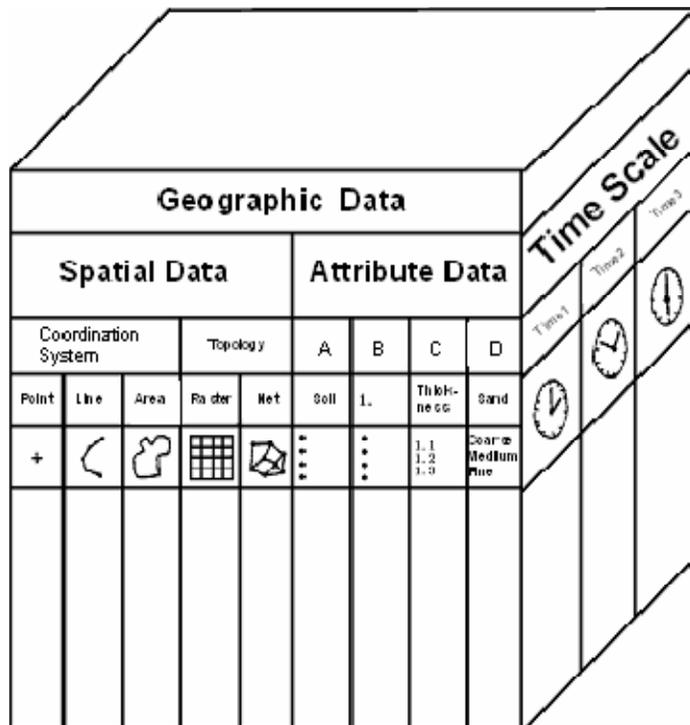


Figure 2.1: Features of geographic data

In many cases, not only attribute data, the change of spatial data and its attributes are also indispensable to record or describe the spatial entities. Therefore, three basic features, Spatial data, Attribute data and Time scale are usually necessary for spatial entities as shown in figure 2.1.

Measurement Scales

The attributes pertaining to spatial objects shown in a thematic map can be recorded by four different levels. They are (1) nominal, (2) ordinal, (3) interval, and (4) ratio.

Nominal Scale: Measurement on a nominal scale consists of simply placing each individual in to one of a number of categories. Since it makes no assumption about the value being assigned to the data, this is the lowest scale of measurement. There are again two types of scales in nominal, namely, Dichotomous (Binary) and Categorical scales. Dichotomous scale is used for mutually exclusive sets represented as yes or no. Examples include male & female, black & white etc. Categorical scale shows descriptive labels such as “wheat region”, “rock types” etc. The nominal scale can be applied over polygons, lines, and points.

Ordinal Scale: This ordinal scale shows ordering or ranks. Here, features are classified based on “names + rank”, while on nominal, features are classified based on names. A map can show a rank of cities, classification of roads into first, second and third class. Countries in the world can be ranked according to their population size as high, medium, low. Soils can be classified according to their drainage condition. Thus, ordinal scale can be expressed in different ways and is applicable to points, lines, and polygons.

Interval Scale: Features on interval scale are classified based on names + rank + specified quantities. Examples include populations of town A=5000, town B=10000, town C=15000, town

D=20000, and so on. Average annual temperature of Addis Ababa=27⁰C, Gondar=30⁰C, Afar=40⁰C. This scale does not have a natural (absolute) zero and uses an arbitrary (relative) one instead.

Ratio Scale: Measurements made on a ratio scale have all the characteristics of measurements on an interval scale. In addition, the ratio of any two values on the ratio scale is independent of the unit of measurement. For example, if a weight of one book is 1 pound and of another is 2 pounds; the ratio is 1:2 which is equal to 0.5. If the weights are converted to the metric system the ratio would be 453/906=0.5. It is not the case for non ratio scale such as temperature in ⁰C or ⁰F. If mean temperature of two successive days for a place are 40⁰F and 68⁰F, the ratio of the two values is 40/68=0.59; but if the figures are converted to ⁰C, the ratio will be 10/20=0.50 which is different from 0.59. This is because, zero ⁰C temperature does not mean that there is no temperature; it is simply a convention for freezing point. Unlike the interval scale, a ratio scale makes use of an absolute zero (or a true origin). Here, zero means non-existent. Zero kilometer (0 km) means no distance, it is just a point.

2.2 Spatial Data Classification

Spatial data can be classified by several categories.

2.3.1 Digital Data and Analogue Data

Spatial data can be divided into digital data and analogue data. The advantages of digital versus analogue data are outlined in the table below:

Table 2.1: Comparison of digital data and analogue data

Digital	Analogue
easy to update	whole map to be remade
easy and quick transfer (e.g. via internet)	slow transfer (e.g. via post)
storage space required is relatively small (digital devices)	large storage space required (e.g. traditional map libraries)
easy to maintain	paper maps disintegrate over time
easy automated analysis	difficult and inaccurate to analyze (e.g. to measure areas and distances)

2.3.2 Text and Imaginary

Spatial data includes text data and imaginary data.

Text data includes

- Reports
- Documents and records
- Statistics and census
- Results of investigation and experiments

Imaginary data includes

- Maps (digital map and analogue map)
- Photos
- RS (remote sensing, satellite images and aerial photographs)

2.3 GIS Data Sources

Spatial data can be obtained from different data source:

- Digitized and scanned data (processing needed existing data)
- Databases (available existing data)
- GPS field sampling of attributes
- Remote sensing and aerial photography

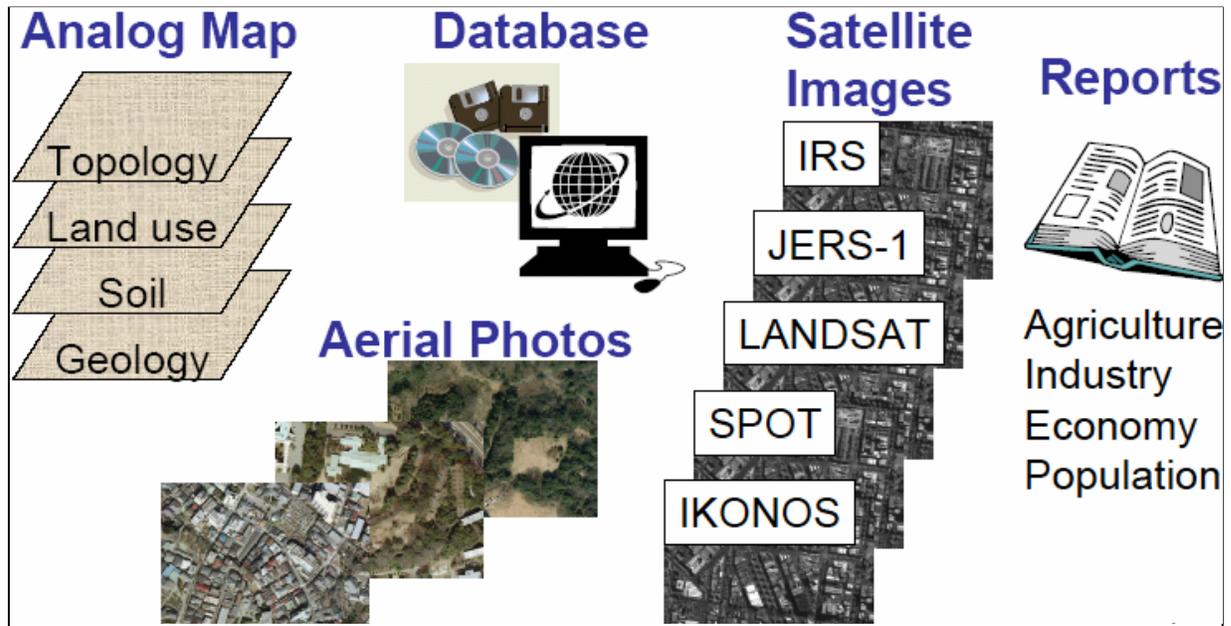


Figure 2.2: Various data sources for GIS

2.4 Data Entry and Presentation

The first step of using GIS is to provide it with data. The acquisition and preprocessing of spatial data is an expensive and time consuming process. And data entry process is error prone. Much of the success of a GIS project depends on the quality of the data that is entered into the system, and thus this data entry phase is critical and must be taken seriously.

Data entries are a procedure of encoding data in to a computer readable format and write to GIS software. Data entries are two types: Graphical data conversion and attribute data conversion. Graphical data conversion creates digital map layers. Attribute data conversion populates tabular data files associated with graphical elements on a layer.

The process of digital data conversion is made up of four phases of work.

1. *Acquisition:* Acquisition of digital data by digitizing existing maps, purchasing from government agencies or commercial data suppliers, or by collecting primary data (new data) using field survey, GPS based attribute and, remote sensing and photogrammetric data.
2. *Editing:* Editing to clean the acquired digital data, if necessary, in order to ensure that they are of an acceptable quality with respect to application objectives.

3. *Formatting or translating*: To convert the digital data into the specific physical data base format of the GIS where the data are used.
4. *Linking graphical data to their associated attribute data*.

The vector based data entry can be done using key board, digitizers, and scanners.

Data Entry by Digitization

Map digitization is carried out with a special data capturing device called 'digitizer' usually referred to as digitizing tablets. Digitizers are produced in different sizes.

A digitizer is made up of three components: a table, a cursor and a controller. The table consists of a large printed circuit board of extremely fine grids that is embedded between flat fiber glass and plastic plates. The spacing of the measurement grids determines the resolution of the table, which is the smallest distance separating two adjacent points. Most digitizers today have a resolution of 0.001 inch (0.025mm). The cursor is a free-moving component connected to the controller by means of a thin cable. It has a tracking cursor around which a field coil is mounted. Whenever the cursor button is pressed, this field coil sends out a signal that is picked by the measuring grid to generate the coordinates of the cursor position on the table surface. The cursor button is one of the buttons (usually from 4 to 16 in number) on the keypad of the puck. The controller is the interface between the digitizer and the host computer, the micro processor accepts the coordinates captured during digitization process and passed them to the host computer for display on the screen.

Map digitizing is a multiple process (Figure 2.3). The typical map digitizing procedures include (1) Preparation, (2) Creating a digitizing template, (3) Map digitizing, and (4) Post digitizing data processing.

Preparation for Digitizing: map digitizing starts by getting the map and the digitizer ready for the process. This includes checking the quality, i.e. accuracy, and completeness of the map data as well as identifying control points, known as 'tic points', for registering the output digital data to map coordinates.

Creating a Digitizing Template: A digitizing template contains the tic points, map neat lines and graphical elements that are common to all layers. For applications that require multiple layers, using a template enables these layers to be registered perfectly with one another, it helps to minimize the amount of work because graphical elements common to all layer need to be digitized only once (figure 2.4).

Map Digitizing: Map digitizing begins with the registration of the map mounted on the digitizing table to the digital map displayed on the screen. This is transformation process that uses the grid coordinates obtained by digitizing the tic points of the map and the ground coordinates of the corresponding points entered through the keyboard.

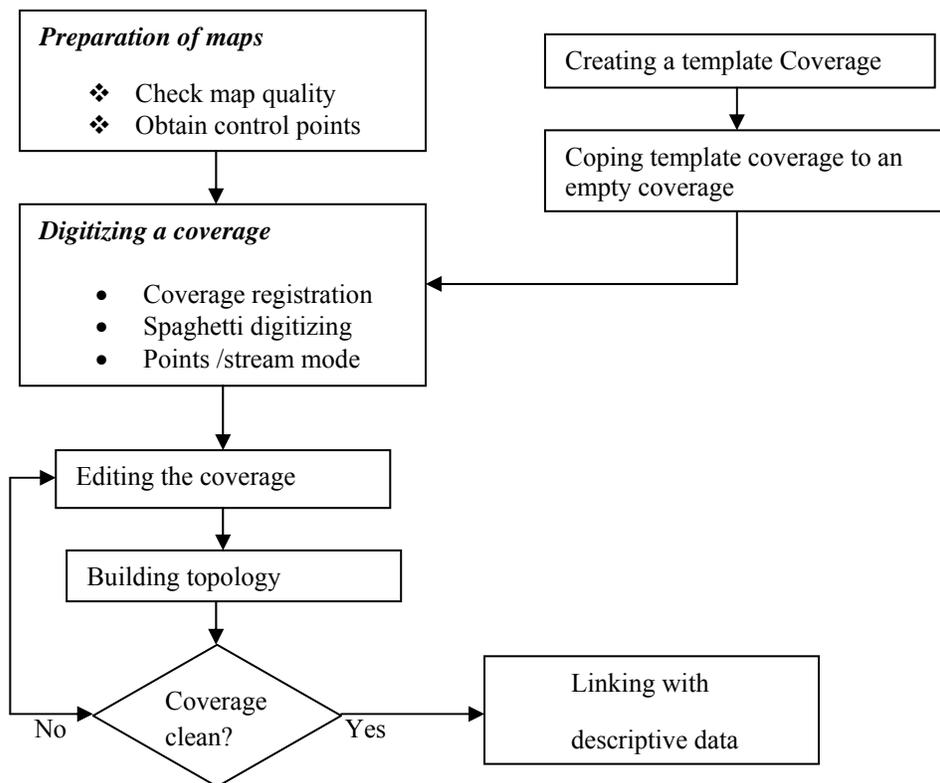


Figure 2.3: Graphical data conversion by map digitizing

On the completion of this process, the coordinates for corresponding objects on the map and on the screen become identical.

Map digitizing can be carried out in stream mode or point mode. In stream mode, the digitizer generates coordinates automatically at predefined time intervals. In the point mode, the digitizer will generate coordinates only when the user presses the button on the keypad of the digitizer. The data file size in point mode digitization is small, when compared to stream mode digitization.

Post digitizing Data processing (Graphical Data Editing): Once the digitizing is complete and data are saved, the post digitizing process should be performed. Post digitizing data processing is to ensure the integrity of the data before they can be used in a geographic database. Integrity of the data means that they are free from errors. The possible errors to be checked after digitization are:

- Lines intersect where they are expected to intersect (i.e., no undershoot or overshoot)
- Node are created at all points where lines intersect
- All polygons are closed
- Each polygon contains a label point
- The topology of the layer is built

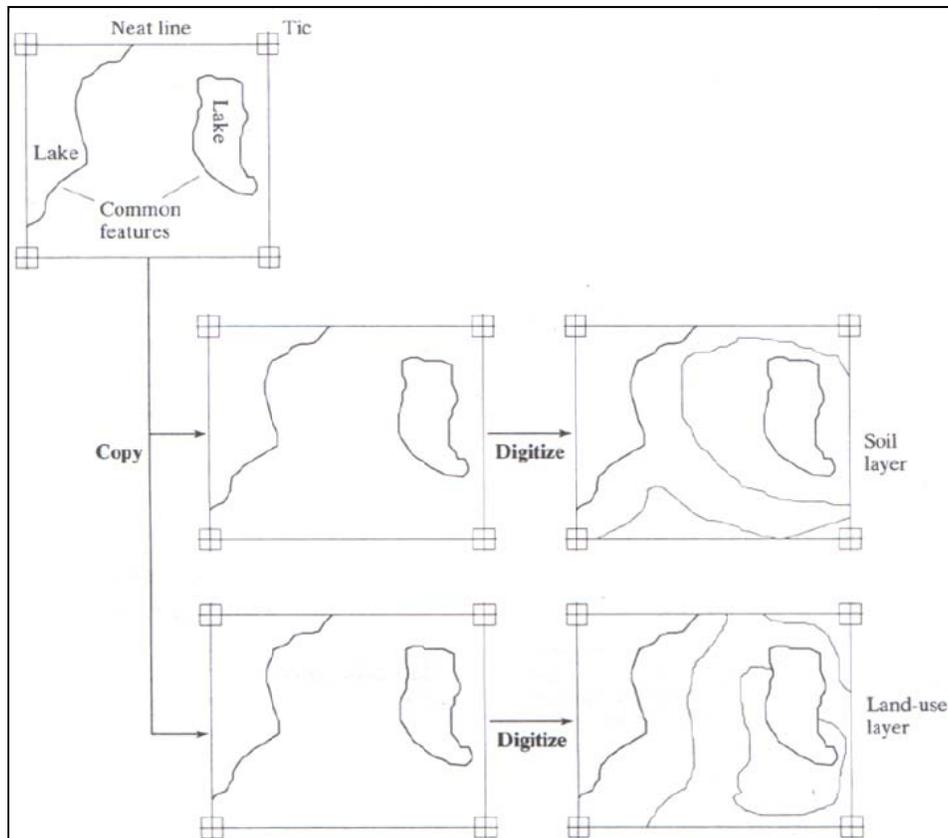


Figure 2.4: The method of using a digitizing template. Using a digitizing template eliminates the need to digitize common features on different layers and ensures perfect registration of these features when the layers are used for overlay analysis.

