GEOGRAPHIC INFORMATION SYSTEMS

IS311- Exercises

Information System Department

FACULTY OF COMPUTERS AND INFORMATION
Minia University

Geographic Information Systems (IS201)

Lecture Notes

Collected and edited for
Information System Students
Computers and Information Faculty

Table of Contents

Chapter 1: Introduction ... 3

1.1 GIS ... 3

1.2 Elements of GIS ... 6

1.3 Capabilities of GIS ... 11

1.4 Applications of GIS ... 14

Chapter 2: GIS Data ... 16

2.1 Geographic data Types ... 16

2.2 Graphic Data Models ... 17

2.3 Remote Sensing ... 19

2.4 Geo-data Sources ... 22

Chapter 3: Geographic Data Input & Output Methods and Data Quality. ... 27

3.1 Data Input ... 27

3.2 Data Input Methods. ... 28

3.3 Data Output methods. ... 29

3.4 Data Quality ... 30

Chapter 4: Data Management ... 35

4.1 File Processing Vs DBMS ... 35

4.2 Three Classic Database Models ... 38

4.3 The nature of Geographic Data ... 42

4.4 Spatial Data Models ... 43

4.5 Run-Length Encoding ... 46

4.6 The Quadtree data model ... 47

4.7 The Vector Data Model ... 48

7.8 Limitation of General Purpose DBMS for GIS Applications ... 52

7.9 Approaches Used to Implement a GIS ... 53

Chapter 5: GIS Analysis Functions ... 55

References ...

Chapter 1: Introduction

A geographic information system (GIS) is a computer system for capturing, storing, querying, analyzing, and displaying geospatial data. One of many applications of GIS is disaster management.

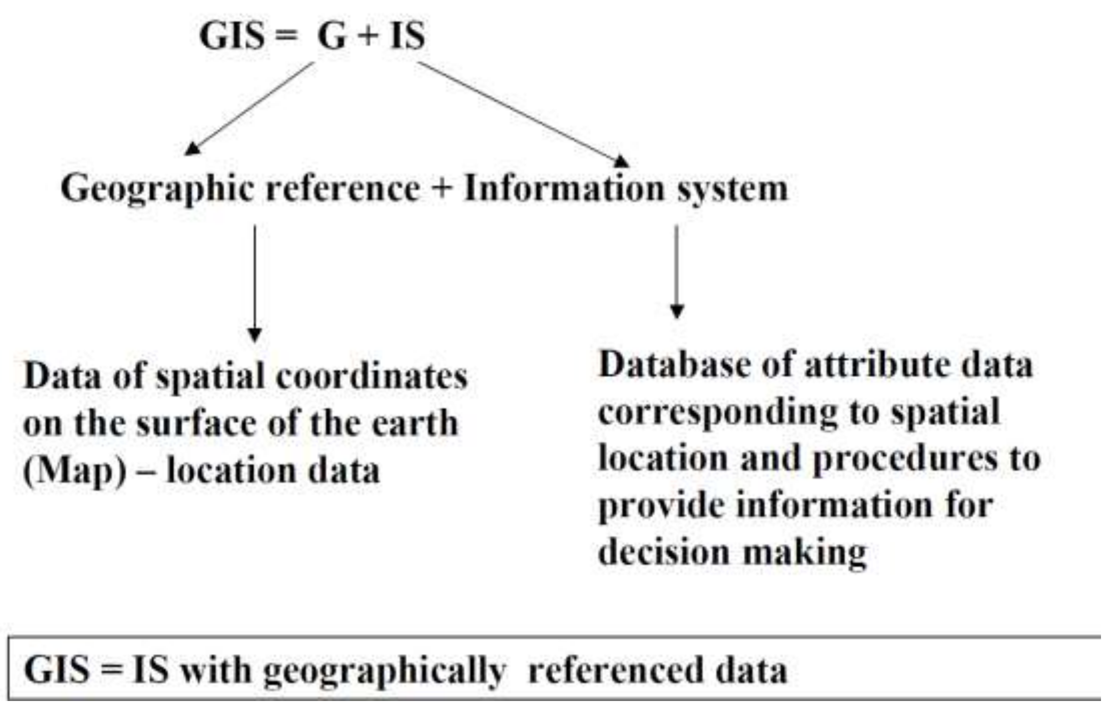
On March 11, 2011, a magnitude 9.0 earthquake struck off the east coast of Japan, registering as the most powerful earthquake to hit Japan on record. The earthquake triggered powerful tsunami waves that reportedly reached heights of up to 40 meters and traveled up to 10 kilometers inland. In the aftermath of the earthquake and tsunami, GIS played an important role in helping responders and emergency managers to conduct rescue operations, map severely damaged areas and infrastructure, prioritize medical needs, and locate temporary shelters. GIS was also linked with social media such as Twitter, YouTube, and Flickr so that people could follow events in near real time and view map overlays of streets, satellite imagery, and topography. In September 2011, the University of Tokyo organized a special session on GIS and Great East Japan Earthquake and Tsunami in the Spatial Thinking and GIS international conference for sharing information on the role of GIS in managing such a disaster.

1.1 GIS

Geospatial data describe both the locations and characteristics of spatial features. To describe a road, for example, we refer to its location (i.e., where it is) and its characteristics (e.g., length, name, speed limit, and direction), as shown in Figure 1.1. The ability of a GIS to handle and process geospatial data distinguishes GIS from other information systems and allows GIS to be used for integration of geospatial data and other data. GIS is used to support decision making processes in public and private institutions around the world and can contribute significantly to the design of administrative and management procedures that are more efficient, transparent and customer-friendly. Any decision that involves spatial aspects such as where or where not to do something can be solved with the help of GIS. Enquiries, which would otherwise take a lot of time, can be automated. Adequate areas for certain activities can be identified as easily as groups of citizens or properties that fall into the same category. Therefore, GIS can significantly increase efficiency and reduce costs. One major out-put of any GIS application is a thematic map visualizing the current situation and/or possible solutions.

DEFINITION OF GIS

A GIS is basically a computerized information system like any other database, but with an important difference: all information in GIS must be linked to a geographic (spatial) reference (latitude/longitude, or other spatial coordinates).



Just to state a few examples of the benefits of GIS:

• High-quality decision making with new possibilities of data analysis
• Faster insight into data
• Better communication between departments/institutions
• Handling of large data volumes
• Increased transparency and efficiency in public procedures
• Better resource allocation
• Needs-oriented regional and municipal planning
• More efficient land tax collection
• Easy identification of appropriate sites for investments and conservation areas

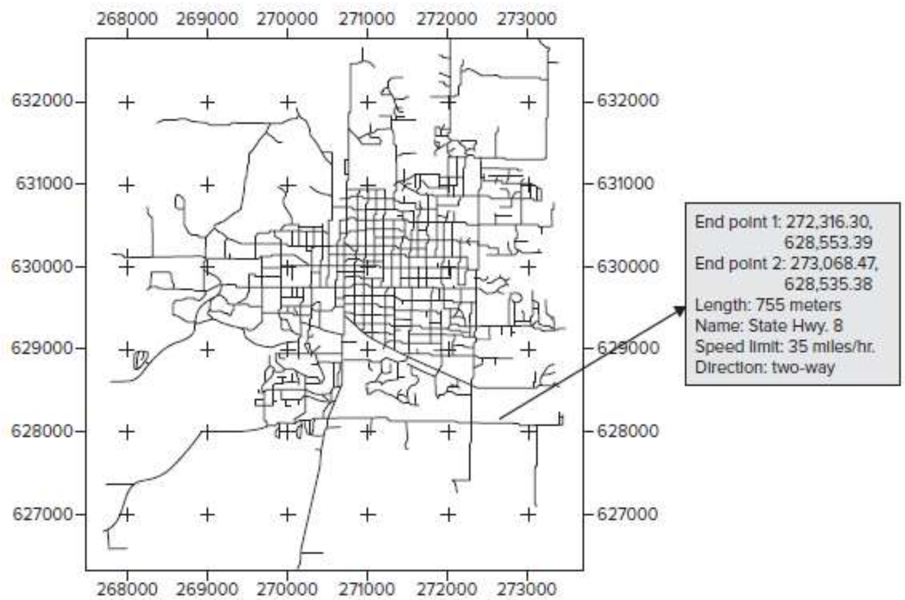


Figure 1.1
An example of geospatial data. The street network is based on a plane coordinate system. The box on the right lists the *x*- and *y*-coordinates of the end points and other attributes of a street segment.

Components of geographic data

1. The phenomenon being reported such as temp, elevation and so on, also called attribute.
2. The spatial location of the phenomenon such as coordinates.
3. Time (change over time)

• Geo-referenced data (Geographic data): spatial data that related to locations on the earth surface.

Components of a GIS

Similar to other information technologies, a GIS requires the following components besides geospatial data:

- Data: sensors, Maps, Records
- Hardware. GIS hardware includes computers for data processing, data storage, and input/ output; printers and plotters for reports and hard-copy maps; digitizers and scanners for digitization of spatial data; and GPS and mobile devices for fieldwork.
- Software. GIS software, either commercial or open source, includes programs and applications to be executed by a computer for data management, data analysis, data display, and other tasks. Additional applications, written in Python, JavaScript, VB.NET, or C++, may be used in GIS for specific data analyses. Common user interfaces to these programs and applications are menus, icons, and command lines, using an operating system such as Windows, Mac, or Linux.
- People. GIS professionals define the purpose and objectives for using GIS and interpret and present the results.
- Methods. A successful GIS operates according to a well-designed plan and business rules, which are the models and operating practices unique to each organization.

1.2 Elements of GIS

Pedagogically, GIS consists of the following elements: geospatial data, data acquisition, data management, data display, data exploration, and data analysis.

1.2.1 Geospatial Data

By definition, geospatial data cover the location of spatial features. To locate spatial features on the Earth's surface, we can use either a geographic or a projected coordinate system. A geographic coordinate system is expressed in longitude and latitude and a projected coordinate system in *x*, *y* coordinates. Many projected coordinated systems are available for use in GIS. An example is the Universal Transverse Mercator (UTM) grid system, which divides the Earth's surface between 84°N and 80°S into 60 zones. A basic principle in GIS is that map layers representing different geospatial data must align spatially; in other words, they are based on the same spatial reference.

A GIS represents geospatial data as either vector data or raster data (Figure 1.3).

- The vector data model uses points, lines, and polygons to represent spatial features with a clear spatial location and boundary such as streams, land parcels, and vegetation stands (Figure 1.4). Each feature is assigned an ID so that it can be associated with its attributes.
- The raster data model uses a grid and grid cells to represent spatial features: point features are represented by single cells, line features by sequences of neighboring cells, and polygon features by collections of contiguous cells. The cell value corresponds to the attribute of the spatial feature at the cell location.

Raster data are ideal for continuous features
such as elevation and precipitation (Figure 1.5).

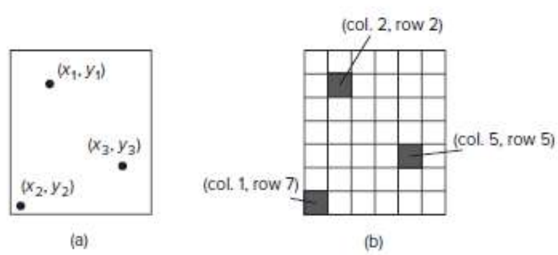


Figure 1.3
The vector data model uses *x*-, *y*-coordinates to represent point features (*a*), and the raster data model uses cells in a grid to represent point features (*b*).

Figure 1.4
Point, line, and polygon features.

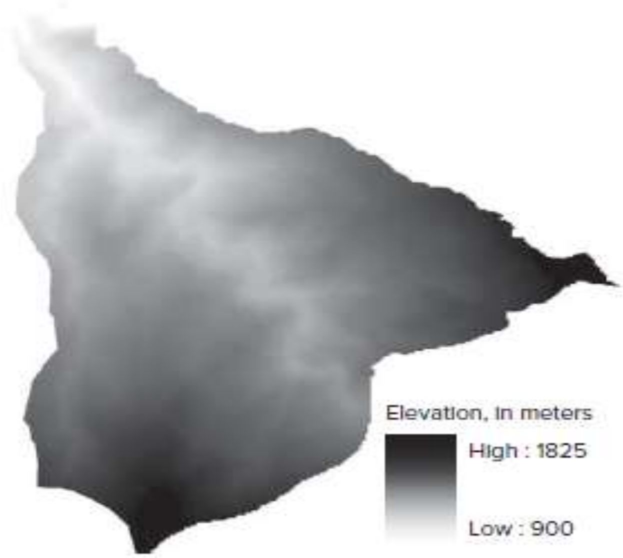


Figure 1.5
A raster-based elevation layer.

1.2.2 Data Acquisition

Data acquisition is usually the first step in conducting a GIS project. The need for geospatial data by GIS users has been linked to the development of data clearinghouses and geoportals. Since the early 1990s, government agencies at different levels in the United States as well as many other countries have set up websites for sharing public data and for directing users to various data sources. To use public data, it is important to obtain metadata, which provide information about the data. If public data are not available, new data can be digitized from paper maps or orthophotos, created from satellite images, or converted from GPS data, survey data, street addresses, and text files with *x*- and *y*- coordinates.

Data acquisition therefore involves compilation of existing and new data. To be used in a GIS, a newly digitized map or a map created from satellite images require geometric transformation (i.e., georeferencing).

Additionally, both existing and new spatial data must be edited if they contain digitizing and/or topological errors.

