DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

LECTURE NOTES ON

OGI352 GEOGRAPHICAL INFORMATION SYSTEM

OGI352 GEOGRAPHICAL INFORMATION SYSTEM

COURSE OBJECTIVES:

To impart the knowledge on basic components, data preparation and implementation of Geographical Information System.

UNIT I FUNDAMENTALS OF GIS

Introduction to GIS - Basic spatial concepts - Coordinate Systems - GIS and Information Systems – Definitions – History of GIS - Components of a GIS – Hardware, Software, Data, People, Methods – Proprietary and open source Software - Types of data – Spatial, Attribute data- types of attributes – scales/ levels of measurements.

UNIT II SPATIAL DATA MODELS

Database Structures – Relational, Object Oriented – Entities – ER diagram - data models -conceptual, logical and physical models - spatial data models – Raster Data Structures – Raster Data Compression - Vector Data Structures - Raster vs Vector Models- TIN and GRID data models.

UNIT III DATA INPUT AND TOPOLOGY

Scanner - Raster Data Input – Raster Data File Formats – Georeferencing – Vector Data Input – Digitizer – Datum Projection and reprojection -Coordinate Transformation – Topology - Adjacency, connectivity and containment – Topological Consistency – Non topological file formats - Attribute Data linking – Linking External Databases – GPS Data Integration

UNIT IV DATA QUALITY AND STANDARDS

Data quality - Basic aspects - completeness, logical consistency, positional accuracy, temporal accuracy, thematic accuracy and lineage – Metadata – GIS Standards –Interoperability - OGC - Spatial Data Infrastructure

UNIT V DATA MANAGEMENT AND OUTPUT

Import/Export – Data Management functions- Raster to Vector and Vector to Raster Conversion - Data Output - Map Compilation – Chart/Graphs – Multimedia – Enterprise Vs. Desktop GISdistributed GIS.

TOTAL:45 PERIODS

COURSE OUTCOMES:

On completion of the course, the student is expected to

CO1 Have basic idea about the fundamentals of GIS.

CO2 Understand the types of data models.

CO3 Get knowledge about data input and topology

CO4 Gain knowledge on data quality and standards

CO5 Understand data management functions and data output

TEXTBOOKS:

1. Kang - Tsung Chang, Introduction to Geographic Information Systems, McGraw Hill Publishing, 2nd Edition, 2011.

2. Ian Heywood, Sarah Cornelius, Steve Carver, Srinivasa Raju, "An Introduction Geographical Information Systems, Pearson Education, 2nd Edition,2007.

REFERENCES:

1. Lo. C. P., Albert K.W. Yeung, Concepts and Techniques of Geographic Information Systems, Prentice-Hall India Publishers, 2006

CHAPTER- 1

FUNDAMENTALS OF GIS

1.1. GIS OVERVIEW

GIS refers to three integrated parts.

a) **Geographic**: Of the real world; the spatial realities, the geography.

b) **Information**: Data and information; their meaning and use.

c) **Systems**: The computer technology and support infrastructure.

GIS therefore refers to a set of three aspects of our modern world, and offers new ways to deal with them. The concept of information is indeed the heart of the rapidly growing field of Geographic Information Systems or GIS. As the world moves into the Information Age, meaningful data and information are becoming the major ‘currency’. With the continuous advances in computer technology, it is easy to concentrate on dazzling systems and software, but the real value of any product is the data and the information such data provides. At the heart of any GIS is information.

1.1.1. Introduction to GIS

(i) A **geographic information system** (**GIS**) is a system designed to capture, store, manipulate, analyze, manage, and present spatial or geographic data. GIS applications are tools that allow users to create interactive queries (user-created searches), analyze spatial information, edit data in maps, and present the results of all these operations. GIS (more commonly GIScience) sometimes refers to geographic information science (GIScience), the science underlying geographic concepts, applications, and systems.

(ii) GIS can refer to a number of different technologies, processes, techniques and methods. It is attached to many operations and has many applications related to engineering, planning, management, transport/logistics, insurance, telecommunications, and business. For that reason, GIS and location intelligence applications can be the foundation for many location-enabled services that rely on analysis and visualization.

(iii) GIS can relate unrelated information by using location as the key index variable. Locations or extents in the Earth space–time may be recorded as dates/times of occurrence, and x, y, and z coordinates representing, longitude, latitude, and elevation, respectively. All Earth-based spatial–temporal location and extent references should be relatable to one another and ultimately to a "real" physical location or extent. This key characteristic of GIS has begun to open new avenues of scientific inquiry.