Data Conversion by Scanning and Vectorization

Map data conversion by scanning and vectorization is often referred to as screen digitizing or heads-up digitizing to distinguish it from conventional map data conversion using a digitizing table. The use of this technology for map data conversion has grown considerably in the last few years as the result of improvement of hardware design, software capabilities, and data compression techniques. This approach of digital data conversion is capable of converting a large number of maps in a relatively short time frame and at a cost comparable to or lower than the conventional methods of map digitizing. As a result, the scanner has gradually replaced the digitizer as the means of digital data conversion in GIS project.

It must be noted, however, that the process of scanning is only a very small part of the whole data conversion process (Figure 2.5). A considerable amount of work is needed to prepare the map for scanning. This includes separating the captured map data into different layers, touching up thin line work, as well as closing gaps in line objects. After scanning, it is necessary to convert the raster image to vector graphics, to build layer topology, and to link the resulting vector graphical elements to their associated attributes.

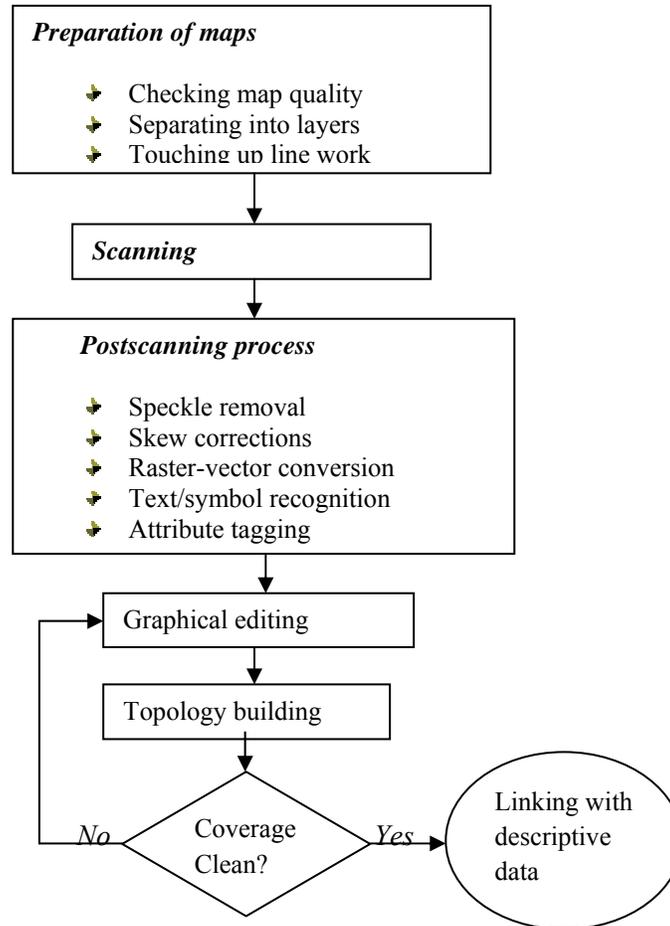


Figure 2.5: Graphical data conversion by scanning. Scanning of the map is only a small part of the whole data conversion process. Much work is needed to prepare the map before scanning and to postprocess the resulting vector data to make them useful for geographic database creation.

The scanner and its specifications

For use in GIS data conversion, a scanner has to meet a number of technical specifications; these include:

- **Resolution:** This is a measure of the sharpness of the image, usually expressed in terms of dots per inch (dpi) or actual pixel size in microns. Experience has shown that a resolution of 300 to 600 dpi is sufficient for general GIS applications involving paper maps; scanning of aerial photographic images for digital mapping requires a resolution as high as 2000 dpi.
- **Accuracy:** The accuracy of scanners is expressed as a ratio between the dimensions of the raster image and the original document. A scanner in good working condition is expected to have an accuracy of $\pm 0.1\%$.
- **Scan Width and Length:** A minimum scan width of 36 inches (91.4 cm) is required. Document length is not always a problem.
- **Output File Format:** It is essential for the output formats to be accepted by the vectorization software to be used.

- *Quality of the Accompanying Software:* This includes the ability of the software to integrate with the computer hardware platform, to perform batch and icon driven processing to remove speckles, and to correct the skew of the resulting image.

Post scanning Data Processing

Scanning is a nonselective data conversion process in the sense that every point on the map is captured. The resulting raster image is of limited use for GIS applications other than simple display on the computer screen. As the technology stands today, it is not yet possible to totally automate the vectorization of raster data on a scanned image. Postscanning data processing, using either computer assisted or manual methods, is still an indispensable part of the data conversion process. The actual amount of work that is required for postscanning data processing depends on a variety of factors, including the quality of the original maps, the complexity of the map contents as well as the functionality of the vectorization software.

In general, postscanning data processing is carried out as a sequence of processes that includes (1) raster-vector conversion, (2) raster text conversion, (3) raster symbol conversion, (4) graphical data editing, and (5) attribute data tagging.

Raster-Vector Conversion: This process, which changes raster images into vector graphics, can be carried out manually or with computer assistance (automated or semi-automated vectorization). In recent years, there has been some success in automating the raster-vector conversion process using artificial intelligence and pattern recognition techniques. The quality of the original maps seems to be the major limiting factor for the adoption of automated methods. For example, automated vectorization fails because of the presence of broken lines that cause gaps in the image, lines that are too thin or do not have adequate contrast to be picked up during image scanning represent another common source of error, special cartographic symbols and widely spaced characters in place names also cause failure in automated symbol and text recognition.

Raster Text Conversion: This process recognizes characters in the raster image and changes them into alphanumeric data. Relatively well-developed character recognition programs can now be used to perform this process automatically. However, verification of the result of character recognition and the corrections of errors that may occur are largely manual tasks.

Raster Symbol Conversion: This process recognized cartographic symbols in the raster image and converts them into alphanumeric codes. The apparent lack of standards in the form, size, and codes of cartographic symbols has made automated symbols recognition a much more difficult task than character recognition. As a result, raster symbol conversion is largely a manual task in practice.

Graphical Data Editing: This process cleans the graphics by removing data conversion errors in the same way as table digitizing errors are corrected.

Attribute Data Tagging: This process, which adds attribute data (e.g. feature identifiers, feature codes, and contour labels) to the graphical data after raster-vector conversion, is usually carried out interactively on the screen.

Global positioning system (GPS)

GPS is a Satellite-based Navigation System developed by the USA. It is used for location accuracy. It helps to fix locations of the real world features on map accurately. The GPS provides data on altitude of the location, geographic coordinates, and barometric pressure. It is also used for tracking.

There are three basic segments of a GPS, namely, Space Segment, Control Segment, and User Segment (Figure 2.6).

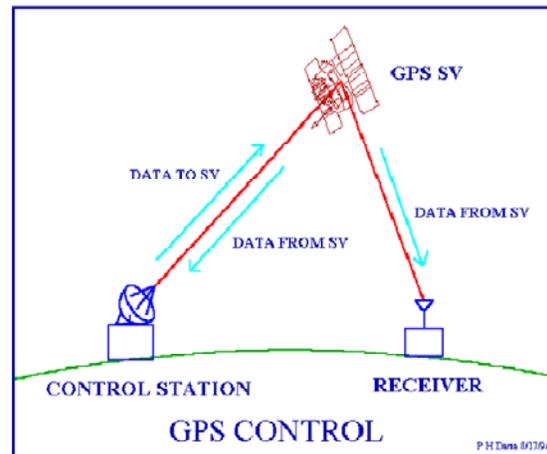


Figure 2.6: Segments of a GPS

Space Segment: The GPS works on constellation of 24 high altitude satellites which are orbited at a height of 26,000 km. The satellites are called NAVSTAR (Navigation System with Time and Ranging). The satellites are positioned in six earth centered orbital planes with four satellites in each plane. The nominal orbital period of a GPS satellite (i.e. the time for the satellite to complete one revolution around earth) is 11 hr 58 min. The orbits are nearly circular and equally spaced above the equator at a 60° separation, with an inclination relative to the equator of 55° .

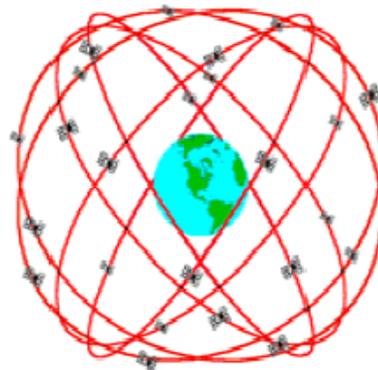


Figure 2.7: Space segment of a GPS



Figure 2.8: GPS. This Picture was taken at San Diego Aerospace Museum. This is the only GPS satellite on public display

Control Segment: The control segment monitor satellite orbits, maintain satellite health, maintain GPS time, update satellite navigation messages, and command small maneuvers of satellites to maintain orbit and relocations to compensate and failures.

User Segment: GPS can be used for civil and military purpose.

– Civilian Users

- Mapping, Surveying
- Navigation
- Search and Rescue (SAR)
- Pleasure, Sports, Hiking

– Military Users

- Navigation
- Guidance
- Artillery



GPS makes use of time of arrival (TOA) of the GPS signal to determine positions of earth's surface. A GPS satellite, which has a known position in space, sends out a signal to a receiver on earth's surface. The time interval, known as the signal propagation time, recorded at the receiver is then multiplied by the speed of the signal to give the emitter-to-receiver distance. By measuring the propagation time of signal broadcast from multiple satellites at known locations, the receiver can determine its position by means of a method "space resection" (Figure 2.9).

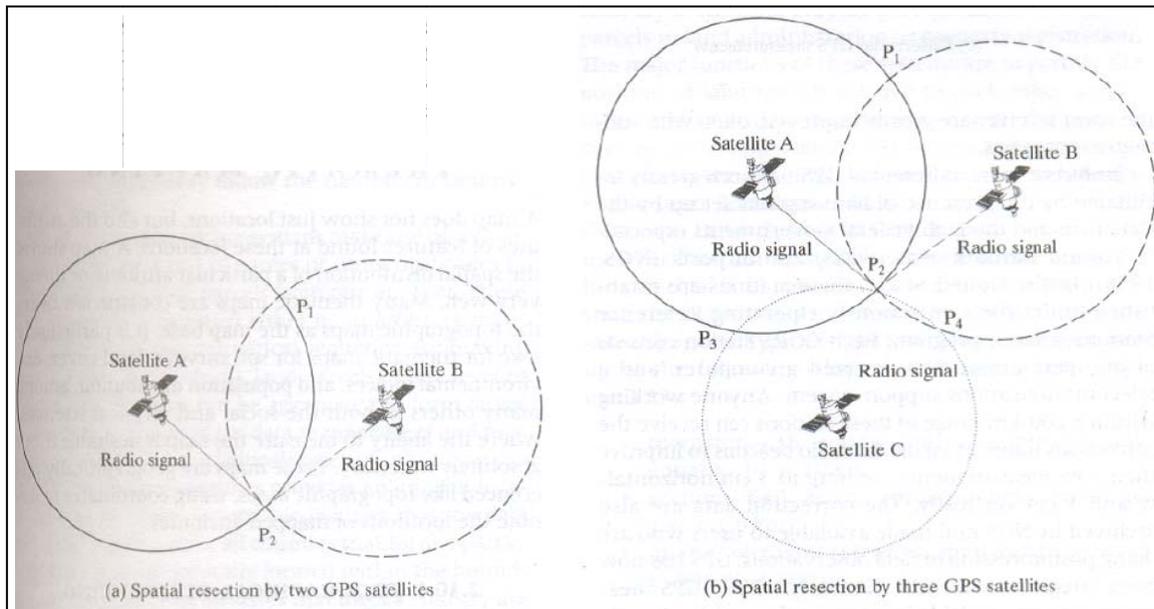


Figure 2.9: Spatial resolution by satellite ranging. Two satellites can not fix a position in space because the circles representing the satellite ranges meet at two locations, P1 and P2 as shown in (a). To fix a position, a minimum of three satellites is needed, as shown in (b). To find three-dimensional elevation, four satellites are required.

A GPS receiver can be employed as a stationary or mobile unit in point positioning, known respectively as static and kinematic GPS surveys. A single receiver in a static point positioning can give moderate accuracy (5 to 10m). Kinematic point positioning, which is used to determine the position of moving objects such as vehicle's trajectory in space, is capable of achieving an accuracy of 10 to 100m. For accurate distance determination, static relative positioning using two stationary receivers is the most commonly employed method by surveyors. The accuracy achievable is 1 PPM (parts per million) to 0.1 PPM. The method of differential GPS (DGPS) has been developed for accurate point determination. In this method, at least two GPS receivers are required. One receiver is placed at a station whose precise position is already known (hence known as the base, or reference station), and the other receiver is moving from one station to another (Figure 2.10). The base station calculates the actual corrections to the observed GPS readings. With the aid of communication links, these corrections are transmitted to the station where the rover receiver is located (or stored for later correction). In this way, the positions fixed by the rover receiver are greatly improved, often with sub-meter accuracies.