1.2.3 Attribute Data Management

A GIS usually employs a database management system (DBMS) to handle attribute data, which can be large in size in the case of vector data. Each polygon in a soil map, for example, can be associated with dozens of attributes on the physical and chemical soil properties and soil interpretations. Attribute data are stored in a relational database as a collection of tables. These tables can be prepared, maintained, and edited separately, but they can also be linked for data search and retrieval. A DBMS offers join and relate operations. A join operation brings together two tables by using a common attribute field (e.g., feature ID), whereas a relate operation connects two tables but keeps the tables physically separate. Spatial join is unique in GIS as it uses spatial relationships to join two sets of spatial features and their attribute data, such as joining a school to a county in which the school is located. A DBMS also offers tools for adding, deleting, and manipulating attributes.

1.2.4 Data Display

A routine GIS operation is mapmaking because maps are an interface to GIS. Mapmaking can be informal or formal in GIS. It is informal when we view geospatial data on maps, and formal when we produce maps for professional presentations and reports. A professional map combines the title, map body, legend, scale bar, and other elements together to convey geographic information to the map reader. To make a "good" map, we must have a basic understanding of map symbols, colors, and typology, and their relationship to the mapped data. Additionally, we must be familiar with map design principles such as layout and visual hierarchy.

After a map is composed in a GIS, it can be printed or saved as a graphic file for presentation. It can also be converted to a KML file, imported into Google Earth, and shared publicly on a web server. For time-dependent data such as population changes over decades, a series of map frames can be prepared and displayed in temporal animation.

1.2.5 Data Exploration

Data exploration refers to the activities of visualizing, manipulating, and querying data using maps, tables, and graphs. These activities offer a close look at the data and function as a precursor to formal data analysis. Data exploration in GIS can be mapor feature-based. Map-based exploration includes data classification, data aggregation, and map comparison.

Feature-based query can involve either attribute or spatial data. Attribute data query is basically the same as database query using a DBMS. In contrast, spatial data query allows GIS users to select features based on their spatial relationships such as containment, intersect, and proximity. A combination of attribute and spatial data queries provides a powerful tool for data exploration.

1.2.6 Data Analysis

A GIS has a large number of tools for data analysis. Some are basic tools, meaning that they are regularly used by GIS users. Other tools tend to be discipline or application specific. Two basic tools for vector data are buffering and overlay: buffering creates buffer zones from select features, and overlay combines the geometries and attributes of the input layers (Figure 1.8). Four basic tools for raster data are local (Figure 1.9), neighborhood, zonal, and global operations, depending on whether the operation is performed at the level of individual cells, or groups of cells, or cells within an entire raster.

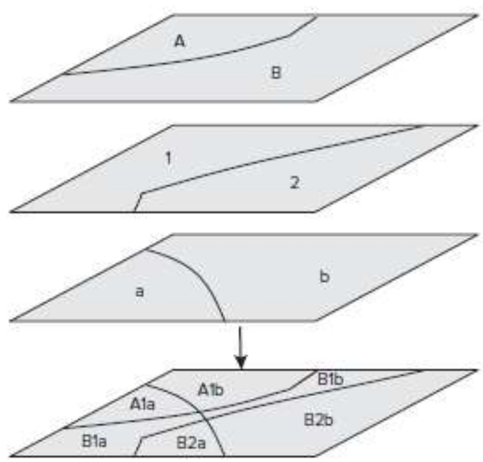


Figure 1.8
A vector-based overlay operation combines geometries and attributes from different layers to create the output.

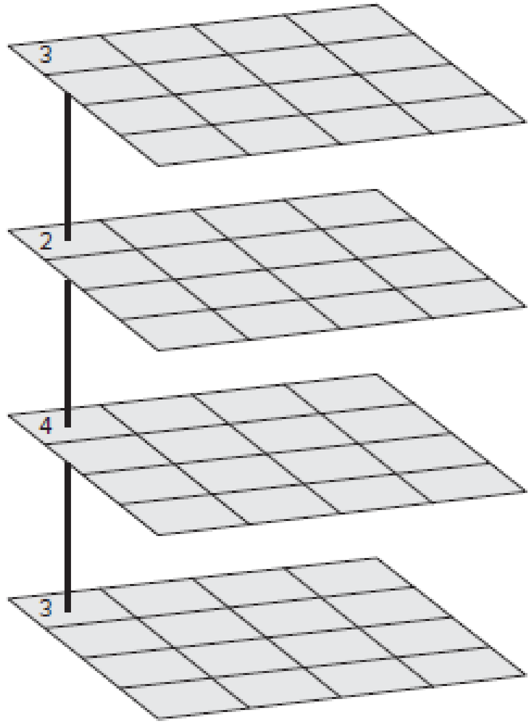


Figure 1.9
A raster data operation with multiple rasters can take advantage of the fixed cell locations. For example, a local average can easily be computed by dividing the sum of 3, 2, and 4 (9) by 3.

1.3 Capabilities of GIS

1. Cartographic capabilities: such as digitizing, graphic display, interactive graphic display and plotting.

2. Data Management capabilities: data storage and retrieval.
3. Analytical capabilities: geo-spatial data analysis.

Maps and Map Scale

- Maps are a way of communicating and visualising geo-referenced information.

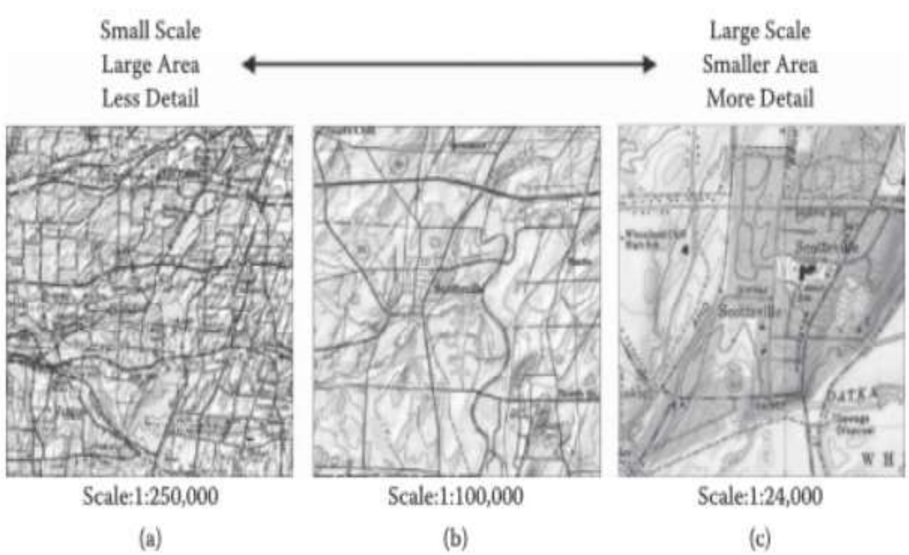
Map Scale: is the ratio (proportion)between measurements on the map and corresponding measurements on the ground.

- The map is a model of the reality.

- **Ways of Representing Map Scale**
    - Representative Fraction: The ratio between the map and the ground. For example 1:100000 or 1/100000 is interpreted as one map unit is equivalent to 100000 units on the ground. Use the same measurement units inches, Kilometres.
    - Verbal Statement: express the scale verbally such as "one inch to sixteen miles"
    - Graphical Bar: Is the most commonly used visual representation of scale.
- **Calculating the ground Distance**

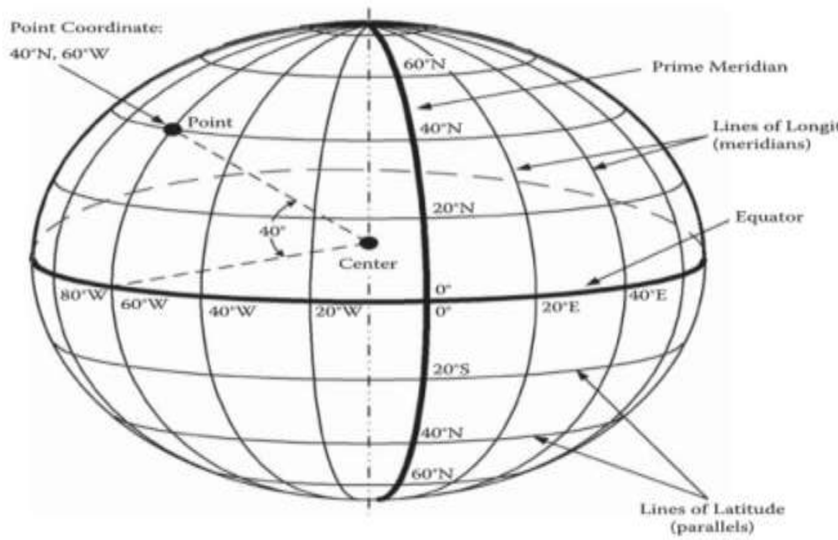
Sm = Dm/Dg Where, Sm is the Map Scale. Dm is the distance on map. Dg is the distance on the ground.

- **Large vs. Small Scale Maps**
    - Small-Scale maps show larger areas with less details.
    - Large-Scale maps: show smaller areas with more details.



Coordinate Systems:

- Geographical Referencing on the earth surface can take many forms such as zip codes, addresses and Latitude and Longitude.
- Lat and Long are spherical coordinates and the basis for generating other 2-D map projections such as UTM.

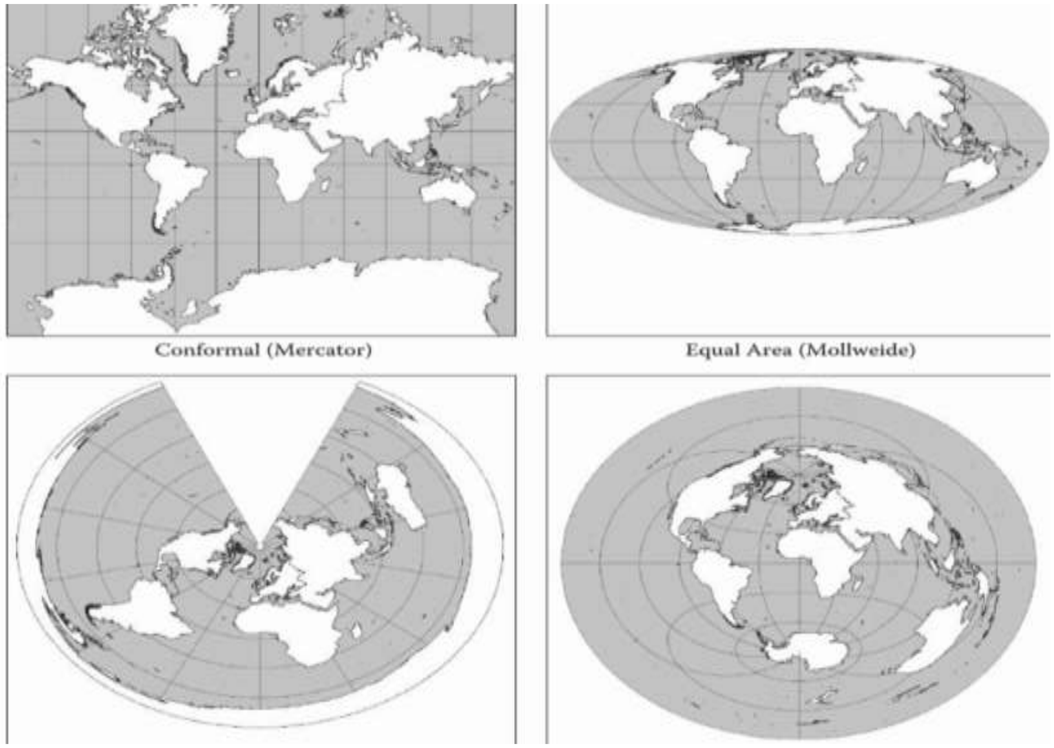


Map Projection

Is a method of representing the earth's three-dimensional surface(Sphere) into a two-dimensional surface.

But we should note that:

1. Projections are mathematical transformations.
2. Scale is true only in certain places.
3. There are many projections.
4. All map projections distort. (lose some details)
5. Some types are better for some applications and for some areas.



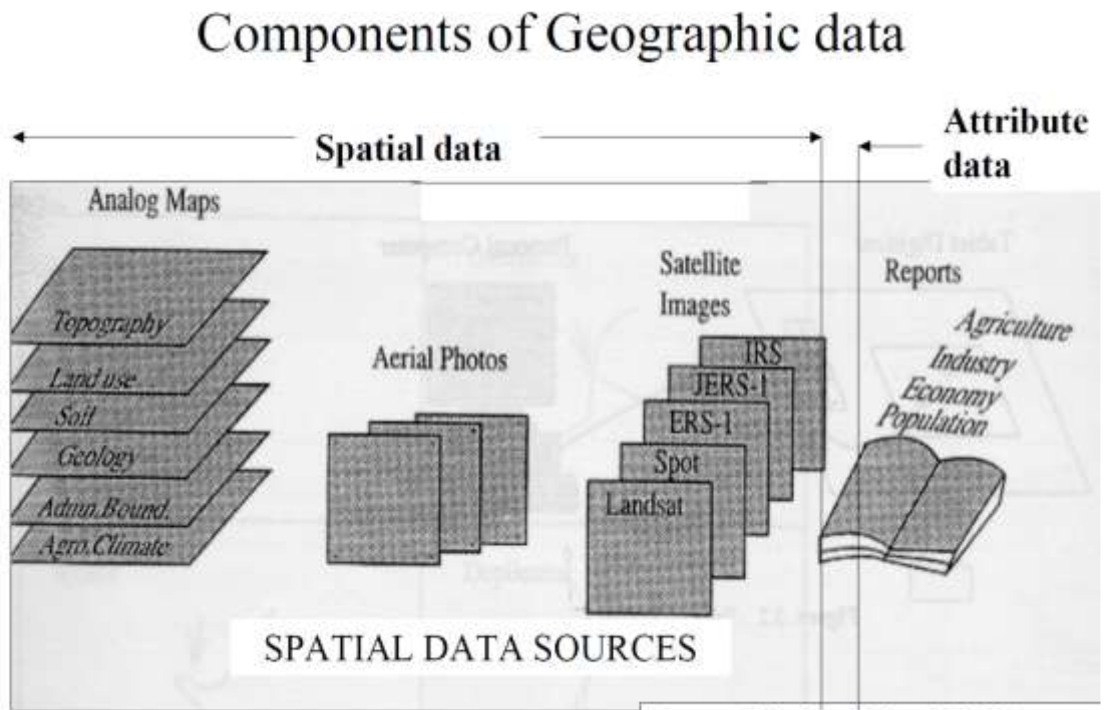
1.4 Applications of GIS

GIS is a useful tool because a high percentage of information we routinely encounter has a spatial component. An often cited figure among GIS users is that 80 percent of data is geographic. To validate the 80 percent assertion, Hahmann and Burghardt (2013) use the German Wikipedia as the data source and report that 57 percent of information is geospatially referenced. Although their finding is lower than 80 percent, it is still strong evidence for the importance of geospatial information and, by extension, GIS and GIS applications. The following are the most popular applications of GIS:

- Renewable resources (water, air, soil, forest).
- Business Applications.
- Election administration.
- Infrastructure Management (security, gas, electricity).
- Maps and data publishing.