1.2. BASIC SPATIAL CONCEPTS

Spatial Concept seeks to position spatial concepts as the driving force for spatial thinking and for the selection and use of spatial tools. Eight concepts are the focus of spatial reasoning in the use of geographical information. These concepts are demonstrable at all levels of space and time (from sub-atomic to galactic, passed through future, and microseconds to ions). They can be rendered understandable through simple illustrations to young children but they are also sufficiently engaging at advanced levels for thinking about scientific and social problems.

Location - Understanding formal and informal methods of specifying “where”

Distance - The ability to reason from knowledge of relative position

Network - Understanding the importance of connections

Neighborhood and Region - Drawing inferences from spatial context

Scale - Understanding spatial scale and its significance

Spatial Heterogeneity - The implications of spatial variability

Spatial Dependence - Understanding relationships across space

Objects and Fields - Viewing phenomena as continuous in space-time or as discrete

These concepts have been a foundation for researchers for centuries (see Classics in Spatial Thinking). They have been augmented in recent decades with computational and visualization tools and with vast and easily accessible information resources. These concepts and tools must be as central to general education as reading, writing, and arithmetic. In conjunction with the appropriate spatial tools, they provide a basic scaffold for designing research, solving problems, and structuring education programs.

Spatial concepts invites contributions about other concepts for spatial thinking (e.g., in design fields and in the humanities and arts). Examples may include the link between form and function in architecture, the search for pattern in speech and text, the use of spatial notation in music, the use of spatial metaphor in the sciences and humanities, the importance of place in cultural and social studies, and the spatial elements of aesthetics in the visual arts. If you wish to add concepts to the listing.

1.2.1. Location:

Every type of spatial data has two components: a location and some attribute(s). In this way, location can be seen as a fundamental trait that both defines spatial data and separates it from other types of information. Broadly speaking, there are two types of locations: absolute and relative. Absolute location refers to an exact position on the Earth’s surface defined by some coordinate system. Street addresses and latitude/longitude coordinates are good examples of absolute locations. Relative locations, on the other hand, are defined in reference to other objects. For example, one could define UCSB’s relative location as 10 miles west of downtown Santa Barbara.

The method of determining or measuring location is called georeferencing. There are many different ways of defining a georeference, but they all must meet a few requirements. First, the georeference must be unique so that only one location is described. Second, georeferences must have an accepted meaning that is shared so that most users understand its implication. Finally, georeferences must be viable throughout time so that their meaning is not lost. Metric georeferences are those which define location by measurement and are of particular importance in analyzing spatial data. In order to properly define a metric georeference, consideration must be given to the shape of the Earth, map projections and coordinate systems, and positional accuracy.

Representing location is also an important concept in the use of spatial data. Spot locations are represented as zero-dimension points, lines as one-dimensional poly-lines, areas as two-dimensional polygons, and volumes as three-dimensional polyhedra.

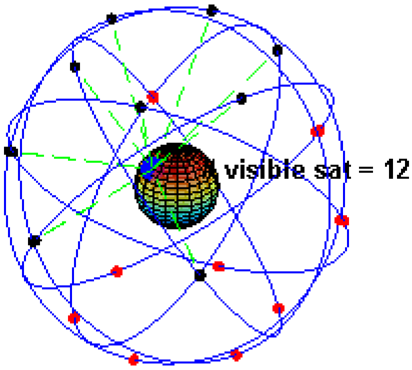


Fig. 1.1. Related Images Location Example (GPS image)

1.2.2. Distance

Distance describes the measurement or separation of two objects or places. At its most basic, measured distance provides a mechanism for describing spatial extent. In spatial reasoning, the classical case is that a positive relationship exists between distance and locational attribute similarity. The relationship, as worded by the Swiss-American cartographer, Waldo Tobler, is, "Everything is related to everything else, but near things are more related to each other." Examining the strength, limitations, and exceptions of the distance-similarity correlation serves as the foundation of much quantitative and qualitative spatial research. Measured distance also provides a necessary dimension for describing position. All real points can be plotted with respect to another by coupling measures of distance with heading and/or time. Common application of distance include spatial analysis and modeling, physics and gravitation, distance decay, buffers, geodesics, route description and optimization, and qualitative comparison of place.