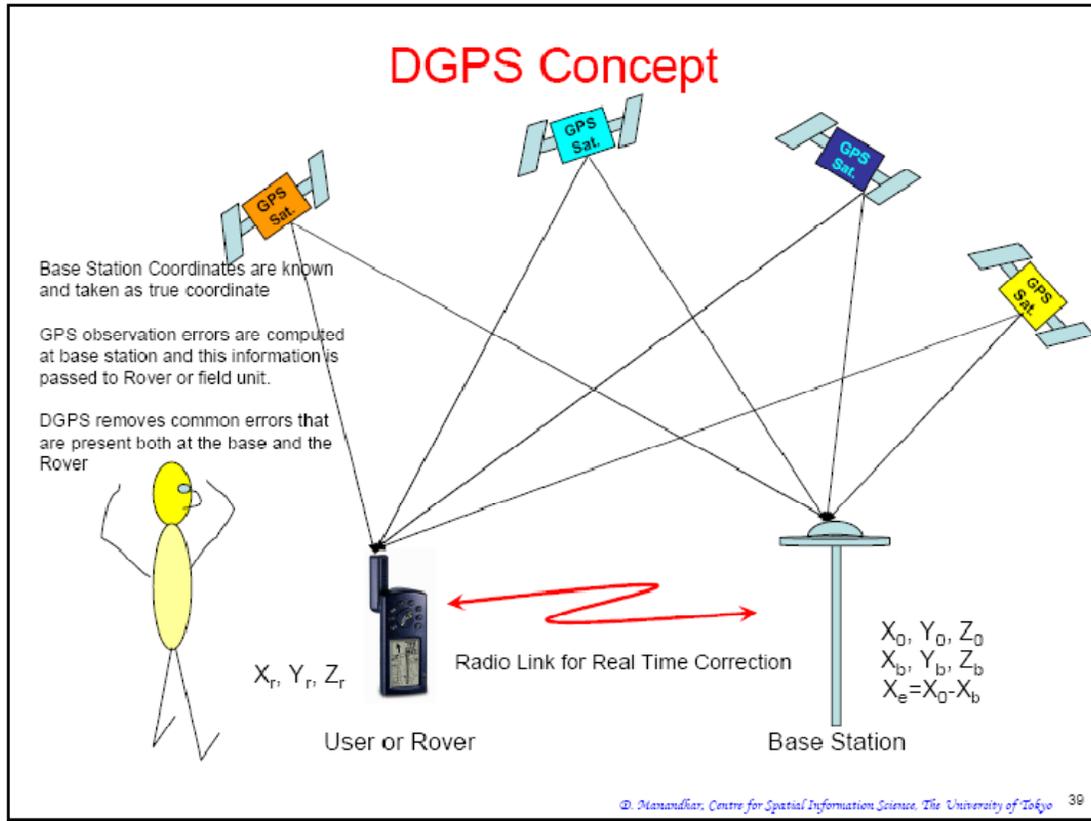


Figure 2.10: DGPS Measurements

Geographic Coordinate System of the Earth

The coordinate system is one of a set of numbers that determines the location of a point in space. In order to locate places on the earth, a three dimensional coordinate reference system has to be developed that takes in to account its shape. The spherical coordinate system in use over 2000 years is known as geographic coordinate system, which makes use of network of latitude and longitude (also known as graticule) to fix the positions of points on earth (Figure 2.11). Compare it with the two dimensional coordinate system of Figure 2.12. The two primary reference points on earth are the north and south poles. Halfway between the two poles is an imaginary line called the equator. The polar axis and the circle containing the equator intersect at right angle at the center of the earth, which is regarded as the origin of the coordinate system. The line of longitude passing through the Greenwich observatory England has value '0°'.

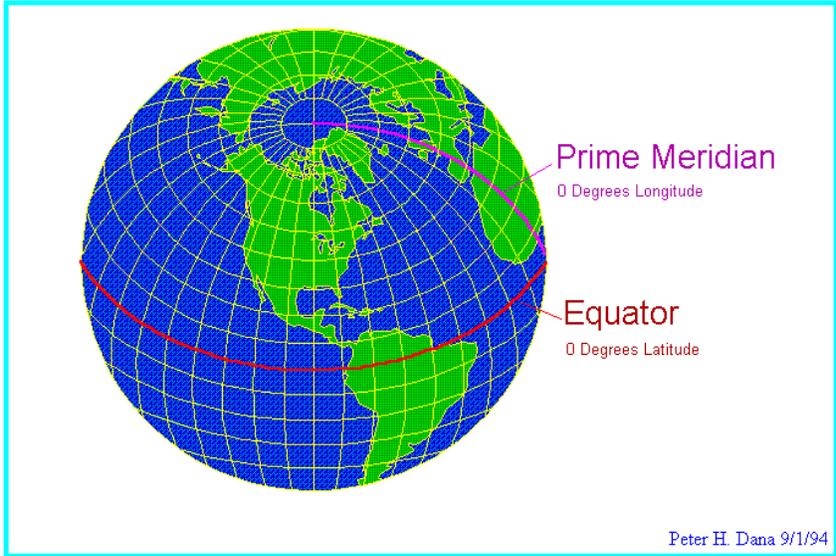


Figure 2.11: Three Dimensional World Coordinate System

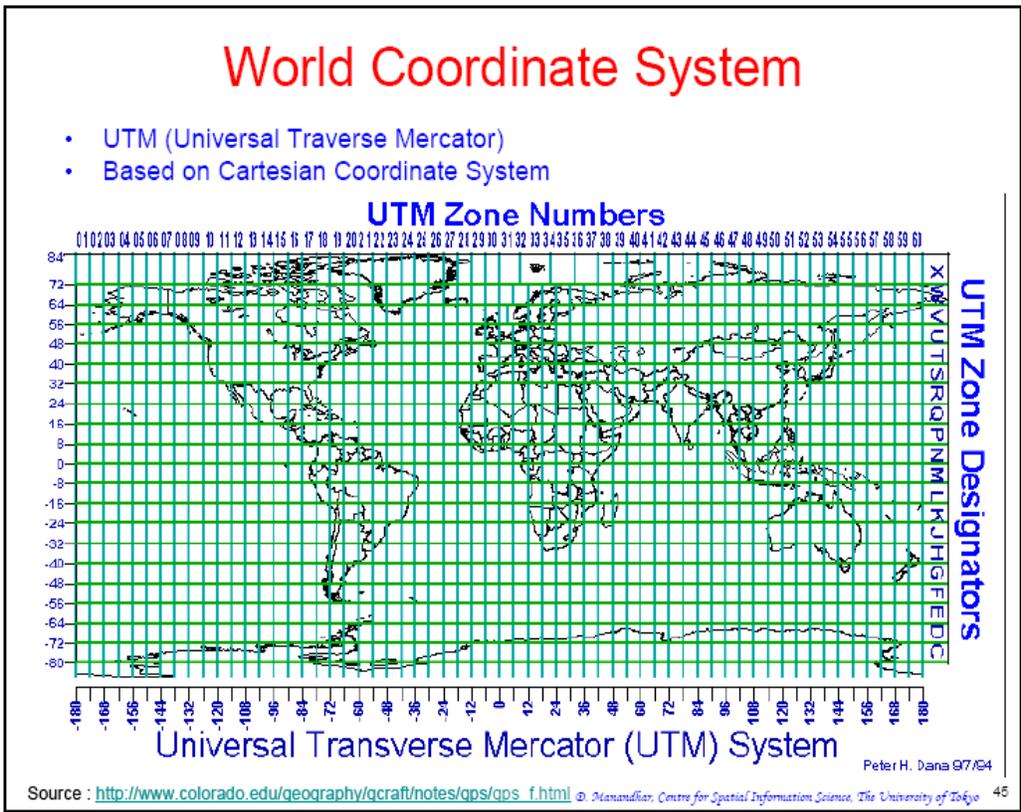


Figure 2.12: Two Dimensional World Coordinate System

Georeferencing

Geographic referencing or georeferencing is defined as the representation of the location of the real world features within the spatial framework of a particular coordinate system. The objective of georeferencing is to provide a rigid spatial framework by which the positions of real world features are measured, computed, recorded, and analyzed. The geoids and ellipsoid model is a mathematical surface that represents the shape of the earth. The ellipsoid is used as a reference surface for horizontal positions and the geoid is the reference surface for elevation.

Geodesists have been attempting to fit the earth in the ellipsoid using a polar flattening value. Numerous ellipsoids were developed to produce earth's best fitting ellipsoid. All these ellipsoids were defined using constants determined by measurements in a particular area of interest, i.e. a country or a continent. Ellipsoids defined in this way could fit well that part of the earth surface in or around the area of interest but not in other parts of the world. Depending upon how well they can fit the local surface of the earth, different ellipsoids have been adopted by different countries. Some 30 ellipsoids are common today. These include Airy 1830, Australian National, Clarke 1866, Clarke 1880, Everest 1956, GRS 80, and WGS 84.

Map projections

Map projection is the principles and techniques of transforming and representing three-dimensional features on Earth's surface on a two-dimensional flat map sheet. There are different ways to classify map projections. One of these methods is to classify map projections according to the type of developable surface on to which the network of meridians and parallels is projected. A developable surface is a surface that can be laid out flat without distortion.

There are three types of developable surfaces:

(1) Cylindrical, (2) Conical, & (3) Azimuthal

Cylindrical projection: A cylinder is assumed to circumscribe a transparent globe, so that the cylinder touches the equator throughout its circumference (Figure 2:13a). Open the cylinder along a meridian and unfolding it, a rectangular-shaped cylindrical projection is obtained. The meridians are vertical and parallel straight lines, intersecting the equator at right angles and dividing it into 360 equal parts. The parallels will be horizontal straight lines at some selected distance from the equator and from each other.

Conical projection: A cone is placed over the globe in such a way that the apex of the cone is exactly over the polar axis (Figure 2.13b). A cone must touch the globe along the parallel of latitude, known as standard parallel. Along the standard parallel, scale is correct and distortion is least. When the cone is cut open along a meridian and laid flat, a fan-shaped map is produced, with meridians as straight lines radiating from the vertex at equal angles, while parallels are arcs of circles, all drawn using the vertex as center.

Azimuthal or Planar projections: A plane is placed so that it touches the globe at the north or south pole. This can be conceived as the cone becoming increasingly flattened until its vertex reaches the limit of 180° .

The resulting project is Polar Azimuthal Projection. It is circular in shape with meridians projected as straight lines radiating from center of the circle, which is the pole. The parallels are complete circles centered at the pole (Figure 2:13c).

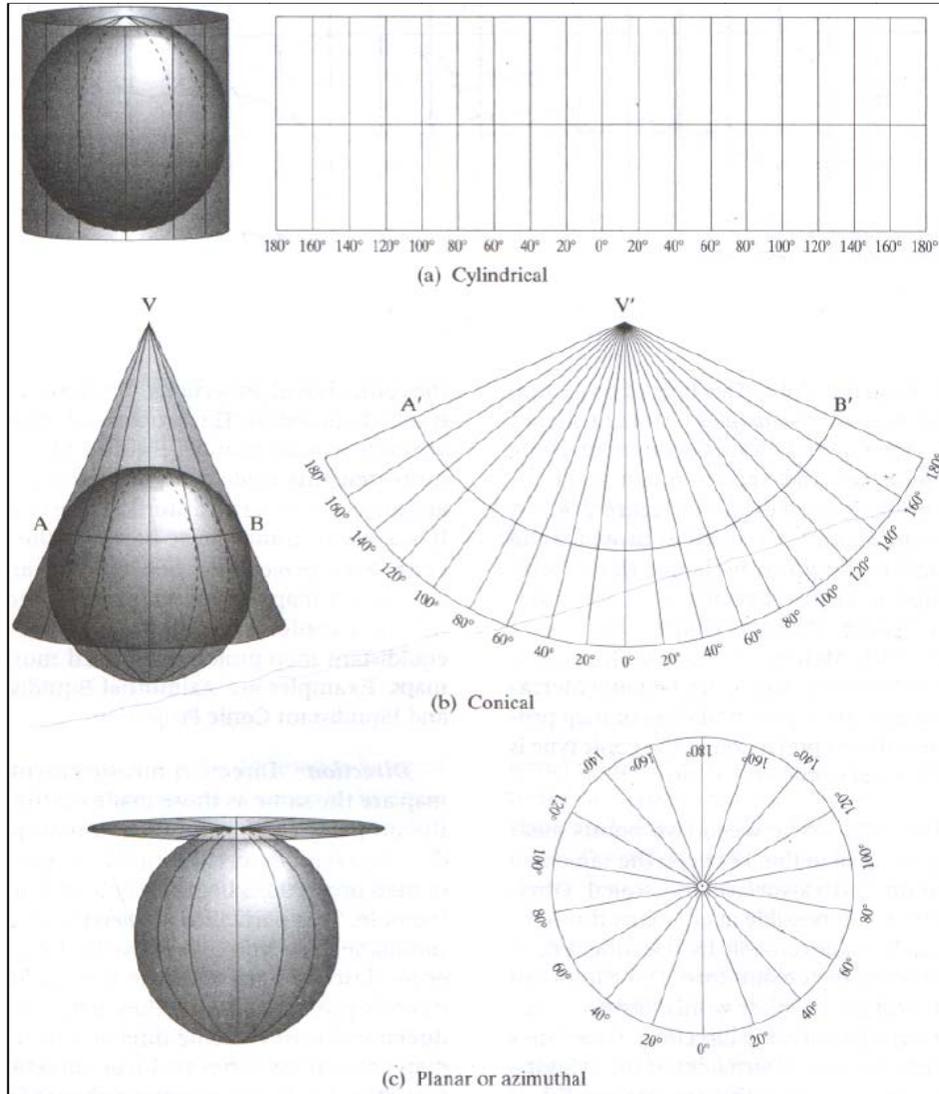


Figure 2.13: Basic Types of Map Projection

2.5 Database and Database Management Systems

A large computerized collection of structured data is called database. In non-spatial domain, databases have been used since 1960s for various purposes like bank account administration, stock monitoring, salary administration, order bookkeeping, and flight reservation. These applications have in common that amount of data is usually quite large, but that the data has itself simple regular structure.

Database setup is a difficult task. One has to consider carefully what the database purpose, who are the users. Then one needs to identify the available data sources and define the format in which the data will be organized within the database. This format is called database structure. After its design, data will be entered into the database. Keeping the data update is also important and some one (database manager) should be responsible for regular maintenance of the database. It is essential to document all the decisions made in creation of database.

The database management system (DBMS) is a software package that allows the user to setup, use and maintain database. A DBMS offers generic functionality for database organization and data handling.

The functions offered by DBMS to support effective storage and processing data are:

1. A DBMS support the storage and manipulation of very large data sets.
Some data sets are so big that storing them in text files or spread sheets files becomes too awkward for use in practice. The result may be that finding simple facts takes minutes, and performing simple calculation perhaps even hours.
2. A DBMS can be instructed to guard over some levels of data correction
For instance, an important aspect of data correctness is data entry checking: making sure that the data that is entered into the database is sensible data that does not contain obvious errors. Since we know in what study are we work, we know the range of possible geographic coordinates, so we cab make DBMS to check them. This is simple example of type or rules, generally known as integrity constraints that can be defined in and automatically checked by a DBMS. More complex integrity constrains are possible, and their definition is part of the development of a database.
3. A DBMS supports the concurrent use of the same dataset by many users
Moreover, for different users of the database, different views of data can be defined. In this way, users will be under the impression that they operate on their personal database, and not on one shared by many people. This DBMS function is called concurrency control.

Large datasets are built up over time, which means that substantial investments are required to create them, and that probably many people are involved in the data collection, maintenance and processing. These datasets are often considered to be of a high strategic value of the owner(s), which is why many want to make use of them within an organization.

4. DBMS provides a high level, declarative query language.

The most important use of the language is the definition queries. A query is a computer program that extract data from the database that meet the conditions indicated in the query. The word declarative means that the query language allows the user to define what data must be extracted from the database, but not how that should be done. It is the DBMS itself that will figure out how to extract the data that is requested in the query. Declarative languages are generally considered user-friendly because the user need not care about 'how' and can focus on the 'what'.