- Exploration of Gas, oil, minerals.
- Public health and safety.
- Real estate Management.
- Surveying and Mapping.
- Transportation and Logistics.
- Urban and regional planning.
- Education and Research.

Chapter 2: GIS Data



Source: (Murai and Murai, 1999))

2.1 Geographic data Types

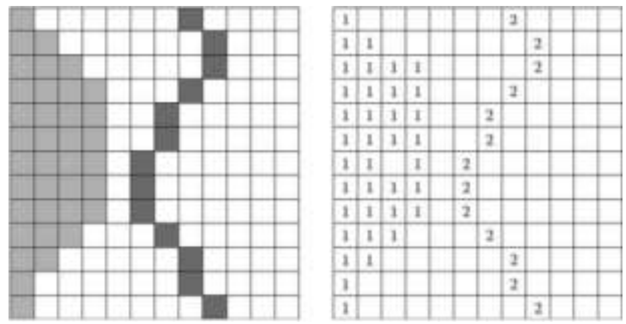
- Geographic data can be classified into two categories:

1. Graphic data: describing features on the map.
2. Non-graphic data: describing values of feature measurements.

Graphic Data

- Point: (●) Is a zero dimensional object that describe a feature on the map such as a house, a tree.
- Line: () is a one dimensional object connecting two points describing a feature on the map such as a river or a road.
- Node: (*) Zero dimensional object describing a topological junction such as intersections of roads.
- Area (Polygon): a feature set of line segments for describing 2 dimensional features such as lake, city boundry.
- Grid/Cell: Each cell has a value.

- Pixel: The same as the cell representation for dealing with images. Each pixel represents the smallest unit in a picture that can not be divided anymore.



2.2 Graphic Data Models

GIS data represents real world objects such as roads, land use, elevation with digital data. Real world objects can be divided into two abstractions: discrete objects (a house) and continuous fields (rain fall amount or elevation).
There are two broad methods used to store data in a GIS for both abstractions:

1. **Vector Data Model:** Defines the geographic features as coordinates such as point, line, area.
2. **Raster Data Model:** Divide the area into a grid of cells. Each cell has a specific value.

Raster
A raster data type is, in essence, any type of digital image. Anyone who is familiar with digital photography will recognize the pixel as the smallest individual unit of an image. A combination of these pixels will create an image, distinct from the commonly used scalable vector graphics, which are the basis of the vector model. While a digital image is concerned with the output as representation of reality, in a photograph or art transferred to computer, the raster data type will reflect an abstraction of reality. Aerial photos are one commonly used form of raster data, with only one purpose, to display a detailed image on a map or for the purposes of digitization. Other raster data sets will contain information regarding elevation, a DEM

(digital Elevation Model), or reflectance of a particular wavelength of light.

Vector

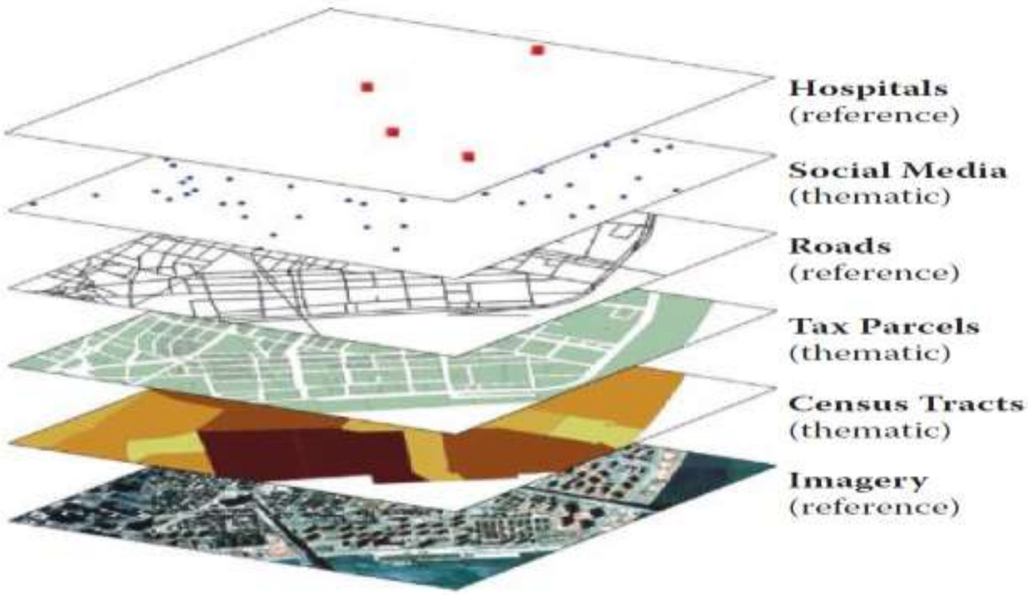
A simple vector map, using each of the vector elements: points for wells, lines for rivers, and a polygon for the lake. In a GIS, geographical features are often expressed as vectors, by considering those features as geometrical shapes. In the popular ESRI Arc series of programs, these are explicitly called shape files.

Nongraphic data

- Attributes: (values)
- Geo-Referenced data: Describe some event or action related to a location on the earth.
- Geographic index: Used as an identifier and sometimes for linking graphic with nongraphic data.
- Spatial relationships: describe relationships between features such as (adjacency, Connectivity, proximity)

Layers

- Is a set of homogenous features registered positionally to other layers in the database.
- The base map (Reference map)is divided into a set of thematic maps. Each one describe only one feature such as temperature, soil type, etc.



2.3 Remote Sensing

Remote sensing is the technique of collecting information from a distance. Data collected from a distance are named Remotely sensed data. Remotely sensed data are normally provided in electronic form, since this is the easiest way to transport the large files representing individual 'scenes' from the various sensors.

Remote Sensing Techniques

1. Photographic Cameras (Aerial Photography).
2. Electro-optical scanners.
3. Land satellite.
4. SPOT (System Pour L'observation de Laterre)

Steps for Analysing Remotely Sensed Data

1. Defining the information needed.
2. Collection of information using remote sensing and other techniques.
3. Information Analysis.

4. Verification of the results of the analysis.
5. Reporting of the obtained results to the users.
6. Taking some action or a decision.

Step 1: Defining the information needed.

Before any data collection, the required information should be specified. Its required accuracy, coverage, time, the data collection cost and the format (either electronic, paper maps or tabulated stats).

Step 2: Collection of information using remote sensing and other techniques.

- Remotely sensed data are not used alone. They are combined with other data sources such as field observations or existing maps.
- The data collection devices detect energy reflected or emitted from earth objects. The reflected light energy is recorded as black-white or coloured images

Step 3: Information Analysis.

There are three basic types of analysis:-

- **Measurement**: Use the sensed values of measurements for calculating environmental conditions such as soil type, moisture, crop conditions.

- **Classification**: Define regions with similar features. They can be coded with the same colour on the map.
- **Estimation**: Is used to estimate the quantity of a material such as the quantity of wheat in an area.

Step 4: Verification of the results of the analysis.

- The verification is important for measuring the level of accuracy of the collected data.
- It is done by taking a sample of points from the map and comparing it with independent measurements such as field surveys. If the accuracy is too low, reject the map.

Step 5: Reporting of the obtained results to the users.

- After checking the quality of the produced information, the information can then be produced in a suitable format. It can be produced as paper maps, annotated images, a computer data file.
- The chosen format should be suitable for the end user of the information.

Step 6: Taking some action or a decision.

The main objective of producing information is helping in decision making. The information is useless if :

  – There is no end user for it.

  – The information never reach the intended user.

– It is not in the suitable format.

2.4 Geo-data Sources

There are many sources for geo-data, the following are the famous of them:

2.4.1 GIS Data Depot

GIS Data Depot (/www.gisdatadepot.com)

- provides free spatial data.
- commercial spatial data that you can purchase.
- You need to be proficient in GIS and have access to GIS software.

2.4.2 Environmental Systems Research Institute (ESRI)

- In addition to being a huge player in the GIS software and consulting industry, ESRI provides a wide range of spatial data for use with its many products.
- ESRI provides the data on DVDs that include HTML help systems that have information about redistribution and a complete set of metadata. ESRI also has an ArcGIS Online Content Sharing Program to allow organizations to share spatial data.

2.4.3 National Geospatial Data Clearinghouse(NGDC)

- A direct outcome of the United States' Federal Geographic Data Committee's (FGDC) collaborative activity was the establishment of a group of about 250 cooperating government bodies in the

United States called the National Geospatial Data Clearinghouse (NGDC).

- Organizations must apply for membership in the network and satisfy the FGDC standards to belong. The EROS Data Centre (a USGS-run geospatial data facility) and the FGDC host the clearinghouse data.
- The NGDC provides well-organized and complete metadata. You can easily access that metadata — but NGDC doesn't commonly offer additional tools.

2.4.4 Centre for International Earth Science Information Network (CIESIN)

CIESIN, housed and operated by Columbia University's Earth Institute, provides data, education, support, and research on data integration. Its interests focus primarily on human and environment interactions, which give CIESIN a decidedly more focused perspective on the datasets it provides and on its mission.

2.4.5 Go-Geo!

- Go-Geo! is a United Kingdom (U.K.) geospatial data port that focuses predominantly on the academic community.
- Funded by the Joint Information Systems Committee (JISC) and operated by the University of Essex and the University of Edinburgh.

- Go-Geo! provides an opportunity to discover, locate, and retrieve data that would otherwise be difficult to find.

2.4.6 Instituto National de Estadistica Geographia e Informatica (INEGI)

INEGI, the National Institute of Statistics, Geography, and Data Processing for Mexico, is a government body that collects and organizes statistical, geographic, and economic information about the country.

2.4.7 **CGIAR Consortium for Spatial Information (CGIAR-CSI)**

CGIAR is an initiative of scientists who have a common interest in international agricultural research. With 15 centres worldwide.

The mission is to apply and advance geospatial sciences, primarily for international sustainable agriculture, natural resource management, conservation of biodiversity, and the alleviation of poverty in developing countries.

2.4.8 **Australian Consortium for the Asian Spatial Information and Analysis Network (ACASIAN)**

- ACASIAN is an applied academic organization interested in the use of GIS databases for Asia (primarily China) and the former Soviet Union.
- 341 datasets are licensed and for sale, and ACASIAN encourages collaborative research that uses its datasets. ACASIAN can provide datasets that include special projections for additional fees.

2.4.9 Geoscience Australia

- Part of the Australian Government's Department of Resources, Energy, and Tourism, Geoscience Australia produces geo-scientific information and knowledge.
- The primary applications are related to decision-making for resource exploration, environmental management, and maintaining the infrastructure critical for the well-being of Australian citizens.
- It provides hardcopy maps and aerial photography, in addition to its digital products.

2.4.10 Canada Geospatial Data Infrastructure

- The Canada Geospatial Data Infrastructure (CGDI) is a combination of technology, geographic data standards, methods, and protocols needed to coordinate and manage Canadian geospatial databases.
- The databases include topographic maps, aerial photography, satellite imagery, and environmental data (including forestry, soil, marine, and biodiversity inventories). They also contain

socioeconomic and demographic (Census) data, and electoral boundaries. The primary purposes for maintaining this managed resource include geospatial analyses to benefit public health, safety and security, the environment.

Chapter 3: Geographic Data Input & Output Methods and Data Quality.

For GIS to be useful it must be able to receive and produce information effectively. Data input and output are the ways that a GIS use to communicate with the world. Input and output technology has seen rapid advances.