1.2.3. Network

A network is a physical or conceptual system of linkages among entities. Networks offer an infrastructure for representing the anisotropic relationships of various constituents and constituent attributes. Typically, a network connection denotes increased accessibility or relatedness along a link, and linkages may override the default notion that closer features are more accessible and related. For instance, driving from one side of a river to the other often requires a circuitous path many orders of magnitude greater than the direct distance from origin to destination.

Networks commonly are represented by the elements: linkages, nodes, and intersections. Carefully crafted associations among these elements facilitate analyses regarding least-cost path optimization, measures of separation and similarity, and emergent spatial structures. Additionally, attributes can be assigned to various network elements to allow for directionality (e.g., a one-way street), intersection policy (e.g., a no U-turn rule), cost (e.g., a speed limit), and system regulation (e.g., synchronized traffic signals to control flow). Some applications of networks include traffic management, delivery systems, social structures, river hydrology, and communication systems.

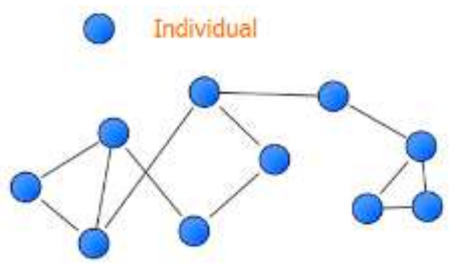


Fig.1.2. Network Example (Social Network Diagram)

1.2.4. Neighborhoods and regions

Neighborhoods and regions define areas surrounding and containing spatial data. They may be formal in nature, such as state and country boundaries, and informal, such as the colloquial use of terms like “downtown.” Regions may also be defined in terms of a particular function. For example, the functional region for a pizza restaurant may be the area within a city to which it delivers.

Utilizing neighborhoods and regions can allow one to make inferences about data from its spatial context. Many fields use neighborhoods and regions in just this way. In remote sensing, neighborhood statistics can be calculated using the values of adjacent pixels. In business, service regions are defined to maximize the number and locations of stores or restaurant franchises. In landscape ecology, metrics have been developed to measure the fragmentation of environmental patches within an ecosystem.

There are problems with making inferences between data at different scales, however. The modifiable area unit problem (MAUP), or ecological fallacy, involves two issues that underscore the use of data aggregated to the neighborhood or region. First, there is a problem of scale. In this way, results of analysis at one scale are not comparable to results at higher or lower levels of aggregation. For example, it does not make sense to use population density calculated at the state level to describe density at the county level. Second, there is a problem of aggregation. Changing region boundaries can have immense effects on the neighborhood statistics. The gerrymandering of voting districts is one example of an aggregation problem.

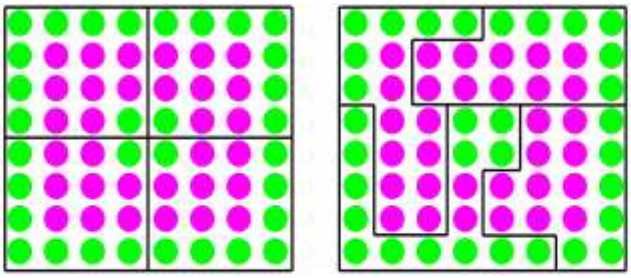


Fig. 1.3. Neighborhood/Region Example (Gerrymandering diagram)

1.2.5. Scale

In spatial reasoning, scale describes the dimensional relationship between a representation and reality. Due to the large variation of all space, scale is used to project reality to more useful and meaningful sizes. For large expanses, scale is reduced (e.g., fitting the entirety of earth's surface on to a paper map), and for miniscule distances, scale is increased (e.g., enlarging and schematizing chemical reactions). Often, scale is denoted as a fraction where a unit of measure in reality is compared to the same unit on the projection. For instance, a paper map of a landscape showing a scale of 1:10,000 would mean that a drawn unit on the map represents 10,000 of the same unit in reality.

Offering spatial data at scales different from their original data collection granularity may imply changes of meaning; thus, data product limitations should be carefully considered by both creators and consumers of spatial information. As a fundamental spatial concept, applications of scale permeate many human activities including data collection, cartography, art, architecture, engineering, and nanotechnology.