5. A DBMS supports the use of a data model. A data model is a language with which one can define a database structure and manipulate the data stored in it.

The most prominent data model is relational data model. Its primitive are tuples (also known as records or rows) with attribute values, and relations, being sets of similarly formed tuples.

6. A DBMS includes data backup & recovery functions to ensure data availability at all times. As potentially many users rely on the availability of the data, the data must be safeguarded against possible calamities. Regular back-up of the dataset, and automatic recovery schemes provides insurance against loss of data.
7. A DBMS allows to control data redundancy
A well-designed database takes care of storing single facts only once. Storing a fact multiple times- a phenomenon known as data redundancy, easily leads to situations in which stored facts start to contradict each other, causing reduced usefulness of the data. Redundancy is not necessarily always an evil, as long as we tell the DBMS where it occurs so that it can be controlled.

2.6. Spatial Data models

Spatial components can be represented by two models. These include Raster model and vector model. Raster model components are represented by square cells and rows and columns. The vector model components are represented by lines, points, and polygons, which are represented by x, y coordinates.

2.6.1. Raster Geographic data representation

The Raster data model is characterized by subdividing a geographic space in to grid cells (Figure 2.14). The linear dimensions of each cell define the spatial resolution of the data, which is determined by the size of the smallest object in the geographic space to be represented. The size is also known as the minimum mapping unit (MMU).

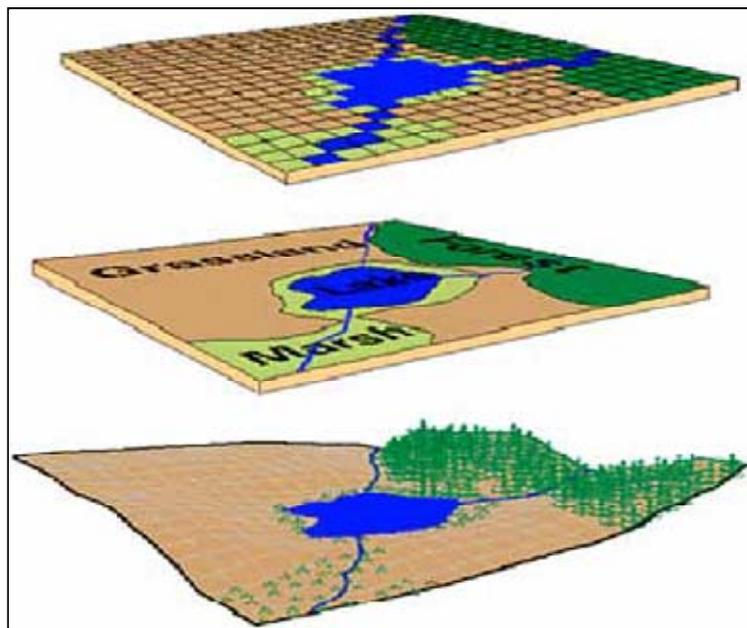


Figure 2.14: Example of raster data

Raster Data compression

It is common for a single raster data file to contain several million grid cells, the actual size of a file dependent on the bit depth, i.e. the number of bits used to represent the value of the pixel. For example, 8 bit data will give values from 0-255(2^8), 16-bit data from 0-65535 (2^{16}) and 32-bit data from 0-4294967295 (2^{32}). A few raster files can easily consume all the space on the hard drive of a small computer. The large file size also poses severe systems performance problems when the data is transmitted across the computer through networks. Data compression is therefore an important feature of digital representation of raster data. Data compression is the representation of data in a more compact form.

There are so many algorithms to handle and compress raster data. These include run-length encoding and quadtree representation.

Run-length encoding: In this method adjacent cells along a row with the same value are treated as a group called a “run”. Instead of repeatedly storing the same value for each cell, the value is stored once together with the number of cells that makes the run. In Figure 2.15, the 6x6 raster data matrix has 36 codes. In run-length code, the matrix becomes a string of “10A2B3A8B1C4B2C3B3C”. This string is made up of 18 codes, which represents 50% reduction in storage requirement.

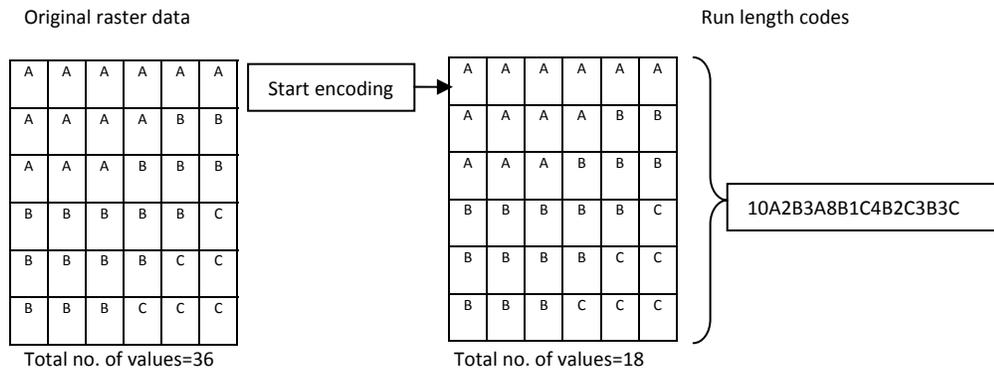


Figure 2.15: The method of run length encoding

Run-length recoding is conceptually simple to understand and technically easy to implement.

Quadtree data representation: The quadtree data model is a hierarchical tessellation model that uses grid cells of variable sizes. In this model, the geographic space is divided by the “process of recursive decomposition”. Instead of dividing the entire geographic space in to grid cells of the same size as in the case of the raster data model, the quadtree model uses finer subdivisions in areas where finer details occur. If the entire map contains only a single landuse type, the whole map will be represented by one single cell. If more than one land use is present, the map will be subdivided into four equal size quadrants. The same test is repeated for each of the four quadrants. Any quadrant that contains more than one land use type will again be divided into four equal parts, and any other quadrant that contains a homogeneous land use type will not be subject to further subdivision (Figure 2.16). The position of individual cells in the resulting quadtree can be determined by the cell identification number.

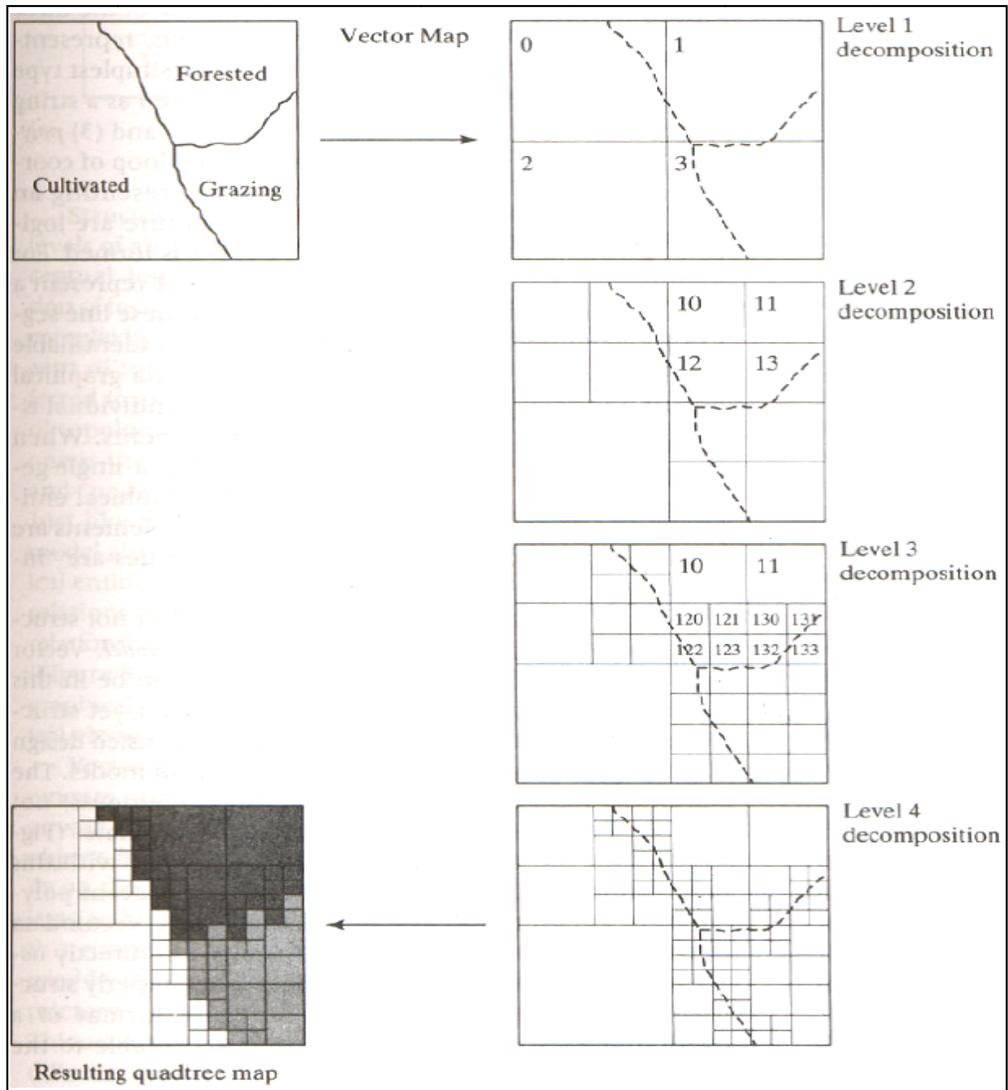


Figure 2.16: Quadtree data representation

2.6.2. Vector Data Representation

Vector data are represented by means of coordinates. When graphical elements (points, lines, polygons) representing an individually identifiable real world features are logically grouped together, a graphical entity is formed. The different line segments that represent a highway is graphical element. When these lines are logically joined together as an identifiable highway in the database, the highway is a graphical entity.

Spaghetti Data Model: Vector data that have been collected but not structured are said to be in the spaghetti data model. Vector data obtained by digitization are in this particular data model because they are not at structured. Spaghetti data model stores graphical elements, but not graphical entities, defined by strings of coordinates. There is considerable redundancy with this data model, as the boundaries between adjacent polygons are stored twice, one for each polygon. Vector data in spaghetti model are usually not directly usable by GIS (Figure 2.17).

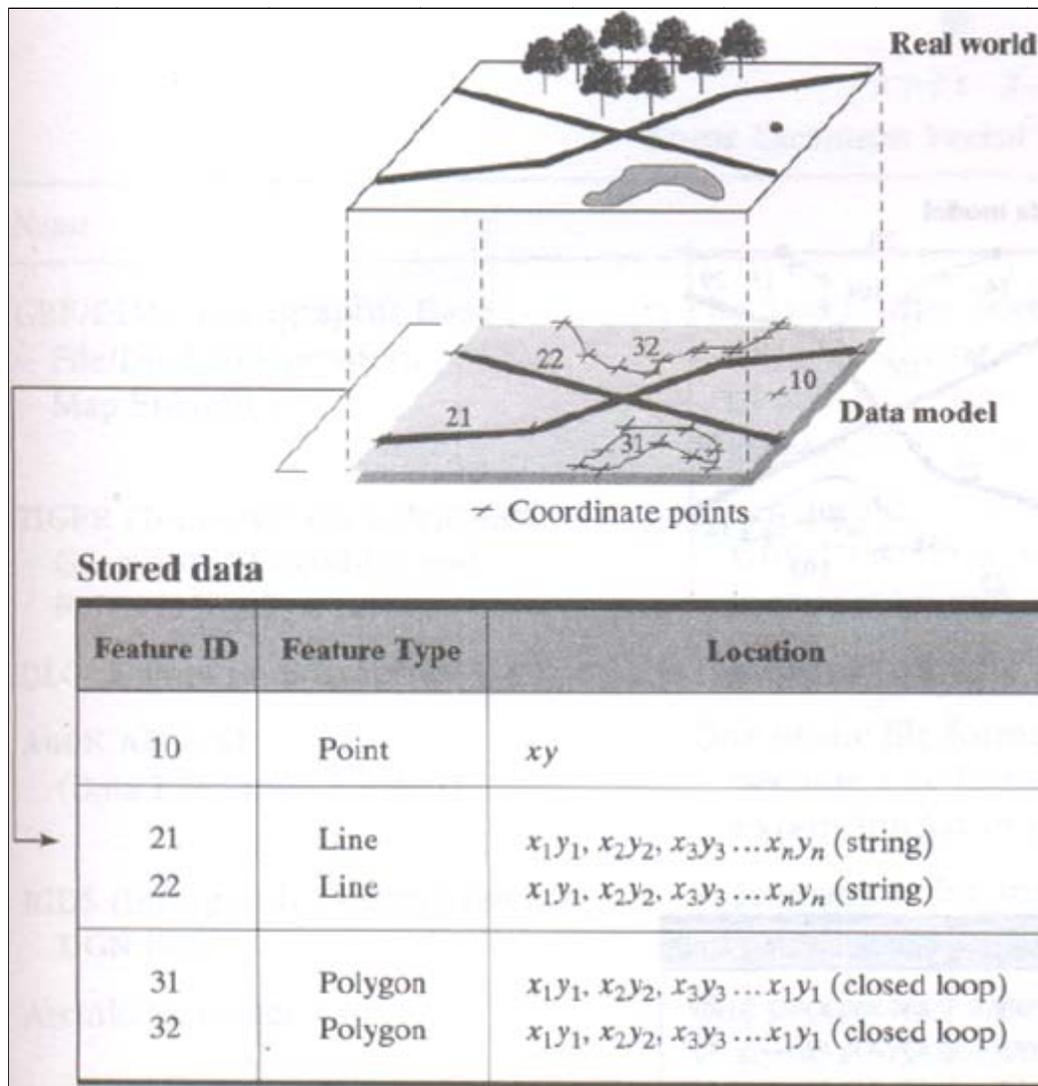


Figure 2.17: The Spaghetti Data Model. The Spaghetti Data Model stores geographic data in the form of graphical elements (point, line, and polygon), but not as graphical entities such as a lake, a wooded area, or a highway.

Topological Data Model: Topological data models are built on the concept of topology, they are referred to as topological data models. Of many variants of topological data models, the most commonly used one is the arc-node data model. “Arc” is a line segment, and “node” refers to the end points of the line segments. Just like spaghetti model, the arc-node data model also stores graphical elements rather than graphical entities. However, this model explicitly (clearly and separately) stores spatial relationship between the graphical elements, as well as relationship between the arcs and their respective nodes. The data stored in the forms of basic graphical elements can be used to represent simple and complex spatial objects (Figure 2.18).