3.1 Data Input

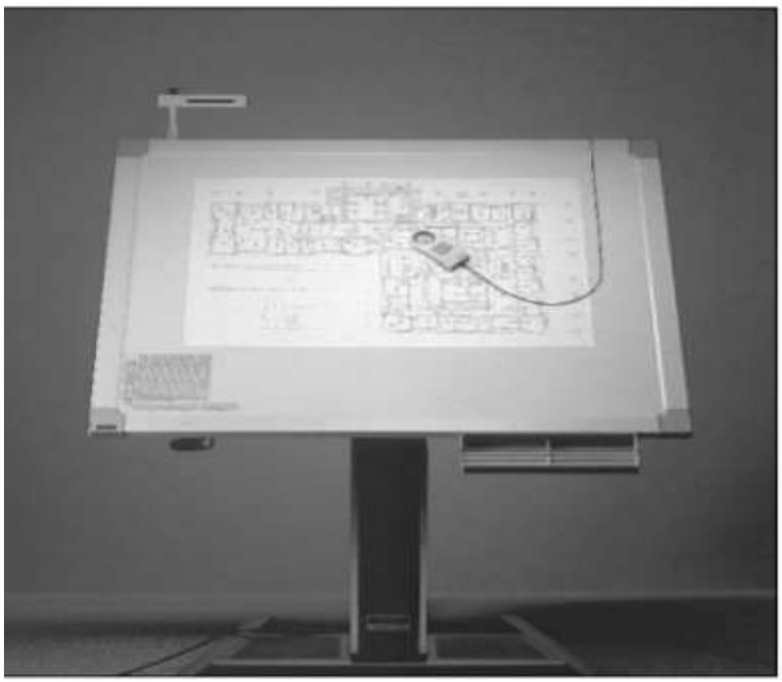
The data input component converts data from their existing form into one that can be used by the GIS. Georeferenced data are commonly provided as paper maps, tab les of attributes, electronic files of maps and associated attribute data, airphotos, and even satellite imagery. The data input procedure can be as straightforward as a file conversion from one electronic format to another, or it can be complex. Data input is typically the major b ottleneck in the implementation of a GIS. Construction of large data b ases can cost five to ten times that of the GIS hardware and software.

- Data input is the procedure of encoding data into a computer-readable form and writing the data to a GIS database.
- The initial cost of building the GIS database is 5 to 10 times the cost of GIS HW and SW.
- The creation of accurate well-documented geo-data is critical to a GIS.
- Documentation describes the quality of the data, which is important to asses its suitability for a particular application.
- Data quality information includes date of collection, accuracy, completeness and the collection method.

3.2 Data Input Methods.

There are many types of methods used for data input of GIS data:

1. **Manual-digitizing:** using digitizing Tablet: Is the most widely used method for map features. The map is mounted to a digitizing table and a pointing device is used to trace the map features. It is time consuming.



2. **Keyboard Data Entry:** The keyboard is used to input the non-spatial data such as field measurements that can be recorded using hand-held computer devices. It is not suitable for entering map features.
3. **Coordinate Geometry Procedures:** Used with spatial features. Collect the geometry (coordinates) of features to create GIS compatible data files. It is more accurate than manual digitizing.
4. **Photogrammetry** (the art and science of obtaining reliable spatial measurements from aerial photography and other remote sensing images).

5. **Scanning:** it is faster than manual digitizing. In scanning you place the map or image edge-first onto the drum inside, and the drum spins while a laser reads (scans) the document. Normally, the scanning software converts the document into a raster image that your GIS software can easily convert to vector format .There are two types of scanners:
   1. Flat-bed scanners.
   2. Drum-scanners.

6. **Translation of existing digital files:** The data might be existing in a digital form, that can be easily converted to another format to be used with the GIS software. The web could be a good source of geo-spatial data.

3.3 Data Output methods.

The output or reporting functions of GISes vary more in quality, accuracy, and ease of use than in the capabilities available. Reports may be in the form of maps, tables of values, or text in hard-copy (such as paper) or softcopy (electronic file). The functions needed are determined by the users' needs, and so user involvement is important in specifying the output requirements.

Output is the procedure by which information from the GIS is presented in a form suitable to the user. Data are output in one of three formats:

1. **Hardcopy:** Permanent means of display.
2. **Softcopy:** Can be viewed on a monitor. It can be interactive.
3. **Electronic:** Computer-compatible files.

Hardcopy Devices

- Pen plotter.
- Line printer black and white.
- Dot matrix printer.
- Colour dot matrix printer.
- Ink jet plotters.
- Thermal plotters.
- Optical film writer.
- Laser printers and plotters.

3.4 Data Quality

The data used in a G1S represent something about the real world at some point in time. They are always an abstraction of reality because we don 't need or want every bit of data, just the ones we think would be useful. The bits we decide to take are the first constraint on the capabilities of the CIS.

Then why not take all the data? First, you could never collect all the data, and second, you wouldn't want all the data even if you could get it. Data are costly to collect. It is costly to collect, store, and sift through large quantities of unnecessary data. Excess data makes it more difficult to use the data you really need. Every expenditure of effort that doesn't contribute to the solution detracts because it represents time, effort, and resources that could have been used elsewhere to improve the analysis. The same argument holds true for the quality of the data.

The most important aspects of data quality are accuracy, precision, time, currency, and completeness. Accuracy measures how often, by how much, and how predictably the data will be correct. Precision measures the fineness of the scale used to describe the data. Time indicates at what point or over what period of time the data were collected. Time can often b e a critical factor of data quality. Some information may quickly become outdated. Currency measures how recently the data were collected. In some cases, the suitability of the data will depend on the season or the year, they were collected. In Canada, for example, summer photography is usually specified for mapping forest cover types. Completeness refers to the portion of the area of interest for which data are available. The term is also used with reference to the classification system that has been used to represent the data.

- **Why Data Quality is important?**

1. Recovery of an error in the data could be more expensive than implementing the whole system.
2. Data error could make a significant error in the output.
3. It is important for gaining the user trust in the system.
4. Knowing the data quality could determine the suitable applications for its use.

Components of Data Quality

- Micro Level components: related to individual data elements.
- Macro Level components: Pertains to the dataset as a whole.
- Usage components: Specific to the resources of the organization.

Micro Level components: include

1. 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.

- It is tested by taking a sample of point coordinates and comparing them to data from an independent data source.
- Positional accuracy has two components:
- Bias: measured by the mean or average error.
- Precision: measured by the STD of the error.
- RMS is an alternative. RMS=SQRT((Xact-Xobs)^2/n)

2. Attribute Accuracy: Variables can be discrete or continuous. Discrete variables can take only finite number of values. Continuous variable can take any value. Similarly, equation can be used to measure its accuracy.

3. Logical Consistency: How well logical relations between data is maintained. For example, data could change over time. These differences are called Sliver.

4. Resolution: The smallest unit represented also called minimum mapping unit.

Macro Level components: includes

1. Completeness: It has three categories Coverage, Classification and Verification.
2. Time: Time is critical in many Geo data. Most of geographic data can change over time.
3. Lineage:
    a. History of data
    b. Source of data
    c. Processing Steps

Usage components: includes

1. Accessibility: Refers to the ease of obtaining and using the data. The accessibility of the data may be restricted because it is privately-held.
2. Direct and Indirect Cost: The direct cost is the cost of buying or collecting the data. The indirect cost includes all the time and materials used to make use of the data.

Sources of Errors

Error is introduced in every step in geo data generation and usage.

1. Errors in data collection

- Inaccuracies in field data collection.
- Errors in existing sources of data (maps).
- Errors in /or misunderstanding of remotely sensed data.

2. Data Input

- Incorrect digitizing of data.
- Errors due to uncertain edges of boundaries.

3.Data Storage

- ✓ Insufficient precision for vector data.
- ✓ Insufficient resolution for raster data.

4. Data Manipulation

- ❑ Boundary Errors.
- ❑ Errors in data classification.
- ❑ Overlay Operations.

5. Data Output

- ❖ Scaling Inaccuracies.
- ❖ Instability of storage media.
- ❖ Errors due to output device.

6. Use of results

- o Errors due to misunderstanding.
- o Inappropriate use.

Chapter 4: Data Management

- Data Management: Is the process of storing and retrieving data.
- There are two approaches for spatial data management:
  1. File Processing.
  2. Database management Environment (DBMS).
- DBMS: Is a set of programs that manipulate and maintain the data in a database.

4.1 File Processing Vs DBMS

- DBMS approach provide data independence. That is, the application program does not need to know how the data is physically stored because all access to the data base is via the DBMS.
- File processing is the most common approach to using a data base. However, it has some serious drawbacks. Since each application program must directly access each data file that it uses, the program must know how the data in each file are stored. This can create considerable redundancy because the instructions to access a data file must be present in each application program. If modifications are made to the data file, these access instructions must be modified in each application program.

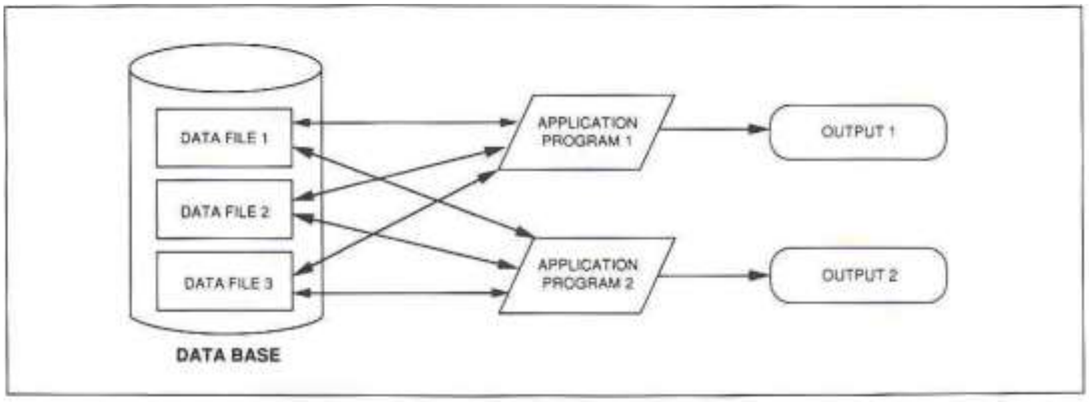


Figure 6.1 Sharing Data Files Among Applications in the File Processing Environment.

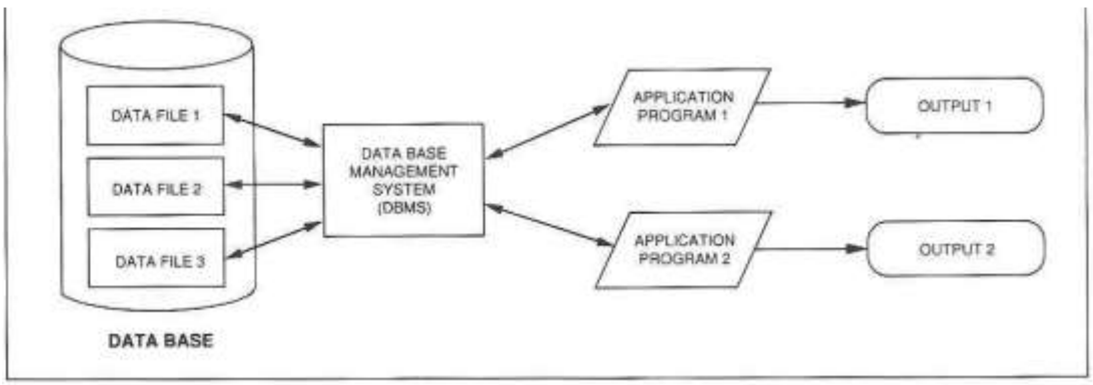
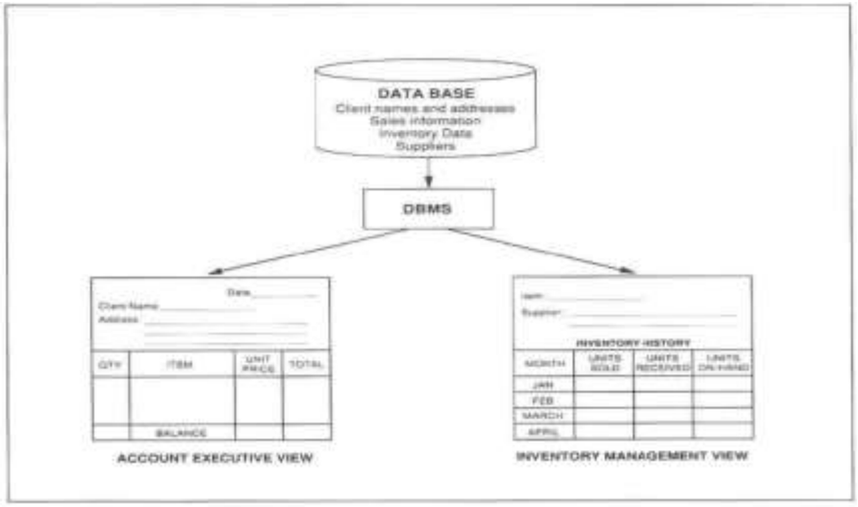


Figure 6.2 Sharing Data Files Among Applications in a Data Base Management System Environment.

- By providing different views, the D BMS tailors the database to each user group — a very valuable function — without storing multiple copies of all the data.



Advantages Of The Database Approach

- Centralized Control. A single DBMS under the control of one person or group can ensure that data quality standards are maintained.
- Data Can Be Shared Efficiently. Using a DBMS, the information in a database can be shared in a flexible yet controlled manner.
- Data Independence. Application programs are independent of the physical form in which the data are stored.
- Easier Implementation of New Data Base Applications. New application programs and unique database searches can be easily implemented using the services provided by a DBMS.
- Direct User Access. Database systems provide a user interface so that non -programmers can perform sophisticated analyses.
- Redundancy Can Be Controlled. In a file processing environment, separate data files are used for each application and data redundancy may result.
- User Views. A DBMS can provide a convenient user interface to create and maintain multiple user views.