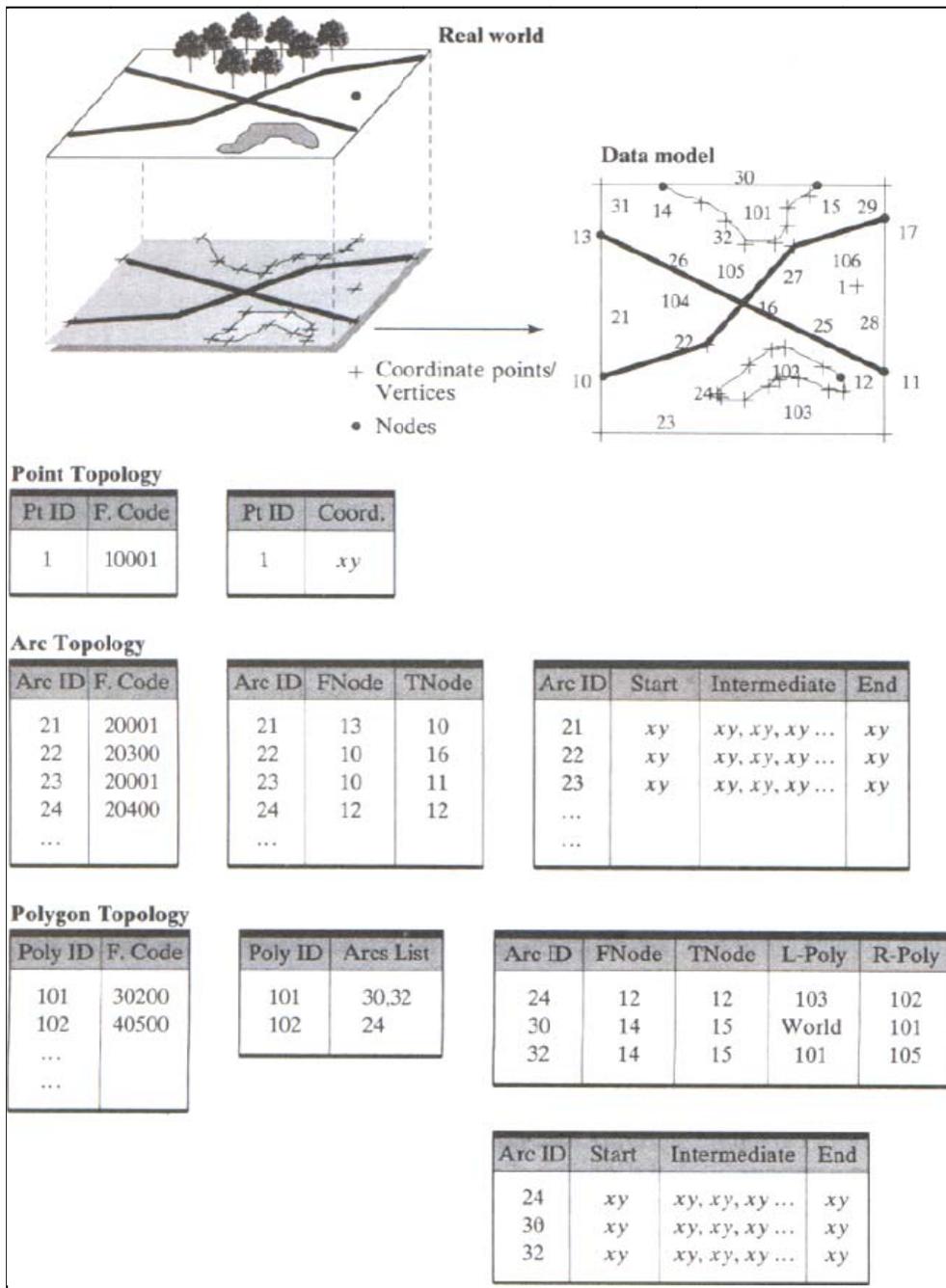


Figure 2.18: The arc-node topological data model

Triangulated Irregular Network (TIN): Triangulated irregular network is another vector based data model to represent terrain data. TIN represent the terrain surface as a set of interconnected triangular facets based on Delaunay triangulation. For each of the three vertices, the x, y coordinates and z values (elevation) are encoded. The coordinate data and topology for TIN are stored in a set of table. Adjacent tangles and nodes of the triangles are stored in separate tables.

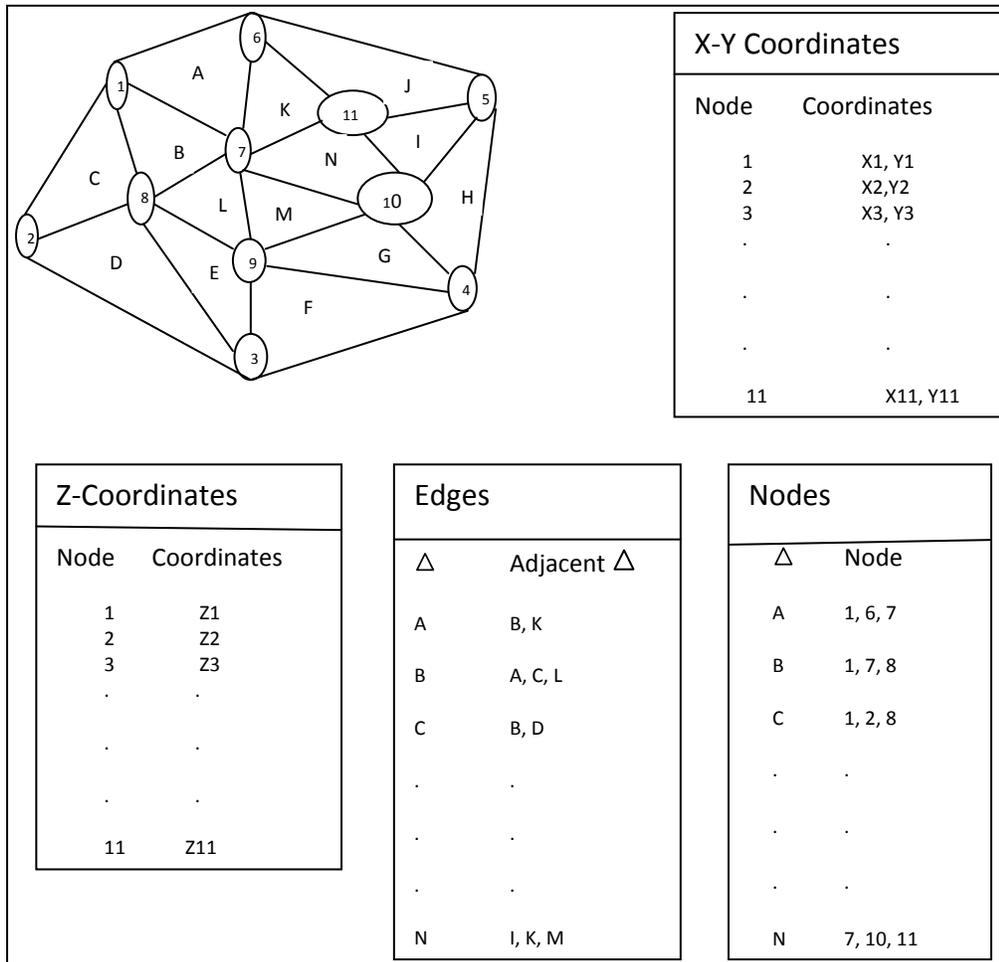


Figure 2.19: The Structure of a TIN

Triangulations on which triangles are most equilateral in shape tend to most accurately represent the surface using Tin model. Terrain parameters like slope, aspect are calculated for each facet and stored as an attribute.

2.7 Comparison of Raster and vector Data

Table 2.2: Relative merits and limitations of Raster and vector Data

<u><i>Raster Model</i></u>	<u><i>Vector Model</i></u>
<i>Merits</i>	<i>Merits</i>
➤ Simple data model	➤ More compact data structure
➤ Use of cheap technology	➤ Topological processing
➤ Ease of data collection	➤ Cartographic quality
➤ Ease of data processing	➤ Sophisticated attribute data handling
<i>Limitations</i>	<i>Limitations</i>
○ No topological relations	○ Complex data model
○ Limited attribute data handing	○ Difficult overlay processing
○ Less compact data structure	○ Difficult presentation of spatial variability
○ Low cartographic output	○ Expensive data collection
	○ Use of expensive technology

3. DATA QUALITY AND STANDARDS

3.1. Concepts and Definitions of Data Quality

Data quality refers to the fitness for use of data for intended applications. Data quality is largely determined by four generic measures of quality, namely, accuracy, precision, error, and the uncertainty that is associated with using data of unknown quality.

Accuracy: By definition, accuracy is the degree to which data agree with the values or descriptions of the real-world features that they represent.

Precision: Whereas accuracy is a measure of how “close” data are to true or accepted values, precision is a measure of how “exact” data are measured and stored. In mathematics, the exactness of representation is the number of significant digits used to record the data. Generally speaking, high precision does not necessarily mean high accuracy. For example, a wrong reading in an angle measurement made to the tenth of a second of arc is precise, but definitely not accurate. On the other hand, high accuracy does not necessarily always require high-precision data representation.

Error: The measure of error is relative to the measure of accuracy in that high-accuracy data are supposed to be free of errors. In practice, however, the two words are usually used in different contexts. Generally speaking, accuracy is used to imply “closeness” between the measured value and the true value of the real-world feature, but error is used to describe the “deviation” between these two values.

Uncertainty: When data of an unknown quality are used, there is always a certain degree of doubt or uncertainty about the validity of the information derived from the data. The basic difference between error and uncertainty is that, whereas, “error” refers to the lack of accuracy and precision in the data, “uncertainty” implies the lack of confidence in the use of the data that is due to the incomplete knowledge of the data. In other words, uncertainty is a measure of what we do not know.

3.2. Components of Data Quality

The characteristics that affect the usefulness of data can be divided in to 9 components. These have been grouped in to 3 categories: micro level components, macro level components, and usage components.

3.2.1. Micro Level Components

Micro level components are data quality factors that pertain to the individual data elements. These components are usually evaluated by statistical testing of the data product against an independent source of higher quality information. They include positional accuracy, attribute accuracy, logical consistency, and resolution.

Positional Accuracy: Positional accuracy is the expected deviance in the geographic location of an object in the dataset (e.g. on a map) from its true ground position. There are two components of positional accuracy: the *bias* and the *precision*. The bias refers to systematic discrepancies between the represented and true position. Bias is commonly measured by the mean or average positional error of the sample points. Precision refers to the dispersion of the positional errors of the data elements. Precision is commonly estimated by calculating the standard deviation of the selected test points. A low standard deviation indicates that the dispersion of the positional errors is narrow, i.e. the error tends to be relatively small. A measure of positional accuracy commonly used in surveying and Photogrammetry is the root mean square error (RMS). It is calculated by determining the positional error of the test points, squaring the individual deviations and taking the square root of their sum.

Attribute Accuracy: Attribute accuracy is defined as the “closeness” of the descriptive data in the geographic database to the true or assumed values of the real-world features that they represent. For example, elevation classification accurate to 1 m, 10m.

Logical consistency: Logical consistency refers to how well logical relations among data elements are maintained. For example, it would not be consistent to map some forest stand boundaries to the center of adjacent roads. Political and administrative boundaries defined by physical features should precisely overlay these features. An unusual problem is encountered when mapping areas with reservoirs. The water level in a reservoir will fluctuate over the year. Different GIS data layers may show the reservoir boundary at different locations, depending on the date of the mapping. As a result, the reservoir boundaries may be accurately delimited, but logically inconsistent among data layers. In this case, the problem is solved by providing a standard outline for each reservoir. The representation of the reservoir on each data layer is then made to conform to the standard outline.

Resolution: The resolution of a dataset is the smallest discernable unit or the smallest unit represented. In the case of images, such as air photos or satellite imagery, resolution refers to the smallest object that can be discerned, also termed spatial resolution.

3.2.2. Macro Level Components

Macro level components of data quality pertain to the data set as a whole. They are not generally amenable to testing but instead are evaluated by judgment (in the case of completeness) or by reporting information about the data, such as the acquisition date. Three macro level components are discussed: completeness, time and lineage.

Completeness: There are several aspects to completeness as it pertains to data quality. They are grouped here in to three categories: completeness of average, classification, and verification.

The completeness of coverage is the proportion of data available for the area of interest. A data set may not provide complete area coverage of the area of interest or attribute data may not be available for some portion of the data set.

Completeness of classification is an assessment of how well the chosen classification is able to represent the data. The completeness of classification may be evaluated with reference to a standard classification on its own merits with reference to specific applications. Table 2.2 is an example of a classification that exhibits several types of incompleteness. For a classification to be complete it should be exhaustive, that is it should be possible to encode all data at the selected level of detail. In this example, the subdivisions of the livestock category are not exhaustive. If the livestock category horses occur, it can not be encoded at this level. The subdivisions of the trunk crop category are exhaustive in that there is an appropriate category for any possible occurrence.

Table 3.1 sample classification to illustrate concepts of completeness

Level 1	Level 2	Level 3
Agricultural	Grains	
	Trunk crops	Broccoli Carrots Tomatoes Others
	Livestock	Cattle Hogs Sheep
Forest	Coniferous	Pine Spruce Fir
	Deciduous	
	Mixed wood	
Urban		
Water		

Completeness of verification refers to the amount and distribution of field measurements or other independent sources of information that were used to develop the data. Geologists indicate this aspect of data quality by using solid lines to map rock types for which they have direct field evidence, such as boundaries they have actually seen. Boundaries that were inferred but could not be verified are shown as dashed or dotted lines.

Time: Time is a critical factor in using many types of geographic information. For geographic information that changes relatively quickly over time, the date of acquisition may be a very important attribute.

Lineage: The lineage of a data set is its history, the source data and processing steps used to produce it.

3.2.3. Usage Components

The usage components of data quality are specific to the resources of the organization. These include data cost and accessibility.

Accessibility: Accessibility refers to the ease of obtaining and using data. The accessibility of a data set may be restricted because the data are privately held. Access to government-held information may be restricted for reasons of national security or to protect citizen rights. Even when the right to use restricted data can be obtained, the time and effort needed to actually receive the information may reduce its overall suitability.

Direct and Indirect costs: The direct cost of a data set purchased from another organization is the price paid for the data. The indirect costs include all the time and materials used to make use of the data. This may include familiarizing the staff with the data (if it is new to them), changing the format of the data if it is not compatible with the needed format (e.g. analog to digital).

3.3 Sources of Error

There is error associated with all geographic information. The following discussion reviews the major types of errors that are introduced at each stage of geographic information processing. Some of the more common errors are listed in table 2.3. The objective of dealing with error should not be to eliminate it but to manage it. Achieving the lowest possible level of error may not be the most cost effective approach. The level of error in a GIS needs to be managed so that data errors will not invalidate the information that the system is used to provide.

Table 3.2: Common sources of error encountered in using a GIS

Stage	Sources of Error
Data collection	Error in field data collection Error in existing maps used as source data Errors in the analysis of remotely sensed data
Data Input	Inaccuracies in digitizing caused by equipment and operator Inaccuracies inherent in the geographic feature (e.g. edges, such as forest edges, that do not occur as sharp boundaries)
Data storage	Insufficient numerical precision Insufficient spatial precision
Data Manipulation	Inappropriate class intervals Boundary errors Error propagation as multiple overlays are combined Slivers caused by problems in polygon overlay procedures
Data output	Scaling inaccuracies Error caused by inaccuracy of the output device Error caused by instability of the medium
Use of results	The information may be incorrectly understood The information may be inappropriately used

Data collection Errors: error exists in the original source materials that are entered in to the GIS. These errors may be a result of inaccuracies in field measurements, inaccurate equipment, or incorrect recording procedures. Air photo or satellite image interpretations introduce a degree of error in the classification and in the delineation of boundaries.

Data Input: The data input devices used to enter geographic data all introduce positional error. For example, digitizing tables are commonly accurate to fractions of a millimeter, but the accuracy varies over the digitizing surface. The center of a digitizing table commonly has a higher positional accuracy than the edges. The operator introduces error in the way the map is registered on the digitizing table, the boundaries are traced, and the accuracy with which the

attributes and label information are entered. Error is introduced in the way spatial information is represented. Curved boundaries are approximated by a series of straight line segments. The smaller the segments used, the more closely the boundary is approximated. Errors in the position of natural boundaries are often introduced because the boundary does not in fact exist as a sharp line. A forest edge, though drawn as a definite line, usually exists as a zone that may be several meters or tens of meters wide.

Data storage: The higher the precision of digital data, the more it consume storage space. And therefore, to save storage space, digital data might be stored with low precession.

Data manipulation: Many GIS analysis procedures involve the combining of multiple overlays. As the number of overlays used in an analysis increase, the number of possible opportunities for error increase. Many manipulation errors arise from the representation of boundaries. As noted previously, the same boundary may be drawn slightly differently in two overlays. The more complex the shape of the boundary, the more of a problem this becomes.

Data output: At the data output stage, error can be introduced in the plotting of maps by the output device and by the shrinkage and swelling of the map material. As paper shrinks and swell, measurements taken from that map will be changed. On a small scale map, the millimeter changes can represent several meters at the ground resolution.

Use of results: Error is also introduced when the reports generated by a GIS are incorrectly used. Results may be misinterpreted, accuracy levels ignored, and inappropriate analyses accepted.

4. GEOGRAPHIC INFORMATION VISUALIZATION AND PRODUCT GENERATION

4.1 GIS and Maps

The relation between maps and GIS is rather intense. Maps can be used as input for a GIS. They can be used to communicate results of GIS operations, and maps are tools while working with GIS to execute and support spatial analysis operations. As soon as a question contains a phrase like “where?” a map can be the most suitable tool to solve the question and provide the answer. “Where do I find University of Gonder?” is an example. Of course, the answer could be in non-map form like “in Amhara Region.” This answer could be satisfying. However, it will be clear this answer does not give the full picture. A map would put the answer in a spatial perspective. It could show where in Amhara Region University of Gonder is to be found (Figure 4.1).

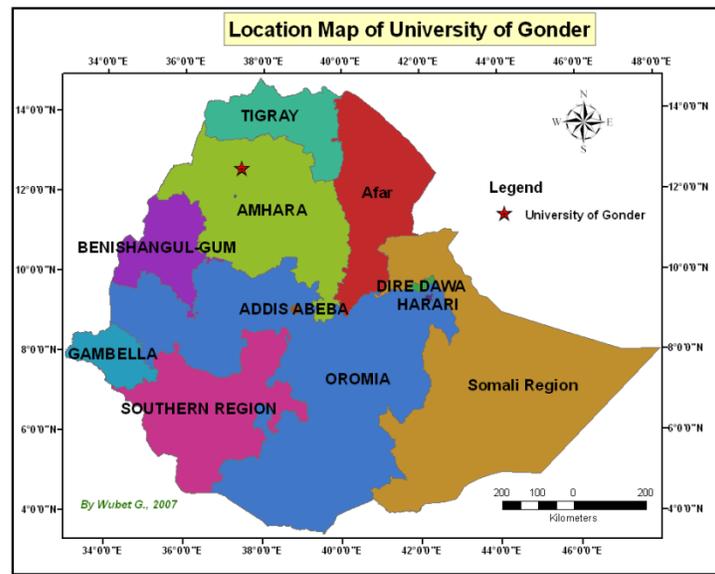


Figure 4.1: Maps and Location. Where is University of Gondar found?

As soon as the location of geographic objects (“where?”) is involved a map is useful. However, maps can do more than just providing information on location. They can also inform about the thematic attributes of the geographic objects located in the map. An example would be “What is the predominant land use in Ribb-Gumara Catchment?” The answer could, again, just be verbal and state “Agriculture.” However, such an answer does not reveal patterns. In Figure 4.2, a dominant southern and western agricultural pattern can be clearly distinguished. Maps can answer the “What?” question only in relation to location (the map as a reference frame). A third type of question that can be answered from maps is related to “When?” For instance, “When did Ribb-Gumara Catchment have greatest flood event, based on the two given landuse/landcover maps?” The answer is “1999,” and this will probably be satisfactory to most people. However, it might be interesting to see how this changed over the years. A set of maps could provide the answer as demonstrated in Figure 4.3. Summarizing, maps can deal with questions/answers

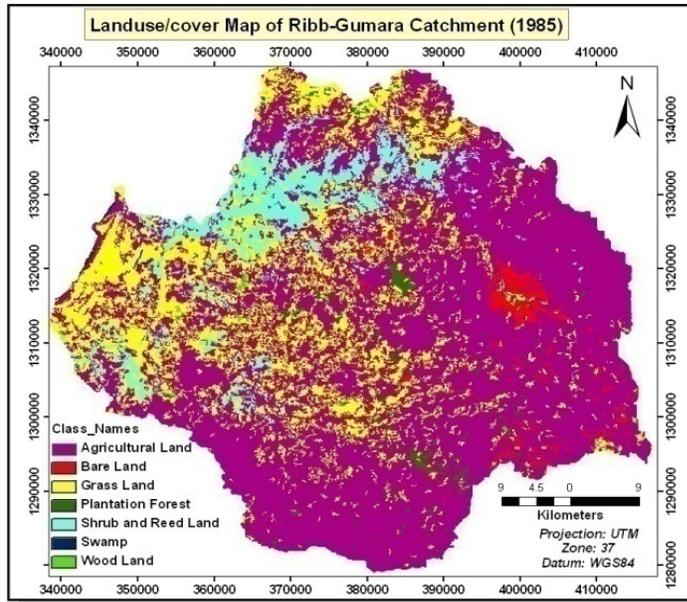


Figure 4.2: Maps and characteristics—“What is the predominant land use in Ribb-Gumara Catchment?””

related to the basic components of spatial or geographic data: location (geometry), characteristics (thematic attributes) and time, and their combination.

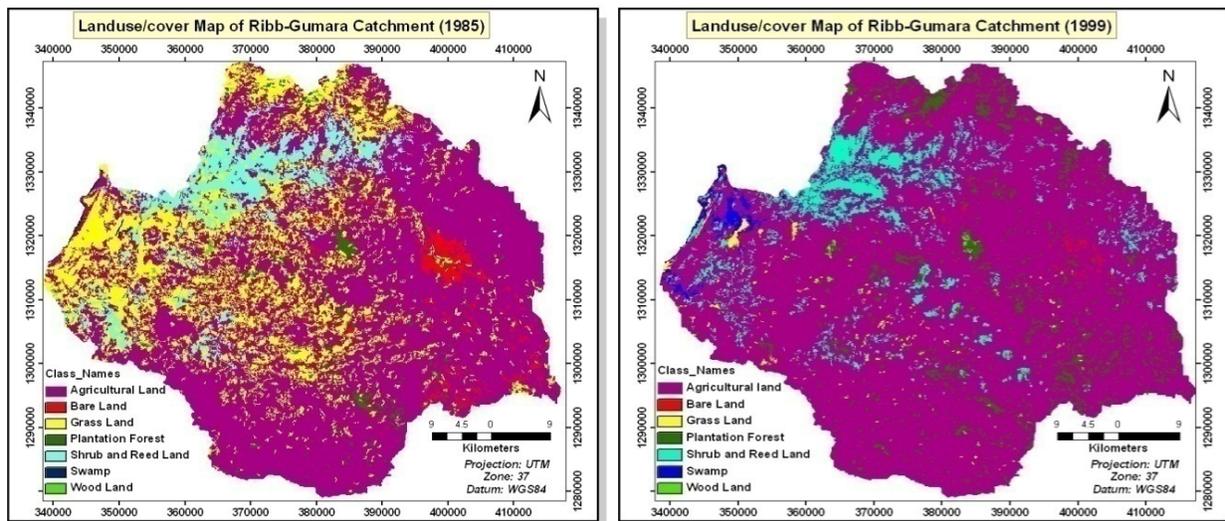


Figure 4.3: Maps and Time-“When did Ribb-Gumara Catchment have greatest flood event?”

As such, maps are the most efficient and effective means to transfer spatial information. The map user can locate geographic objects, while the shape and color of signs and symbols representing the objects inform about their characteristics. They reveal spatial relations and patterns, and offer the user insight in and overview of the distribution of particular phenomena. An additional characteristic of on-screen maps is that these are often interactive and have a link to a database, and as such allow for more complex queries.

Looking at the maps in this paragraph's illustrations demonstrates an important quality of maps: the ability to offer an abstraction of reality. A map simplifies by leaving out certain details, but at the same time it puts, when well-designed, the remaining information in a clear perspective. The map in Figure 4.1 only needs the boundaries of regions with their names, and a symbol to represent the position of University of Gonder. In this particular case there is no need to show cities, mountains, rivers or other phenomena.

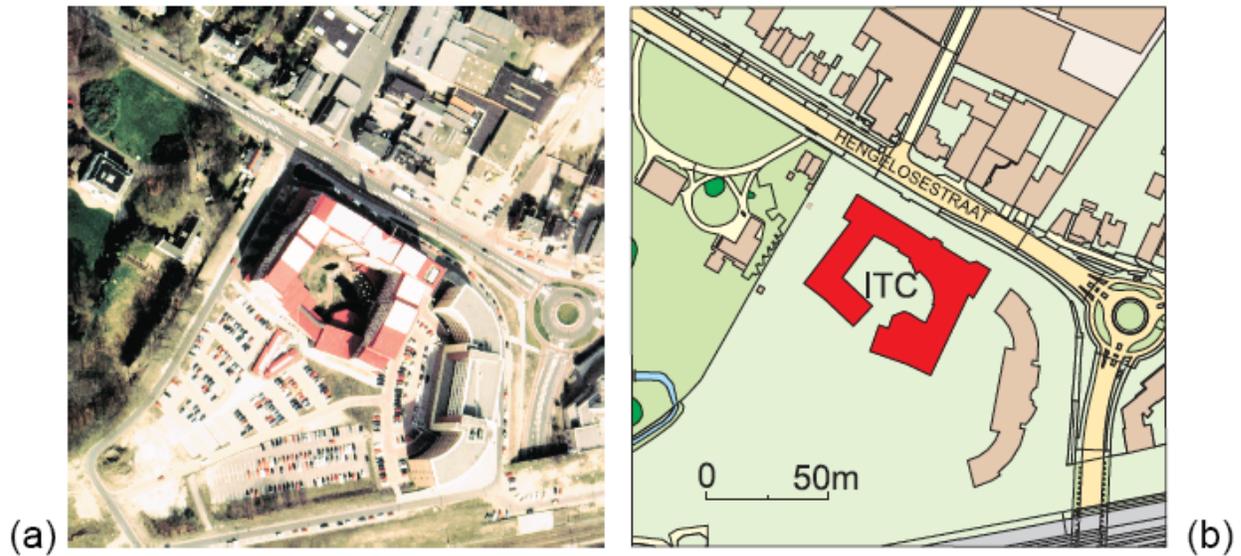


Figure 4.4: Comparing an aerial photograph (a) and a map (b)

This characteristic is well illustrated when one puts the map next to an aerial photograph or satellite image of the same area. Products like these give all information observed by the capture devices used. Figure 4.4 shows an aerial photograph of the Information Technology Communication (ITC), the Netherlands, building and a map of the same area. The photographs show all objects visible, including parked cars, small temporary buildings. From the photograph, it becomes clear that the weather as well as the time of the day influenced its contents: the shadow to the north of the buildings obscures other information. The map only gives the outlines of buildings and the streets in the surroundings. It is easier to interpret because of selection/omission and classification. The symbolization chosen highlights our building. Additional information, not available in the photograph, has been added, such as the name of the major street: Hengelosestraat. Other non-visible data, like cadastral boundaries or even the sewerage system, could have been added in the same way. However, it also demonstrates that selection means interpretation, and there are subjective aspects to that. In certain circumstances, a combination of photographs and map elements can be useful.

Apart from contents, there is a relationship between the effectiveness of a map for a given purpose and the map's scale. The Public Works department of a city council cannot use a 1:250,000 map for replacing broken sewer-pipes, and the map of Figure 4.1 cannot be reproduced at scale 1:10,000. The *map scale* is the ratio between a distance on the map and the corresponding distance in reality. Maps that show much detail of a small area are called *large-scale maps*. The map in Figure 4.4 displaying the surroundings of the ITC-building is an

example. The map of Ethiopia in Figure 4.1 is a *small-scale map*. Scale indications on maps can be given verbally like ‘one-centimeter-to-the-kilometer’, or as a representative fraction like 1: 200,000,000 (1 cm on the map equals 200,000,000 cm (or 2,000 km) in reality), or by a graphic representation like a scale bar as given in the map in Figure 4.4(b). The advantage of using scale bars in digital environments is that its length changes also when the map zoomed in, or enlarged before printing.’ Sometimes it is necessary to convert maps from one scale to another, but this may lead to problems of (cartographic) *generalization*.

Having discussed several characteristics of maps it is now necessary to provide a definition. A *map is* “a representation or abstraction of geographic reality. A tool for presenting geographic information in a way that is visual, digital or tactile.” The first sentence in this definition holds three key words. The geographic reality represents the object of study, our world. Representation and abstraction refer to models of these geographic phenomena. The second sentence reflects the appearance of the map. Can we see or touch it, or is it stored in a database. In other words, a map is a reduced and simplified representation of (parts of) the Earth’s surface on a plane (two dimensional surface).

Traditionally, maps are divided in *topographic* and *thematic maps*. A topographic map visualizes, limited by its scale, the Earth’s surface as accurately as possible. This may include infrastructure (e.g., railroads and roads), land use (e.g., vegetation and built-up area), relief, hydrology, geographic names and a reference grid. Figure 4.5 shows a small scale topographic map of Fogera Woreda. Thematic maps represent the distribution of particular *themes*. One can

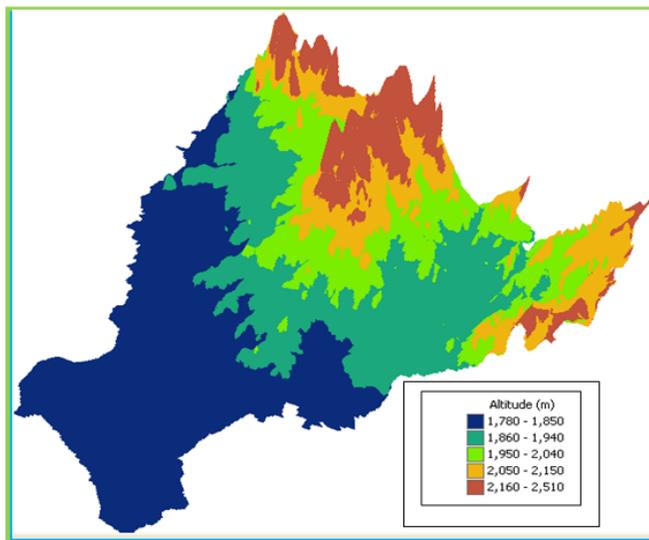


Figure 4.5: The Three Dimensional Topographic Map of the Fogera Woreda

distinguish between *socio-economic themes* and *physical themes*. The map in Figure 4.6(a), showing population density in Fogera Woreda, is an example of the first and the map in Figure 4.6(b), displaying the Woreda's drainage areas, is an example of the second. As can be noted, both thematic maps also contain information found in a topographic map, so as to provide a geographic reference to the theme represented. The amount of topographic information required depends on the map theme. In general, a physical map will need more topographic data than most socioeconomic maps, which normally only need administrative boundaries. The map with drainage areas should have added rivers and canals, while adding relief would make sense as well. Today's digital environment has diminished the distinction between topographic and thematic maps. Often, both topographic and thematic maps are stored in the database as separate data layers. Each layer contains data on a particular topic, and the user is able to switch layers on or off at will.