Disadvantages Of The Database Approach

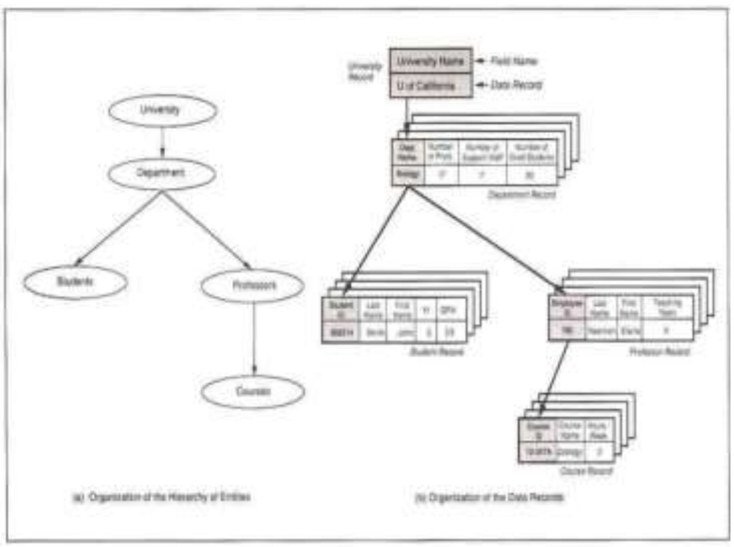
- Cost. The database system software and any associated hardware can be expensive.
- Added Complexity. A database system is more complex than a file processing system.
- Centralized Risk. In centralizing the location of data and reducing data redundancy, there is a greater risk of loss or corruption of data. However, the backup and recovery procedures normally provided in a DBMS minimize these risks.

4.2 Three Classic Database Models

- The conceptual organization o f a database is termed the data model.
- There are three classic data models that are used to organize electronic databases:
  - The Hierarchical Model.
  - The Network Model.
  - The Relational Models.

4.2.1 The Hierarchical Model

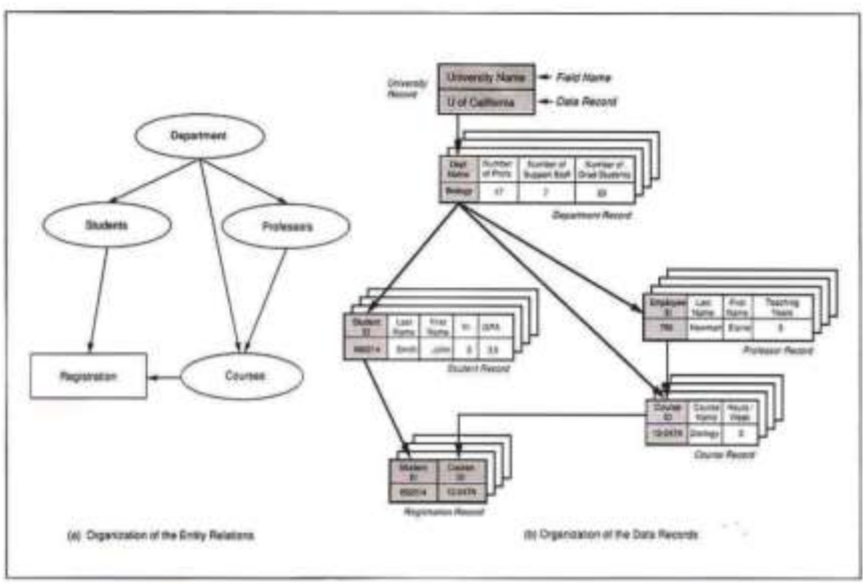
In the hierarchical data model, the data are organized in a tree structure as shown in in the following figure. The relations among the five entities *(University, Department, Students, Professors*, and *Courses)* are defined by the organization of the hierarchy. The organization is encoded in the data records for each entity. The field names are shown in the top half of each box and a sample data record is shown in the lower half. There is one field that is designated as the key field. It is used to organize the hierarchy. The top of the hierarchy is termed the **root.** It is comprised of one entity, in this case a *Unive rs ity* , the University of California. The root may be represented b y a record containing a single data field (as shown here), or by a record containing many fields. Except for the root, every element has one higher level element related to it, termed its **parent,** and one or more subordinate elements, termed **children.** An element can have only one parent but can have multiple children. In the hierarchical data model, every relation is a many-to-one relation or a one-to-one relation. The many departments belong to one university, there are many students in each department, and so on. In the figure the *many* sides of the relation have an arrow head, the *one* side does not. In a hierarchical data model, information is retrieved by traversing the tree structure. Retrieval of all the students or all the professors in a specific department is a very efficient search because there is a direct link between student and department entities and between professor and department entities.



4.2.2 The Network Model

The network data model overcomes some of the inflexibility of the hierarchical model. In the network data model, an entity can have multiple parents as well as multiple child relations and no root is required. As a result, data records can be directly searched without traversing the entire hierarchy above that record. The following figure shows the university database organized as a network model. The *Course* entity can now have two parents and is related to both the *Department* and *Professor* entities. A search of all courses in a specified department can now be done more directly than in the hierarchical example. The *Student -Course* relation is a many-to-many relation. That is, each student can be enrolled in many courses and each course can have many students. While the network model does not allow many-to-many relations, this relation can be handled indirectly by using an intermediate relation, often termed an **intersection record.** As shown in the Figure, the intersection records represent the *Student - Course* combinations, i.e. the registration of students in courses. Each *Student -Course* combination is unique. One *Course* entity can have many *Registration* entities, and one *Student* entity can have many *Registration* entities, so both of these relations are one-to-many and are permitted. Intersection records can also be used in the hierarchical model, although the other restrictions of that model make the implementation somewhat more complex.

Network models tend to have less redundant data storage than the corresponding hierarchical model. However, more extensive linkage information must be stored, adding to the size and complexity of the data files. In a complex database, the linkage information can be substantial and the time needed to update the linkages when changes are made can be significant.



3.2.3 *The Relational Data Model*

The following Figure illustrates the university database organized using the relational data model.

In the relational data model, there is no hierarchy of data fields within a record; every data field can be used as a key. The data are stored as a collection of values in the form of simple records, termed **tuples.**
Each tuple represents a fact, i.e. a set of permanently related values. The tuples are grouped together in two-dimensional tables, with each table usually stored as a separate file. The table as a whole represents the relationships among all the attributes it contains and so it is often termed a **relation.** Using the relational model, a search can be made of any single tab le using any of the attribute fields, singly or together. For example, the Student Information Table can be searched for all students in year 4. Just as easily, the tab le could be searched for all students with the last name Johnson. Searches of related attributes that are stored in different tables can be done by linking two or more tables using any attribute they share in common. This procedure is termed a join operation. The shared attribute need not itself be part of the relation being analyzed.

1. Course Information

Professor ID	Course Dept	Course Name	Course Hours	Course ID
790	Biology	Zoology	3	12-247A
745	Chemistry	Organic	4	14-200A
807	Chemistry	Organic	4	14-200B
642	Chemistry	Biochem	5	14-280A
689	English	Medieval	3	17-340A

2. Registration Information

Course ID	Student ID
12-247A	692214
14-200B	692214
17-340A	692214
17-340A	728437
14-200B	728437
14-280A	728437
14-200B	745870

3. Student Information

Student ID	Last Name	First Name	Yr	GPA	Dept
692214	Smith	John	3	3.5	Biology
728437	Green	John	2	2.4	English
745870	Thomas	Randy	4	3.7	Physics

4. Department Information

Dept Name	Number of Professors	Number of Support Staff	Number of Graduate Students
Biology	17	7	23
Chemistry	10	8	7
English	11	3	20
French	5	1	15
Physics	6	3	8

5. Professor Information

Professor ID	Last Name	First Name	Teaching Years	Dept
745	Brown	Al	5	Chemistry
790	Newman	Elaine	5	Biology
807	Ross	Grant	4	Chemistry
642	Gold	Val	8	Biology
689	Colwell	Bob	8	English

Advantages of the Relational model over the Hierarchical and Network models:

- ✓ The relational model is more flexible than the other models. The way the data are stored in the relational tables does not restrict the kinds of processing that can be done.
- ✓ The organization of the relational model is simple to understand.
- ✓ The same database can generally be represented with less redundancy using the relational model.

4.3 The nature of Geographic Data

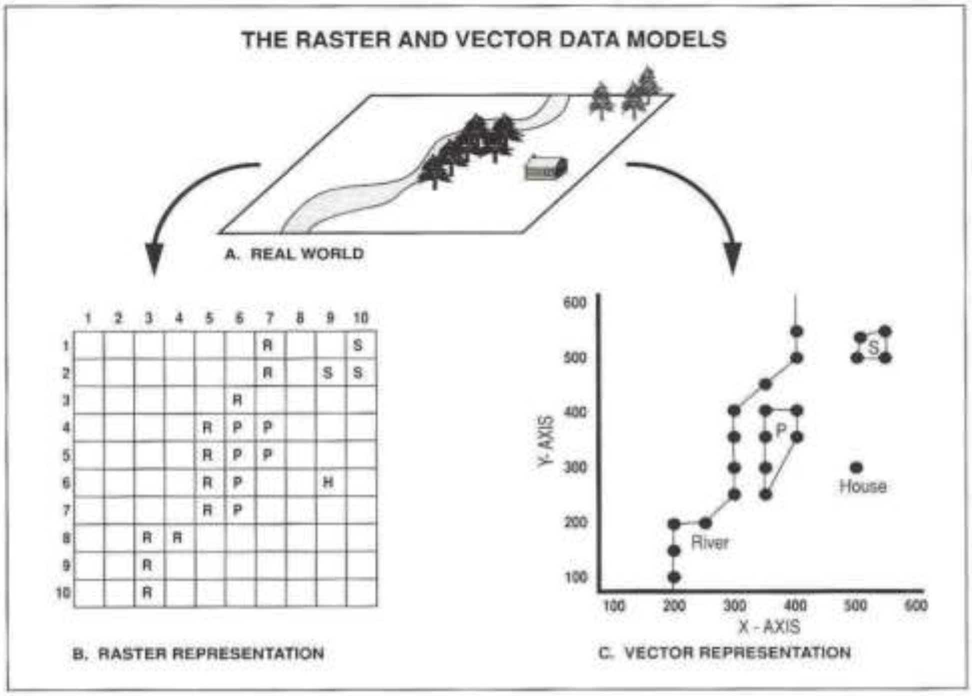
The map is perhaps the most familiar form in which geographic data are represented. A map consists of a group of points, lines, and areas that are positioned with reference to a common coordinate system. It is usually represented in two dimensions so that it is easily portrayed on a flat sheet of paper. The map legend links the non-spatial attributes, such as place names, symbols, and colors, to the spatial data i.e. the locations of the map elements.

The information for a geographic feature has four major components:

1. Locations.
2. Non-spatial attributes.
3. Time.
4. Spatial relationships among geographic features.

4.4 Spatial Data Models

There are two fundamental approaches to the representation of the spatial component of geographic information: the vector model and the raster model. In the vector model, objects or conditions in the real world are represented by the points and lines that define their boundaries, much as if they were being drawn on a map. The position of each object is defined by its placement in a map space that is organized by a coordinate reference system, as shown in the following Figure part C. Every position in the map space has a unique coordinate value. Points, lines, and polygons are used to represent irregularly distributed geographic objects or conditions in the real world. (A polygon is an area bounded b y a closed loop of straight-line segments.) A line may represent a road, a polygon may represent a forest stand, and so on. The spatial entities in the vector model correspond more or less to the spatial entities that they represent in the real world. In the raster model, the space is regularly subdivided into cells (usually square in shape), as shown in the following Figure part B. The location of geographic objects or conditions is defined by the row and column position of the cells they occupy. The area that each cell represents defines the spatial resolution available. Because positions are defined by the cell row and cell column numbers, the position of geographic features is only recorded to the nearest cell.



Raster Vs Vector Models

- ✓ In both models, the spatial information is represented using homogeneous units.
- ✓ In the raster approach, the homogeneous units are the cells. (The area within a cell is not subdivided and the cell attribute applies to every location within the cell.)
- ✓ In the vector approach, the homogeneous units are the points, lines, and polygons. Relative to the raster approach, these units are relatively few in number.