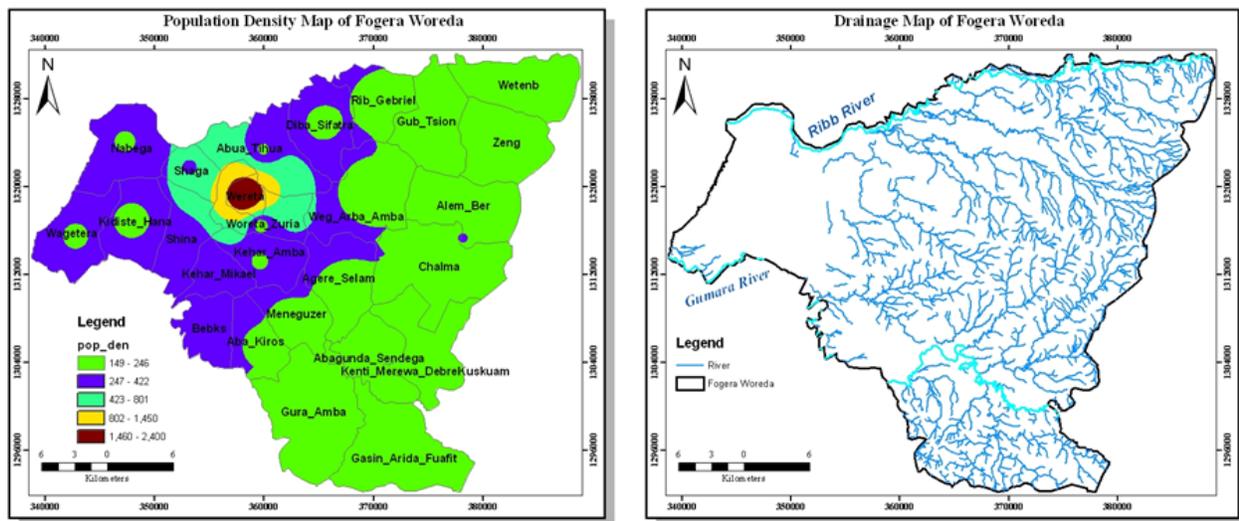


Figure 4.6: Thematic maps: (a) socio-economic thematic map, showing population density of Fogera Woreda (higher densities in Woreda town); (b) physical thematic map, showing drainage of same area.

The design of topographic maps is mostly based on conventions, of which some date back to centuries ago. Examples are water in blue, forests in green, major roads in red, urban areas in black, etc. The design of thematic maps, however, should be based on a set of cartographic rules, also called *cartographic grammar*, which will be explained in the coming subtopics. Nowadays, maps are often produced through a GIS. If one wants to use a GIS to tackle a particular geo-problem, this often involves the combination and integration of many different data sets. For instance, if one wants to quantify land use changes, two data sets from different periods can be combined with an overlay operation. The result of such a spatial analysis can be a spatial data layer from which a map can be produced to show the differences. The parameters used during the operation are based on computation models developed by the application at hand. It is easy to imagine that maps can play a role during this process of working with a GIS. From this perspective, maps are no longer only the final product they used to be. They can be created just to see which data are available in the spatial database, or to show

intermediate results during spatial analysis, and of course to present the final outcome.

The users of GIS also try to solve problems that deal with three-dimensional reality or with change processes. This results in a demand for other than just two-dimensional maps to represent geographic reality. Three-dimensional and even four-dimensional (namely, including time) maps are then required. New visualization techniques for these demands have been developed. Figure 4.7 shows the dimensionality of geographic objects and their graphic representation. Part (a) provides a map of the ITC building and its surroundings, while part (b) shows a three-dimensional view of the building. Figure 4.7(c) shows the effect of change, as two moments in time during the construction of the building.

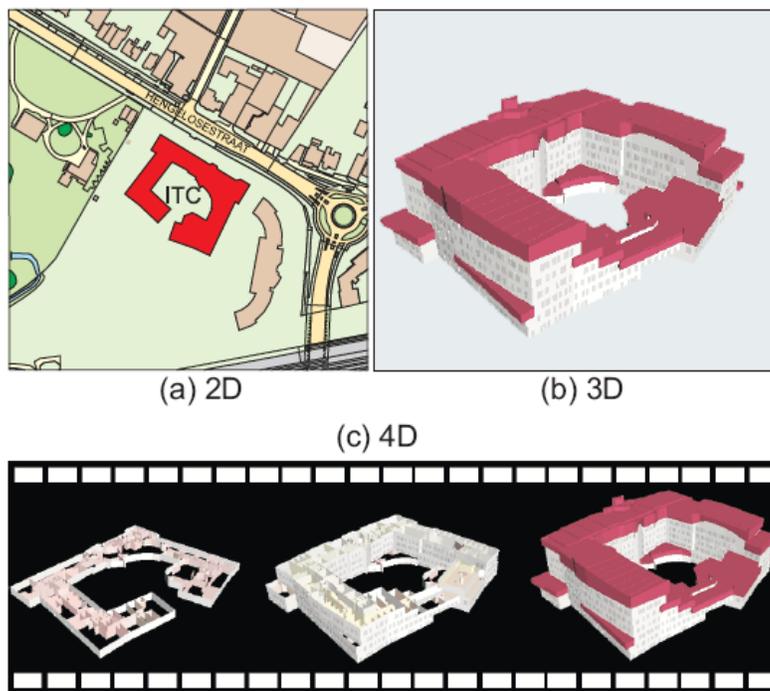


Figure 4.7: The dimensions of spatial data: (a) 2D, (b) 3D, (c) 3D with time.

4.2 Cartography in the Context of GIS

Cartography and GIS are linked by virtue of their common focus on maps. Traditionally, GIS users relied on concepts and techniques drawn from cartography to produce maps. However, since the technologies for producing maps in conventional cartography and GIS are quite different from one another, it is now generally recognized that conventional principles and techniques are no longer adequate for GIS applications.

The renewed interest of GIS researchers and users in cartography was prompted primarily by the disappointment in the poor quality of many maps produced by GIS. Many GIS users have come to the realization that it is the cartographic expertise of people, not computer functionality that plays the most important role in ensuring good quality maps, even in a computer based technology such as GIS. The current interest of GIS users in cartography reflects the common

understanding that it is not possible to rely entirely on the computer to produce maps, but they have to take personal responsibility themselves.

4.3 Visualization of Geographic Information

Cartographic cognition is the process by which the human brain recognizes spatial patterns and relationships. The principles and techniques of using dreamed computer graphics for the analysis and interpretation of a large volume of numerical data are called scientific visualization (SciVis) also referred to as visualization for scientific computing (ViSC). Visualization of geographic information can be done using a variety of techniques. These include (1) two-dimensional plots; (2) three-dimensional plots; (3) two-dimensional planimetric views; (4) three-dimensional perspective views; and (5) animation (Figure 4.8).

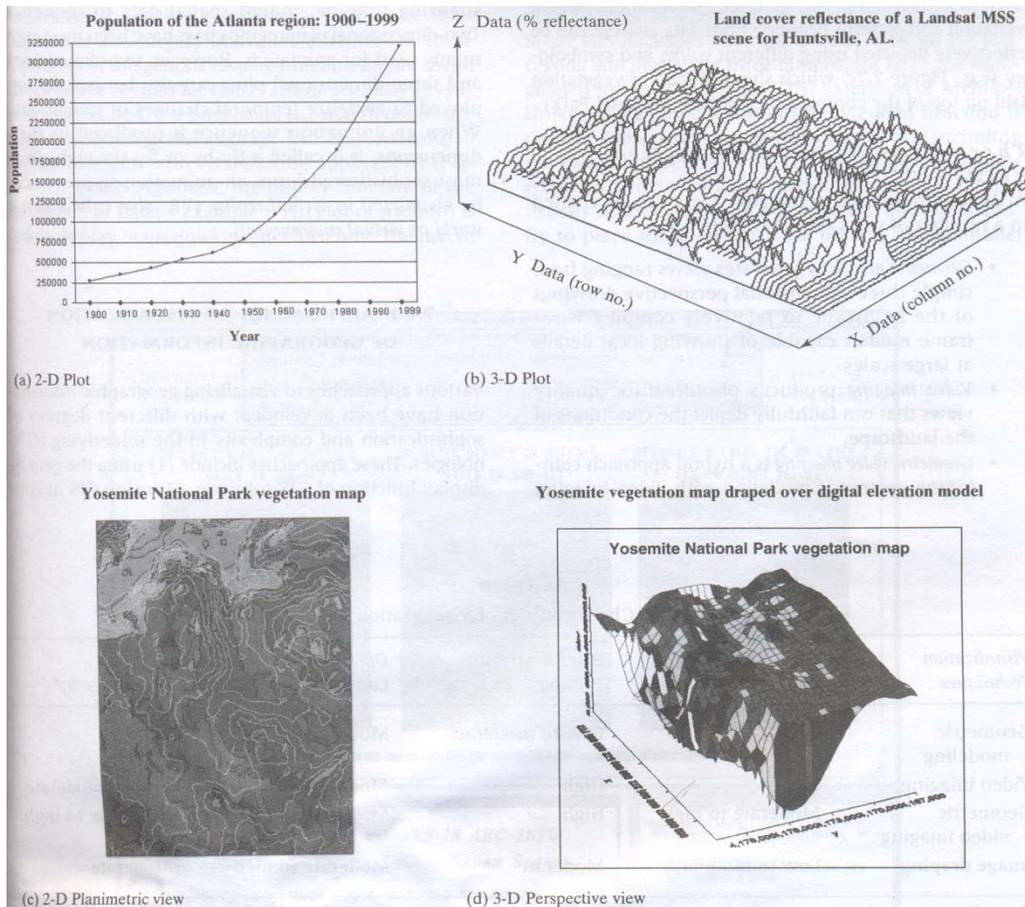


Figure 4.8: Examples of Four Techniques for Visualization of Geographic Information

Two Dimensional plots: These plots are relatively simple to construct and are useful for visualizing the relationship between two numerical variables (e.g. population against time).

Three-Dimensional plots: These plots (sometimes referred to as surface plots) are used to visualize the relationships among three numerical variables (e.g. Landsat MSS scene at each pixel position for a given area).

Two-Dimensional planimetric views: A two-dimensional planimetric view is the technique used in conventional cartographic visualization. Spatial variation and patterns resulting from data analysis can be effectively depicted using different colors and symbology.

Three- Dimensional perspective view: are those that show length, width and height.

Animation: It is the computer graphics techniques for visualizing time-dependent spatial data in sequence.

4.4 Principles of Cartographic Design in GIS

Unlike conventional cartography, maps in GIS can be generated both as soft-copy products on the screen of the computer monitor and hard copy products using a plotter or a printer. The major concepts in cartographic design include (1) use of color, (2) use of text, (3) symbols and symbol sets, and (4) map-to-page transformation.

Use of Color: The primary function of color is to make information on a map visually distinguishable.

Any color on a map can be described by three dimensions: *hue*, the dominant wavelength, which is what we usually think of as "color" such as red, green, or blue; *value* (or lightness), which is the description of how light or dark a color is when holding hue constant; and *saturation*, which is the purity of a hue or the range of wavelengths reflected (the narrower the range the purer the saturation) (Figure 4.9). As a general rule, changes in hue are used to indicate qualitative or nominal differences (e.g. land cover types and administrative units), where as changes in value and saturation are used to represent quantitative or hierarchical differences (e.g.) population density and amount of rainfall).

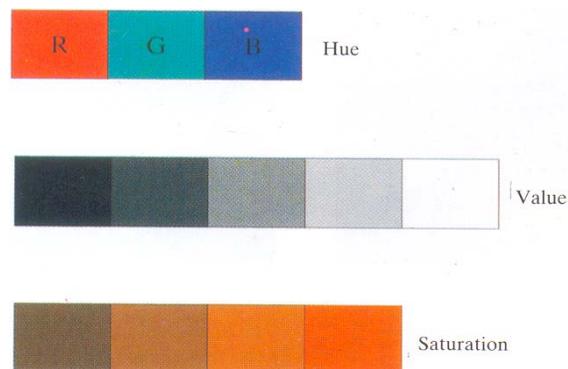


Figure 4.9: Dimensions of Color

Use of Text: Descriptive text is used to give a map its title and to explain the legends. Text can be used in the forms of different character sets that are described by three typographical characteristics, including family, which refers to a set of type faces with variations based on design; face or style, which describes the specific variation based on weight, width and angle; and font, which refers to a character set with a particular face and at a specific size (Figure 4.10).

- **Family:** variation based on design
Times New Roman **Impact** CourierNew **Script** Albertville
- **Face:** variation based on weight, width, and angle
Wide Latin **Brittanic Bold** *italic*
- **Font:** specific size and style
Times New Roman (16) **Times New Roman (24)**

Figure 4.10: Characteristics of Text

Symbols and Symbol Sets: A symbol is a graphic pattern that is used to represent a feature on map. According to the types of features they represent, symbols in GIS are classified in to four categories (Figure 4.11). These include marker *symbols* representing point and node features;

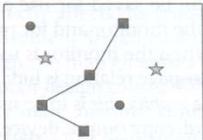
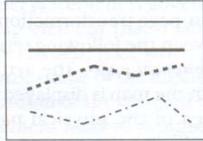
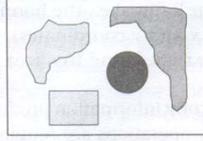
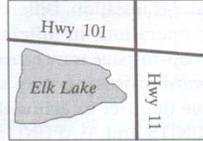
Symbol type	Geographic features
<p>Markers</p> 	Point and node features
<p>Lines</p> 	Line features and polygon boundaries
<p>Shades</p> 	Polygon features
<p>Text</p> 	Annotation and labels

Figure 4.11: Symbols and Symbol Sets

line symbols, representing arcs, routes, and sections using lines of different colors, types, and widths; *shade symbols* for filling polygons and regions using solid color or shade patterns; and text symbols descriptive text used to label features that can be points, lines, and polygons.

Map-To-Page Transformation: Map-to page transformation is the placement of coverage features onto an output medium of a specific size.

Map composition is the process by which maps in GIS are produced. Map composition can be perceived as being made up of three components: (1) Map layout design; (2) geographic contents; and (3) label placement.

Map layout Design: The process of map composition usually starts with an initial layout. There is no single standard to a map layout design, but all output products should include the basic elements as shown in Figure 4.12. Map surround elements (marginal information) included in a given map include title, legend, north arrow, scale, grid, neat line as well as name of the producer of the map.