Comparison between Raster & Vector Data Models

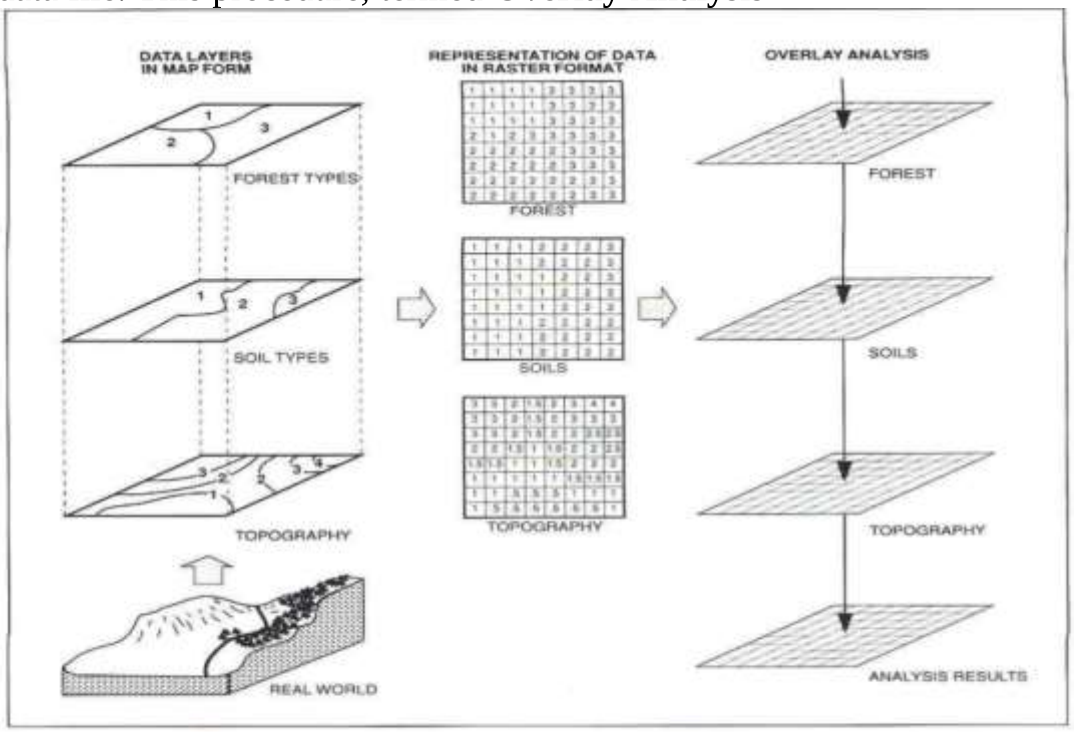
RASTER MODEL	VECTOR MODEL
Advantages:	**Advantages:**
1. It is a simple data structure.	1. It provides a more compact data structure than the raster model.
2. Overlay operations are easily and efficiently implemented.	2. It provides efficient encoding of topology, and, as a result, more efficient implementation of operations that require topological information, such as network analysis.
3. High spatial variability is efficiently represented in a raster format.	3. The vector model is better suited to supporting graphics that closely approximate hand-drawn maps.
4. The raster format is more or less required for efficient manipulation and enhancement of digital images.	
Disadvantages:	**Disadvantages:**
1. The raster data structure is less compact. Data compression techniques can often overcome this problem.	1. It is a more complex data structure than a simple raster.
2. Topological relationships are more difficult to represent.	2. Overlay operations are more difficult to implement.
3. The output of graphics is less aesthetically pleasing because boundaries tend to have a blocky appearance rather than the smooth lines of hand-drawn maps. This can be overcome by using a very large number of cells, but may result in unacceptably large files.	3. The representation of high spatial variability is inefficient.
	4. Manipulation and enhancement of digital images cannot be effectively done in the vector domain.

The Raster Data Model

The raster data model consists of a regular grid of square or rectangular cells. The location of each cell or pixel (for picture element) is defined by its row and column numbers. The value assigned to the cell indicates the value of the attribute it represents.

Overlay Analysis Using Raster Representation

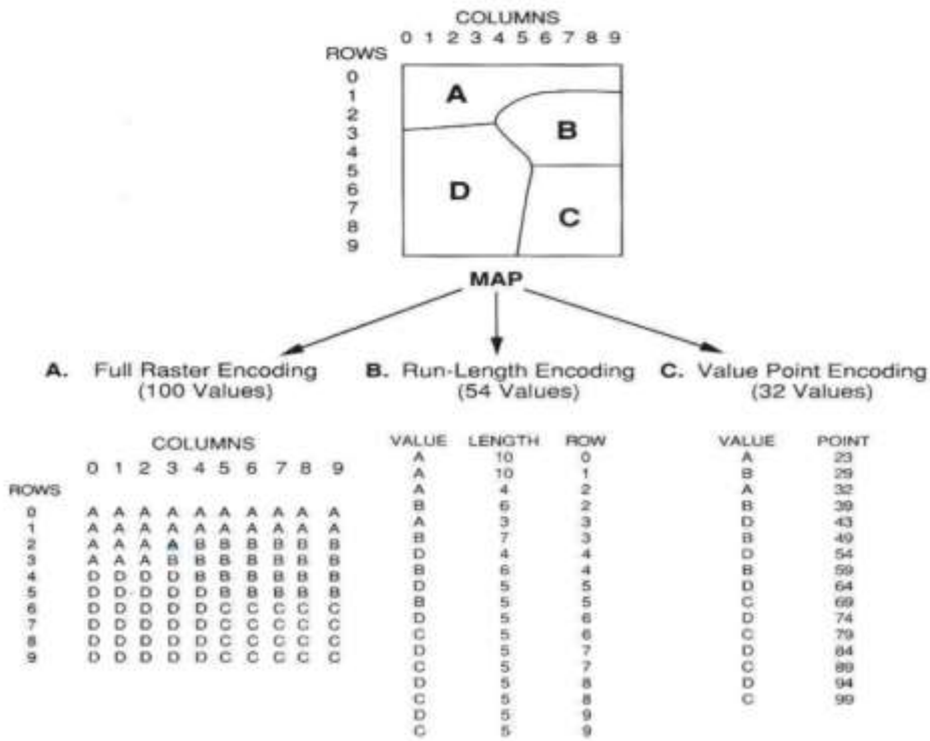
Each cell in a raster file is assigned only one value. So, different attributes are stored in separate files. The soil types and forest cover for an area would be stored as separate soil and forest data files. Operations on multiple raster files involve the retrieval and processing of the data from corresponding cell positions in the different data files. Conceptually, the process is like stacking the files as shown in the following Figure and using the vertical stack of cell values to analyze each cell location. For example, in order to find all the cells with a *Pine* Forest cover and a *Sandy* soil type, each cell in the soil file and each corresponding cell in the forest file would be retrieved and evaluated. All those cells that w ere coded as *Pine* Forest and also as *Sand y* soil would be identified and could be output to a new data file. This procedure, termed **Overlay Analysis**



4.5 Run-Length Encoding

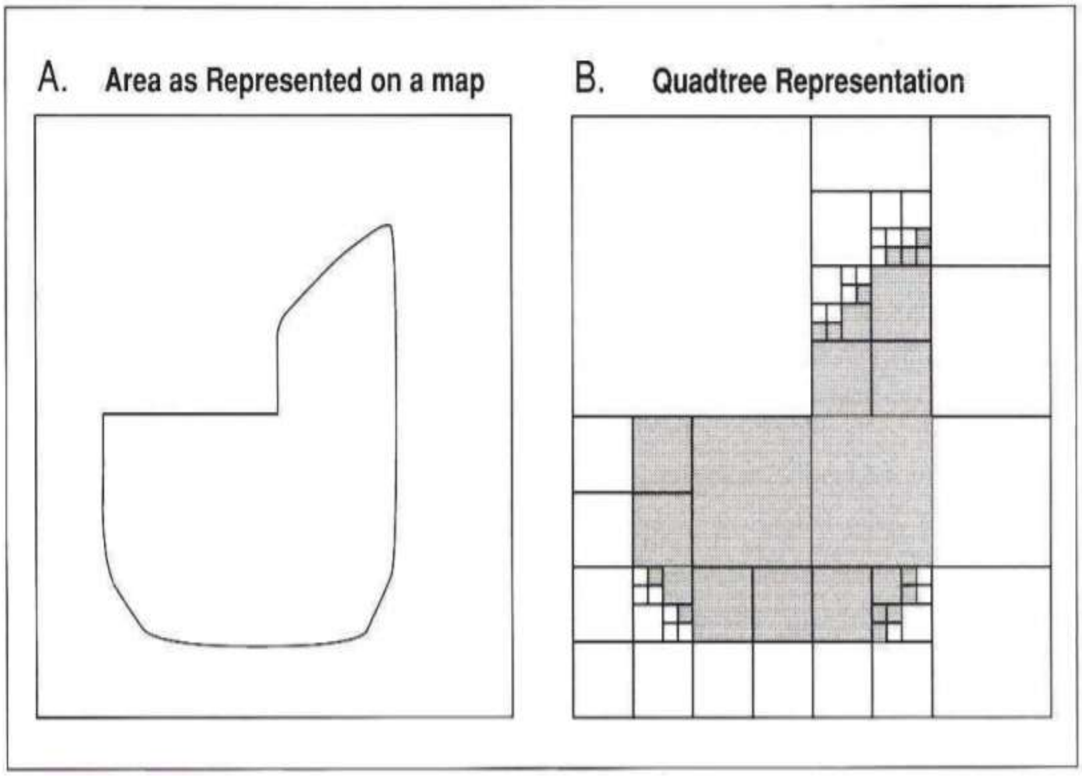
If the data are highly variable from cell to cell, as with digital terrain data or a photographic image, then the large number of cells serve to capture the high spatial variability. If the number of values were reduced, some of the spatial information would be lost. In **run-length encoding,** adjacent cells along a row that have the same value are treated as a group termed a **run.** Instead of repeatedly storing the same value for each cell, the value is stored once, together with information about the size and location of the run. Several run-length encoding strategies have been developed , two of which are illustrated in the following Figure. In **standard run length encoding** the value of the attribute, the number of cells in the run, and the row number are recorded. In this example, the 100 cell values have been reduced to a file of 54 values, see part B.

The second data compression technique shown in part C is termed **value point encoding.** Here the cells are assigned position numbers starting in the upper left corner, proceeding from left to right and from the top to bottom. The position number for the end of each run is stored in the point column. The value for each cell in the run is in the value column. Using value point encoding, only 32 entries w ere needed to encode the same data. However, larger values (requiring more digits and hence more storage) occur in the point column.



4.6 The Quadtree data model

The quadtree data model provides a more compact raster representation by using a variable-sized grid cell. Instead of dividing an area into cells of one size, finer subdivisions are used in those areas with finer details.



4.7 The Vector Data Model

The vector data model provides for the precise positioning of features in space. The approach used in the vector model is to precisely specify the position of the points, lines, and polygons used to rep resent features of interest. The map area is assumed to b e a continuous coordinate space where a position can be defined as precisely as desired. The vector model assumes that position coordinates are mathematically exact. In fact, the level of precision is limited by the number of bits used to represent a single value within the computer, although it is a very fine resolution compared with the cell sizes generally used in raster systems. The location of features on the earth's surface are referenced to map positions using an XY coordinate system (termed a **Cartesian coordinate system).** Geographic features are commonly record ed on two dimensional maps as points, lines, and areas. The vector model uses a similar approach. A point feature is recorded as a single XY coordinate pair, a line as a series of XY coordinates, and an area as a closed loop of XY coordinate pairs that define the boundary of the area. (An area bounded b y a closed loop of straight-line segments is termed, a **polygon.)**

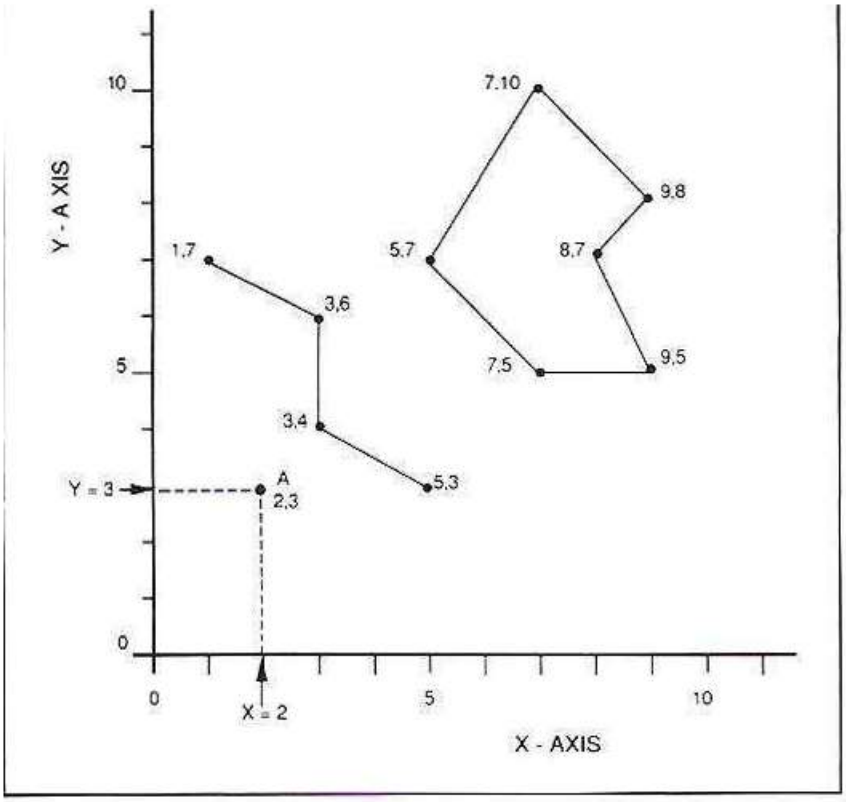
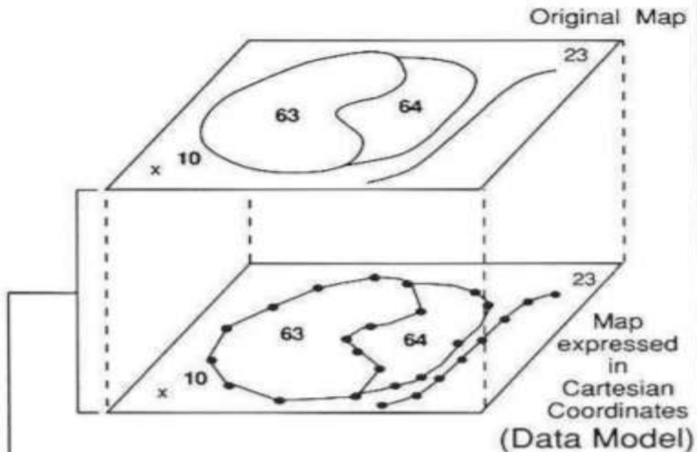


Figure 6.15 Representing Points, Lines, and Polygons as XY Coordinate Strings.

4.7.1 The Spaghetti Data Model

The spaghetti data model is illustrated in Figure 6.16. In this model the pa per map is translated line- for- line into a list of XY coordinates. A point is encoded as a single XY coordinate pair and a line as a string of XY coordinate pairs. An area is represented b y a polygon and is recorded as a closed loop of XY coordinates that define its boundary. The common boundary between adjacent polygons must be recorded twice, once for each polygon. A file of spatial data constructed in this manner is essentially a collection of coordinate strings with no inherent structure — hence the term **spaghetti model.** The structure of this model is very simple and easy to understand. The data model is really the map expressed in Cartesian coordinates. The data file of XY coordinates is actually the **data structure,** the form in which the spatial data are stored in the computer. Although all the spatial features are recorded, the spatial relationships between these features are not encoded.

THE "SPAGHETTI" DATA MODEL

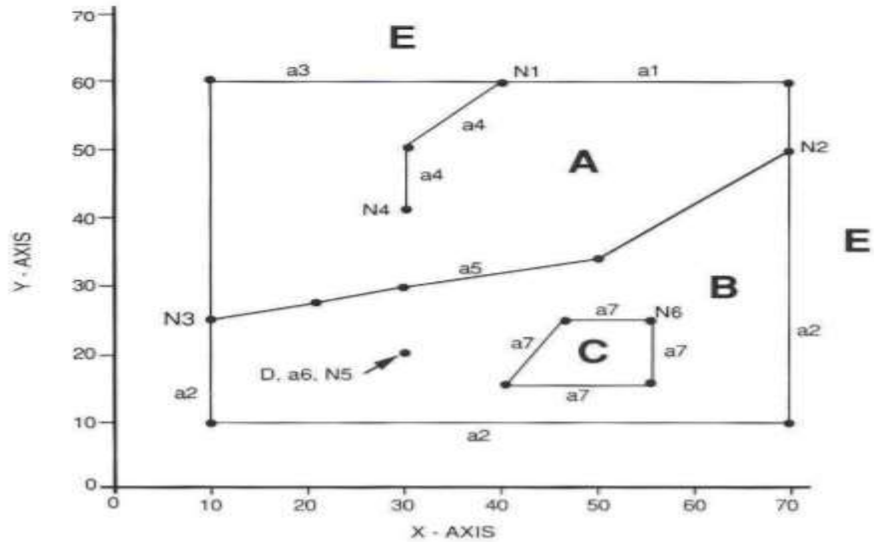


Data Structure

FEATURE	NUMBER	LOCATION
Point	10	X Y (Single Point)
Line	23	$X_1\ Y_1,\ X_2\ Y_2$ ...... $X_n\ Y_n$ (String)
Polygon	63	$X_1\ Y_1,\ X_2\ Y_2$ ...... $X_1\ Y_1$ (Closed Loop)
	64	$X_1\ Y_1,\ X_2\ Y_2$ ...... $X_1\ Y_1$ (Closed Loop)

4.7.2 The Topological Data Model

The topological model is the most widely used method of encoding spatial relationships in a GIS. **Topology** is the mathematical method used to define spatial relationships.



SPATIAL DATA ENCODING

POLYGON TOPOLOGY	
POLYGON	ARCS
A	a1, a5, a3
B	a2, a5, 0, a6, 0, a7
C	a7
D	a6
E	area outside map coverage

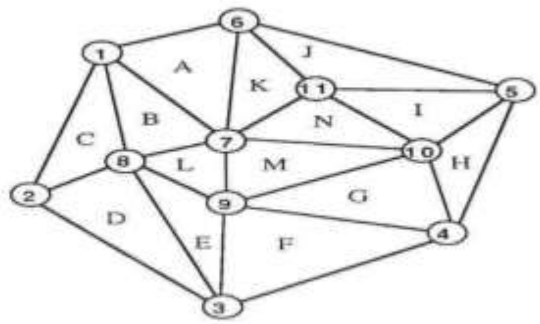
NODE TOPOLOGY	
NODE	ARCS
N1	a1, a3, a4
N2	a1, a2, a5
N3	a2, a3, a5
N4	a4
N5	a6
N6	a7

ARC TOPOLOGY				
ARC	START NODE	END NODE	LEFT POLYGON	RIGHT POLYGON
a1	N1	N2	E	A
a2	N2	N3	E	B
a3	N3	N1	E	A
a4	N4	N1	A	A
a5	N3	N2	A	B
a6	N5	N5	B	B
a7	N6	N6	B	C

ARC COORDINATE DATA			
ARC	START X, Y	INTERMEDIATE X, Y	END X, Y
a1	40, 60	70, 60	70, 50
a2	70, 50	70,10; 10,10	10, 25
a3	10, 25	10,60	40, 60
a4	40, 60	30,50	30, 40
a5	10, 25	20,27; 30,30; 50,32	70, 50
a6	30, 20		30, 20
a7	55, 27	55,15; 40,15; 45,27	55, 27

4.7.3 The Triangulated Irregular Network-TIN

The Triangular Irregular Network or TIN is a vector- based topological data model that is used to represent terrain data. A TIN represents the terrain surface as a set of interconnected triangular facets. For each of the three vertices, the XY coordinate (geographic location) and the Z coordinate (elevation) values are encoded.



X–Y COORDINATES

node	coordinates
1	x1, y1
2	x2, y2
3	x3, y3
:	
11	x11, y11

Z COORDINATES

node	coordinate
1	z1
2	z2
3	z3
:	
11	z11

EDGES

Δ	adjacent Δ
A	B,K
B	A,C,L
C	B,D
D	C,E
E	D,F,L
F	E,G
G	F,H,M
H	G,I
I	H,J,N
J	I,K
K	A,J,N
L	B,E,M
M	G,L,N
N	I,K,M

NODES

Δ	node
A	1,6,7
B	1,7,8
C	1,2,8
D	2,3,8
E	3,8,9
F	3,4,9
G	4,9,10
H	4,5,10
I	5,10,11
J	5,6,11
K	6,7,11
L	7,8,9
M	7,9,10
N	7,10,11

4.8 Limitation of General Purpose DBMS for GIS Applications

The data model most widely accepted for handling non-spatial attribute data in GIS applications has been the relational model. The functions need ed to handle geographic data are not done well in the tabular database environment of a standard relational DBMS. Some of the major difficulties are listed below:

1. The spatial data records used in a GIS are varia b le length records which are needed to store variable numbers of coordinate points, whereas general purpose data base systems are designed to handle fixed length records. Also, a rather complex top ology that is interrelated with the spatial coordinate data must be correctly

maintained. To provide these database functions, additional software is needed to extend the capabilities of the relational DBMS.
2. Manipulation of geographic data involves spatial concepts, such as proximity, connectedness, containment, and overlay, that are not easily accommodated by general purpose database query languages.
3. A G1S requires sophisticated graphics capabilities that are not normally supported b y a general purpose DBMS.
4. Geographic information is complex. The representation of a single geographic feature requires multiple records in multiple files. It may involve geodetic networks, feature coordinates, topology, measurements of spatial features, keys to the non-spatial attribute data, and the non-spatial attributes themselves.
5. The highly interrelated nature of G1S data records requires a more sophisticated security system than the record locking approach used by general purpose DBMS. To ensure the integrity of the geographic database, the security system must protect the integrity of the multiple files in which the spatial data are stored. A change in one record can create multiple errors in multiple files.

4.9 Approaches Used to Implement a GIS

Over the past decade, various practical approaches were taken to provide data management services for a GIS. They may be broadly grouped into the following four, somewhat overlapping, strategies (see Figure 6.22):

1. Develop a proprietary system providing the individual data management services required by the different application modules. This is the file processing approach.
2. Develop a hybrid system using a commercially available DBMS (usually a relational one) for storage of the nonspatial attributes. Develop separate software to manage the storage and analysis of the spatial data, using the services of the relational DBMS to access the attribute data.
3. Use an existing DBMS, usually a relational one, as the core of the GIS. Then develop extensions to the system where needed. Although the spatial and attribute data may be managed by the DBMS, a significant amount of software is generally a d d ed to the DBMS to provide the spatial functions and graphics display used in geographic analysis.
4. Start from scratch and develop a spatial database capable of handling

the spatial and non-spatial data in an integrated fashion.

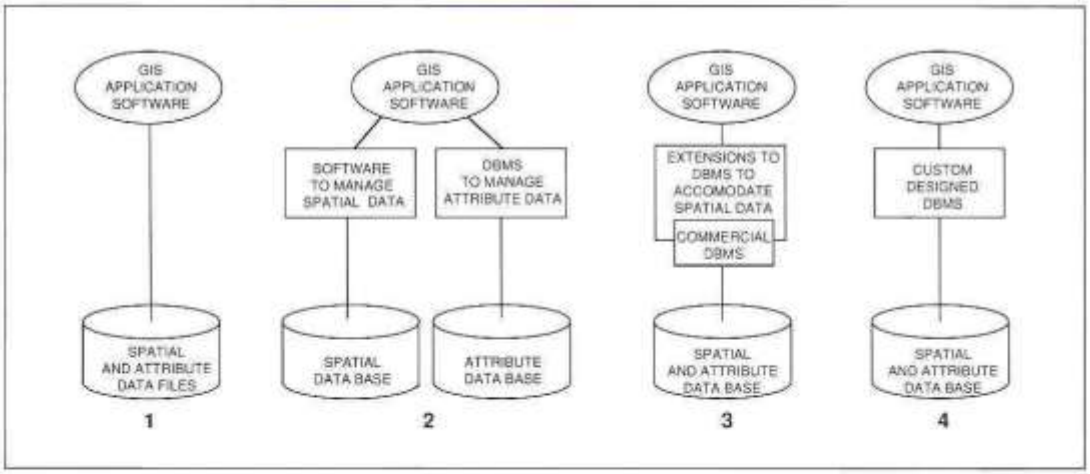


Figure 6.22 Four Approaches to GIS System Design.

Chapter 5: GIS Analysis Functions

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 data base to answer questions about the real world.

To develop the best answers from the information available requires a systematic framing of the questions to be addressed. There is often a strong ten den cy to begin the analysis with only a general idea of the questions to be answered and the data needed. This is a particularly common tendency whenever a new computer-based system, such as a GIS, is introduced.

The answers provided b y a G1S can be categorized into three types as illustrated in Figure 7.1:

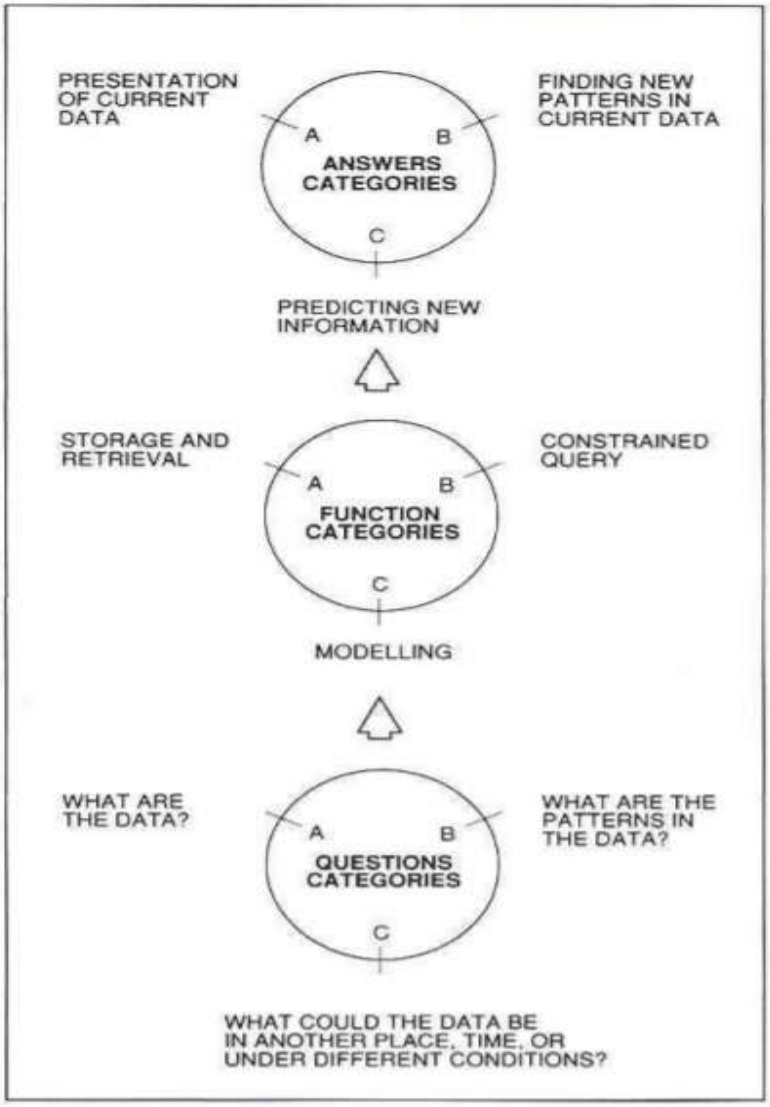
1. a presentation of the current data, i.e. the data in the database such as a map of the city streets,
2. a pattern in the current data, such as all houses valued at over $100,000, and;
3. a prediction of what the data could be at a different time or place. For example, predicting the services that would be lost in the event of an earthquake. This type of analysis might be used to develop emergency response plans.

The types of questions to be answered can also be characterized by three categories:

1. What are the data? i.e., what is the information currently stored in the data base. For example, what is the name and address of the owner of a specified property?
2. What is the pattern in the data? This type of question is a search for entities that possess a specified set of characteristics. For example, plotting a map of all lots with houses valued at over $100,000 would be

defining a pattern in the data that may not be obvious when all the data are view ed together the pattern for this type of house.
3. What could the data be? This type of question implies that a predictive model will be used. The model may be as simple as predicting that a field will produce the same crop next year as this year. It may be as complex as predicting the change in stream flow after a forest has been removed from a watershed.

The functions used to produce these answers can similarly be categorized by the types of answers they provide;
1. storage and retrieval functions,
2. constrained query functions, and
3. modelling functions.