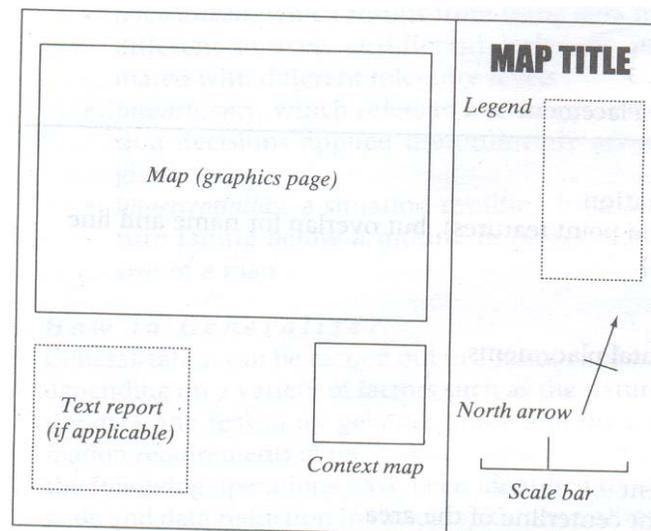


Figure 4.12: Elements of Map Layout

Geographic Contents of the Map: This particular component of cartographic design is concerned with the selection features to be included in a particular map. In essence, the geographic contents of any map are basically governed by three factors: the theme or related themes to be presented, the area to be covered, and the scale of the presentation.

Label Placement: Labels on maps provide the attribute data associated with graphical map elements. Label placement is an important component of cartographic design because it directly affects the readability of the map. A good label placement design enables the reader to associate labels with the map elements that they describe. In contrast, a poor design causes difficulty and uncertainty in using the map.

5. REMOTE SENSING AND GIS INTEGRATION

5.1 Definitions

Three definitions of remote sensing are given below:

- Remote sensing is the science of acquiring, processing and interpreting images that record the interaction between electromagnetic energy and matter.
- Remote sensing is the science and art in obtaining information about an object, area, or phenomenon through the analyses of data acquired by a device that is not in contact with the object, area, or phenomenon under investigation.
- Remote sensing is the instrumentation, techniques and methods to observe the Earth's surface at a distance and to interpret the images or numerical value obtained in order to acquire meaningful information of particular object, on Earth.

Common to the three definitions is that data on characteristics in the Earth's surface are acquired by a device that is not in contact with the objects being measured. The result is usually stored as image data.

5.2 Principles of Electromagnetic Remote Sensing

5.2.1 Introduction

Remote sensing relies on the measurement of electromagnetic (EM) energy. EM energy can take several different forms. The most important source of EM energy at the Earth's surface is the Sun, which provides us, for example, with (visible) light, heat (that we can feel) and UV light, **which can** be harmful to our skin.

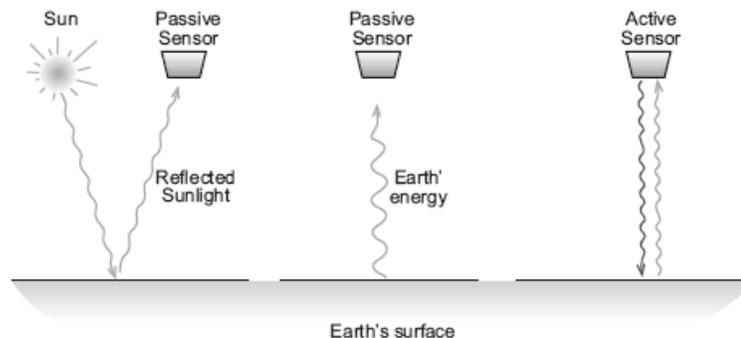


Figure 5.1: A remote sensing sensor measures reflected or emitted energy. An active sensor has its own source of energy.

Many sensors used in remote sensing measure reflected sunlight. Some sensors however, detect energy emitted by the Earth itself or provide their own energy (Figure 5.1). A basic understanding of EM energy, its characteristics and its interactions is required to understand the principle of the remote sensor. This knowledge is also needed to interpret remote sensing data correctly. In between the remote sensor and the Earth's surface is the atmosphere that influences the energy that travels from the Earth's surface to the sensor.

5.2.2 Electromagnetic Energy

5.2.2.1 Waves and Photons

Electromagnetic (EM) energy can be modeled in two ways: by waves or by energy bearing particles called photons. In the wave model electromagnetic energy is considered propagating through space in the form of sine waves. These waves are characterized by two fields: electrical (E) and magnetic (M) fields, which are perpendicular to each other. For this reason, the term electromagnetic energy is used. The vibration of both fields is perpendicular to the direction of travel of the wave (Figure 5.2). Both fields propagate through space at the speed of light c , which is 299,790,000 m/s and can be rounded off to $3 \cdot 10^8$ m/s.

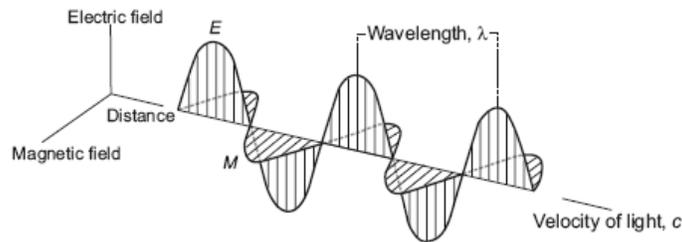


Figure 5.2: Electric (E) and magnetic (M) vectors of electromagnetic wave

One characteristic of electromagnetic waves is particularly important for understanding remote sensing. This is the wavelength λ that is defined as the distance between successive wave *crests* (Figure 5.2). Wavelength is measured in meter, (m) or some factor of meters, such as nanometers (nm, 10^{-9} m) or micrometers (μm , 10^{-6} m). The frequency, ν , is the number of cycles of a wave passing a fixed point over a specific period of time. Frequency is normally measured in hertz (Hz), which is equivalent to one cycle per second. Since the speed of light is constant, wavelength and frequency are inversely related to each other:

$$c = \lambda \cdot \nu$$

5.1

In this equation, c is the speed of light ($3 \cdot 10^8$ m/s), λ is the wave length in m, and ν the frequency (cycles per second, Hz). The shorter the wavelength, the higher the frequency. Conversely, the longer the wavelength the lower the frequency (Figure 5.3)

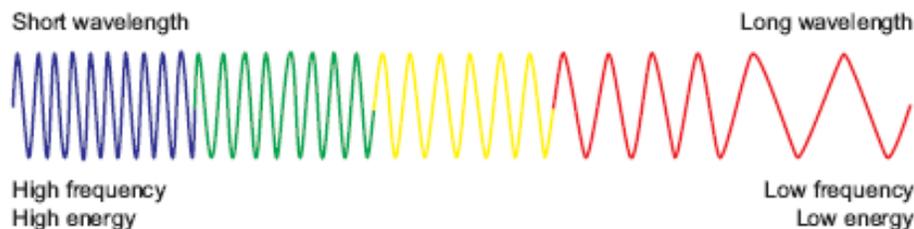


Figure 5.3: Relationship between wavelength, frequency, and energy

Most characteristics of EM energy can be described using the wave model as described above. For some purposes, however, EM energy is more conveniently modeled by the particle theory, in

which EM energy is composed of discrete units called photons. This approach is taken when quantifying the amount of energy measured by multispectral sensor. The amount of energy held by a photon of a specific wavelength is then given by:

$$Q=h*v=h*c/\lambda, \tag{2.2}$$

where Q is the energy of a photon (J), h is Planck's constant ($6.6262*10^{-34}$ Js), and v the frequency (Hz). From Equation 2.2 it follows that the higher the wavelength, the lower its energy content. Gamma rays (around 10^{-9}) are the most energetic, and radio waves (> 1 m) the least energetic. An important consequence for remote sensing is that it is more difficult to measure the energy emitted in longer wavelengths than in shorter wavelengths.

5.2.2.2 Sources of EM energy

All matter with a temperature above absolute zero (0K, where $n^{\circ}\text{C}=n+273$) radiates EM energy due to molecular agitation. Agitation is the movement of the molecules. This means that the Sun, and also the Earth, radiates energy in the form of waves. Matter that is capable of absorbing and re-emitting *all* EM energy is known as a blackbody. For blackbodies, both the emissivity, ϵ , and the absorptance, α , are equal to (the maximum value of) 1. The amount of energy radiated by an object depends on its absolute temperature, its emissivity and is a function of the wavelength ($E = \sigma T^4$ (E is in Watts/m²)). In physics, this principle is defined as Stefan-Boltzmann's law. A black body radiates a continuum of wavelengths. The radiation emitted by a blackbody at different temperatures is shown in Figure 5.4. Note the units in this figure: the x-axis indicates, the wavelength and the y-axis indicate the amount of energy per unit area. The area below the curve, therefore, represents the total amount of energy emitted at a specific temperature. From Figure 5.4 it can be concluded that a higher temperature corresponds to a greater contribution of

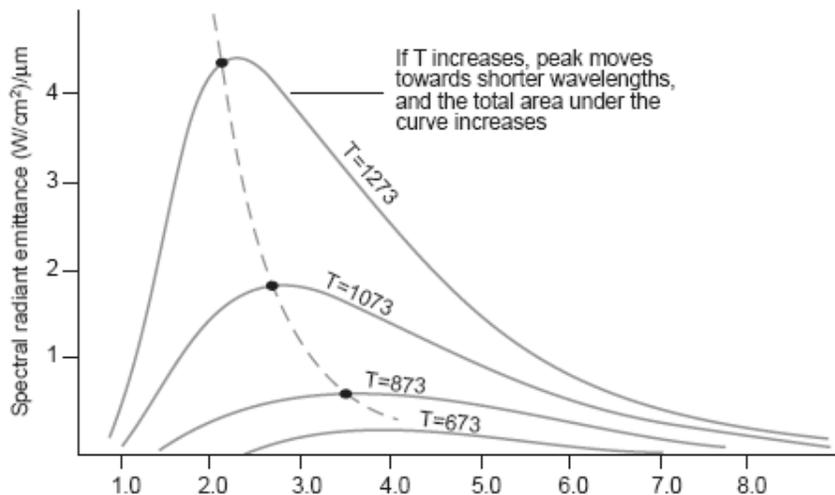


Figure 5.4: Blackbody radiation curves based on Stefan-Boltzmann's law (with temperature in k)

shorter wavelengths. The peak radiation at 400°C is around $4\ \mu\text{m}$ while the peak radiation at 1000°C is at $2.5\ \mu\text{m}$. Figure 5.4 can be more explained by Wein's law which is given by $\lambda_m=A/T$ (λ_m is maximum wave length, $A=2898$, and T temperature). The emitting ability of a real material compared to that of the blackbody is referred to as the material's emissivity. In reality, blackbodies are hardly found in nature; most natural objects have emissivities less than one. This means that only part, usually between 80-98%, of the received energy is re-emitted. Consequently, part of the energy is absorbed. This physical property is relevant in, for example, the modeling of global warming processes.

5.2.2.3 Electromagnetic Spectrum

All matter with a certain temperature radiates electromagnetic waves of various wavelengths. The total range of wavelength is commonly referred to as the electromagnetic spectrum (Figure 5.5). It extends from gamma rays to radio waves.

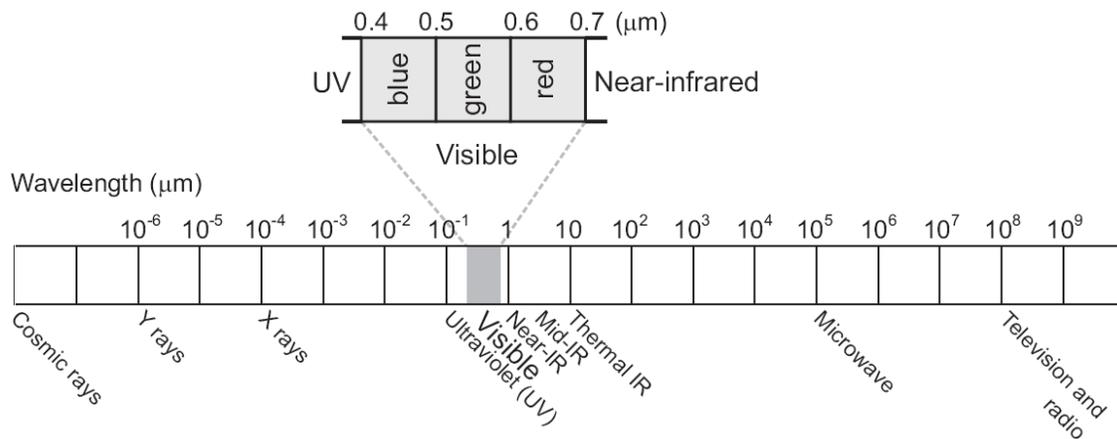


Figure 5.5: The Electromagnetic Spectrum

Remote sensing operates in several regions of the electromagnetic spectrum. The optical part of the EM spectrum refers to that part of the EM spectrum in which optical laws can be applied. These relate to phenomena, such as reflectance and refraction that can be used to focus the radiation. The optical range extends from X-rays ($0.02\ \mu\text{m}$) through the visible part of the EM spectrum up to and including far-infrared ($1000\ \mu\text{m}$). The ultraviolet (UV) portion of the spectrum has the shortest wavelengths that are of practical use for remote sensing. This radiation is beyond the violet portion of the visible wavelengths. Some of the Earth's surface materials, primarily rocks and minerals, emit or fluoresce visible light when illuminated with UV radiation. The microwave range covers wavelengths from $1\ \mu\text{m}$ to $1\ \text{m}$.

The visible region of the spectrum (Figure 5.5) is commonly called 'light'. It occupies a relatively small portion in the EM spectrum. It is important to note that this is the only portion of the spectrum that we can associate with the concept of color. Blue, green and red are known as the primary colors or wavelengths of the visible spectrum. The longer wavelengths used for remote sensing are in the thermal infrared and microwave regions. Thermal infrared gives information about surface temperature. Surface temperature can be related, for example, to the

mineral composition of rocks or the condition of vegetation. Microwaves can provide information on surface roughness and the properties of the surface such as water content.

5.3 Remote Sensing System Classification

In remote sensing, the sensor measures energy. Passive and active techniques are distinguished. Passive remote sensing techniques employ natural sources of energy such as the Sun. Active remote sensing techniques, for example radar and laser, have their own source of energy. Active sensors emit a controlled beam of energy to the surface and measure the amount of energy reflected back to the sensor (Figure 5.1). Passive sensor systems based on reflection of the Sun's energy can only work during daylight. Active sensor systems that measure the longer wavelengths related to the Earth's temperature do not depend on the Sun as a source of illumination and can be operated at any time. Passive sensor systems need to deal with the varying illumination conditions of the Sun which are greatly influenced by atmospheric conditions. The main advantage of active sensor systems is that they can be operated day and night and have a controlled illuminating signal.

5.4 Integration of Remote Sensing and GIS

A fundamental technical requirement for integrating remote sensing with GIS data is the need to have both types of data in the same georeferencing system (i.e. UTM or Geographic Coordinate system). Grate progress has been made toward the total integration of GIS and remote sensing largely because of the rapid advancement in computer hardware and software technology. The raster-vector dichotomy has become blurred as powerful high-speed computer and ingenious computer programming have made the data structure conversion almost transparent. New display technology, which features fly-bys or drapes a vector data set on a raster image, promotes the integration of remote sensing and GIS. The emergence of very high-resolution (submeter) satellite images and high-speed computer power will spearhead the complete integration of remote sensing and GIS.