These categories of questions, functions, and answers are not mutually exclusive. A given answer, function, and question will have aspects of each category. This is rep resented conceptually by their position in the circles. The following examples illustrate this idea. The letter representing each example is shown in the Figure.

A. Retrieving the street map for an area primarily involves the retrieval of existing information. The map exists in the G1S and need only be recalled and output.
B. Retrieving those lots with houses valued over $ 100,000 is an example of the second category of use. H

ere the value of each lot must be considered and only those satisfying the constraint are accepted.

C. Determining the optimum routing of a powerline is an example of modelling. Multiple layers of information are used together to weigh different alternatives in order to optimize the design.

5.1 Classification of GIS Analysis Functions

1. Maintenance and analysis of spatial data.

2. Maintenance and analysis of attribute data.

3. Integrated analysis of spatial and attribute data.

4. Output formatting.

5.1.1 Maintenance and analysis of spatial data.

- Format transformation: Transforming the data into a suitable format for the GIS (raster or vector).

- Geometric transformation: Used to assign coordinates to a map or a data layer so that it can be correctly overlayed on another of the same area. This process is called Registration.

- Transformation between map projections: The data layers to be used together for analysis by GIS should be using the same map projection.

- Conflation: is the process of reconciling the position of the features in different data layers to overlay precisely.

- Edge matching: a procedure to adjust the position of features that extend across map sheet boundaries.

- Editing Functions: are used to add, delete and change the geographic position of features.

- Slivers or splinters are thin polygons that are often created during digitizing and overlay operations.

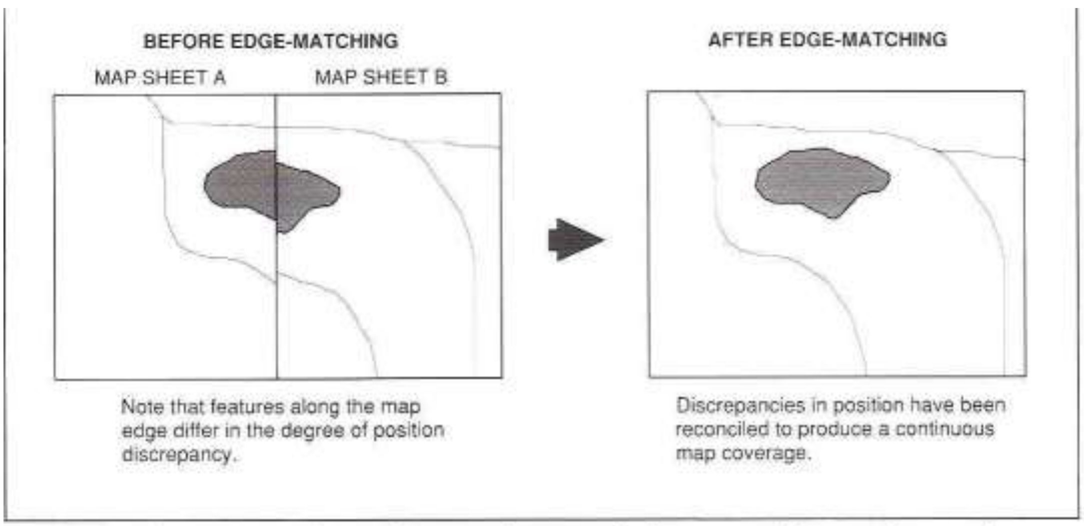


Figure 7.8 Edge Matching. Edge matching is used to reconcile the position of features that extend onto an adjacent map but are not correctly aligned at the map boundary.

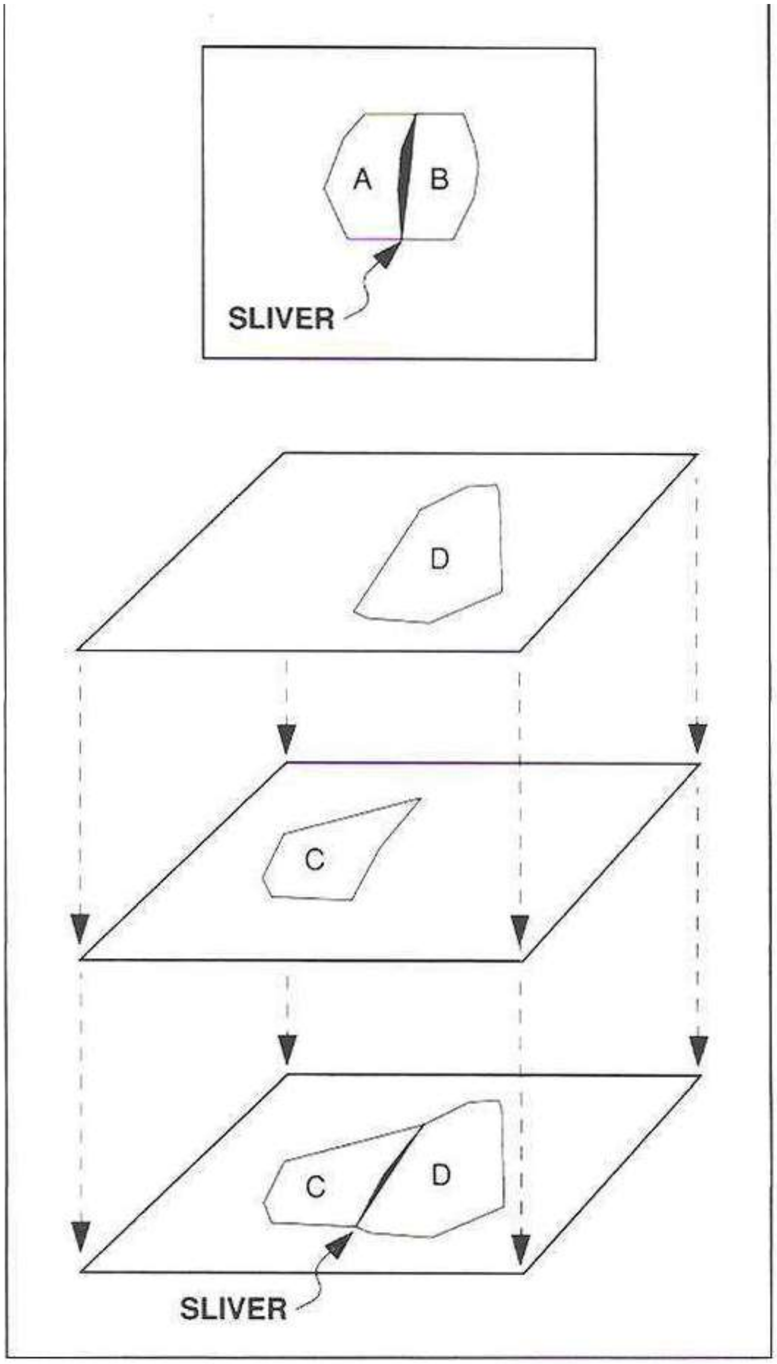


Figure 7.9 Slivers. Slivers may be created during digitizing and overlay operations.

- Line coordinate thinning: used to reduce the quantity of coordinate data that must be stored by GIS. The thinning function reviews all the coordinates and removes the un-necessary ones.

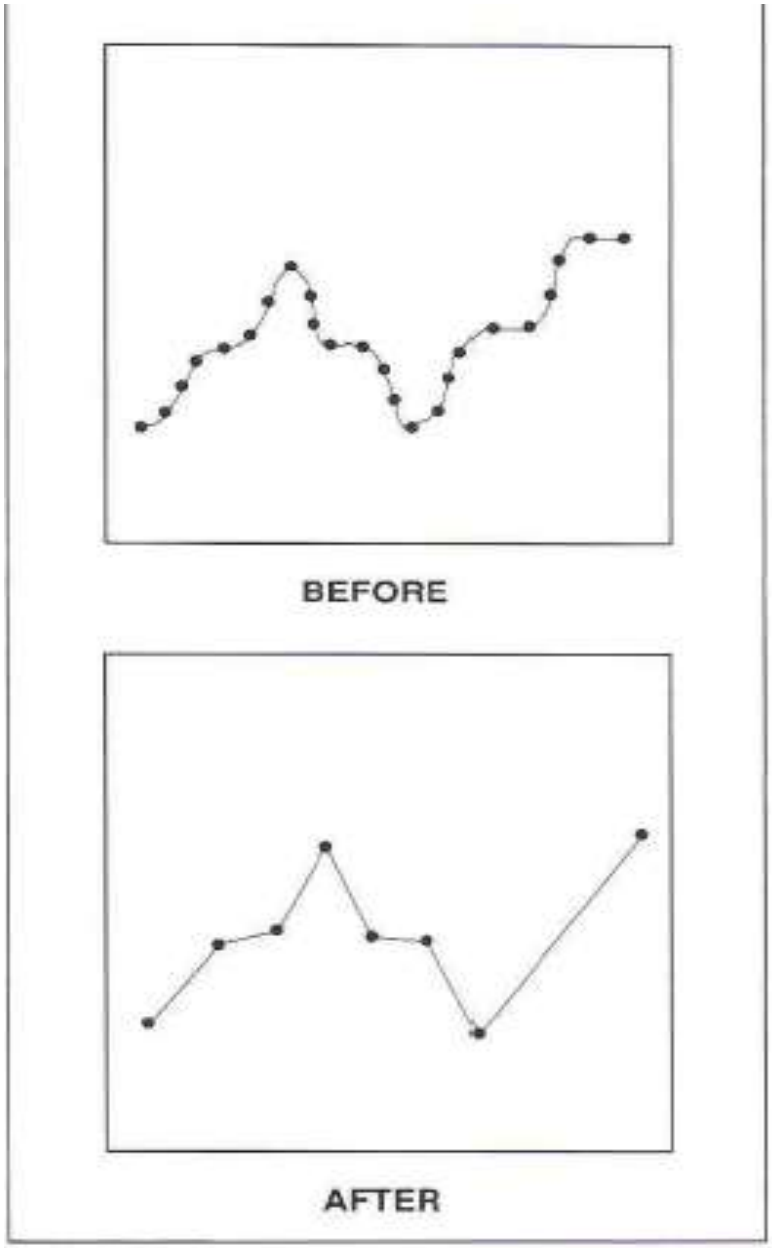


Figure 7.10 Line Coordinate Thinning. Coordinate thinning reduces the number of coordinate pairs used to store a line segment within the GIS.

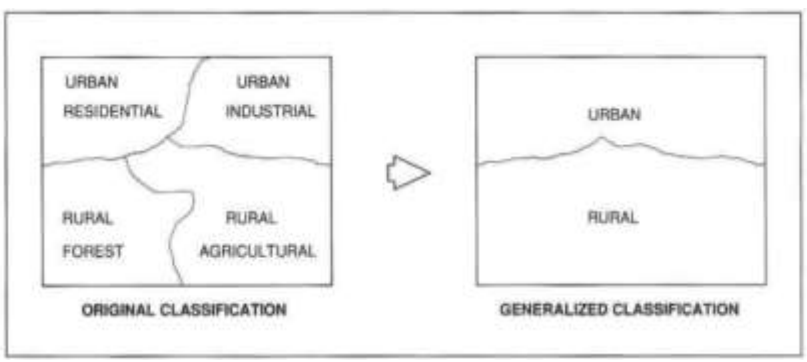
5.1.2 Maintenance and analysis of attribute data.

- Attribute editing functions: allow the attributes to be retrieved, examined and changed.

- Attribute query functions: retrieve records from the database relevant to a query.

5.1.3 Integrated analysis of spatial and attribute data.

- Retrieval operations: involve the search, manipulation and output of the data without modifying the locations of the features.
- Classification and generalization:
    - Classification: the procedure of identifying a set of features as belonging to a group.
    - Generalization: called map dissolve, is the process of making a classification less detailed by combining classes.
- Measurement: Spatial measurements include distance measurement between points, line length, perimeters and areas of polygons.



5.1.4 Output formatting (map annotation)

- Map design principles

1. The names should be logical and close to the feature they describe.
2. The association between the name and the object should be easily recognized.
3. Labels should not overlap.
4. The format and positioning of the name labels should reflect its relative importance.

References

1. Ian, Heywood. An introduction to geographical information systems. Pearson Education India, 2010.
2. Aronoff, Stan. "Geographic information systems: a management perspective." (1989): 58-58.
3. Scholten, Henk J., and Marion JC de Lepper. "An introduction to geographical information systems." The added value of geographical information systems in public and environmental health. Springer, Dordrecht, 1995. 53-70.
4. Burrough, Peter A., Rachael A. McDonnell, and Christopher D. Lloyd. Principles of geographical information systems. Oxford university press, 2015.
5. Chang, Kang-Tsung. Introduction to geographic information systems. Vol. 4. Boston: McGraw-Hill, 2008.

Exercises

Chapter 1: Introduction

1) What are the Elements of GIS?
2) Define geospatial data.
3) Define geometries and attributes as the two components of GIS data.
4) Name *two* examples of vector data analysis.
5) Name *two* examples of raster data analysis.

Chapter 2: GIS Data

1) What are the basic Remote Sensing Techniques?
2) What are the sources for geo-data?
3) What is the Vector Data Model?
4) What is the Raster Data Model?

Chapter 3: Geographic Data Input & Output Methods and Data Quality.

1) What are the types of methods used for data input of GIS data?
2) What are the types of methods used for data output of GIS data?
3) What is a Data quality?

Chapter 4: Data Management

1) What are the Advantages of The Database Approach?
2) What are the Disadvantages of The Database Approach?

3) The information for a geographic feature has four major components, take about them.

Chapter 5: GIS Analysis Functions

1) What is the Classification of GIS Analysis Functions?
2) What is the Edge matching?
3) What is the Classification?
4) What is the Generalization?