1.2.6. Spatial heterogeneity

Spatial heterogeneity refers to the degree of variation in some attribute across places and region. For example, a satellite image of the Pacific Ocean would show little variation and thus would have a low level of spatial heterogeneity; whereas, an image of the patchwork of agricultural fields in the Midwestern U.S. may be considered highly spatially heterogeneous. In the same way that biodiversity defines species variation in biology, spatial heterogeneity defines variation of an attribute in spatial studies.

There are many research implications caused by the fundamental heterogeneity of spatial data. In most cases, spatial data can be said to have a non-constant mean and variance throughout a study area. That is, local statistical parameters change with location and are thus not uniformly distributed. This characteristic of spatial data is termed non-stationarity and directly impacts research in areas such as sampling design. For example, it is very difficult to obtain a representative sample of a region because of the heterogeneous nature of many spaces.

Another characteristic of spatial data that impacts research is the general rule that spatial data tend to become more heterogeneous as the study area gets larger. This means the variance observed in a small region is less than that in a larger region. That is, expected variation around the mean in small regions underestimates and, thus, is not applicable in subsequently larger regions. This uncontrolled variance in spatial data has impacts in the study design of projects involving large areas and times, such as global warming.

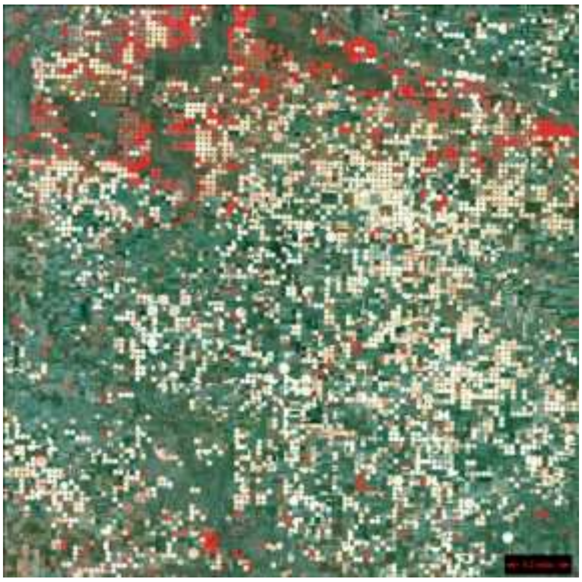


Fig.1.4. Spatial Heterogeneity Example (Landscape Patch image)

1.2.7. Spatial dependence:

Spatial dependence is the manifestation of Tobler's first law of geography, which states, "Everything is related to everything else, but near things are more related to each other." This seemingly simple principle applies to many types of spatial attributes. For example, one would expect the temperature of Santa Barbara to be more similar to the temperature in Los Angeles than to the temperature in Seattle. This relatedness between data based on the distance is also termed spatial autocorrelation. If spatial data were truly random, there would be zero spatial autocorrelation. Applying this idea to the previous example would mean that no inference of similarity could be made between the temperatures in Santa Barbara and Los Angeles. Fortunately, most attributes in the world are not distributed in this manner and thus display a certain degree of spatial dependence.

Researchers can quantify spatial correlation by using indices such as Moran's I and Geary's C. Researchers can also model spatial dependence by using the methods developed in the field of geostatistics. Techniques such as kriging allow researchers to quantify the changes in attribute variance versus distance. This model of spatial dependence is called a semi-variogram, or just variogram. Variograms offer a mechanism to predict attribute values in locations where data are not present. These powerful interpolation tools have many real-world applications in fields such as mining and petroleum discovery, epidemiology, atmospheric science, oceanography, and soil science.

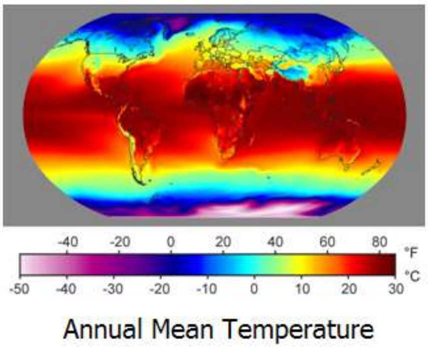


Fig. 1.5. Spatial Dependence Example (Temperature Map)

1.2.8. Objects and fields:

Objects and fields describe two fundamental, dichotomous, conceptualizations of space. Objects are collections of discrete, bounded entities, usually composed of geometric primitives, such as points, lines, curves, and polygons. Conceptually, object-based reality is considered empty space populated by distinct entities. Attributes are associated with bounded features to define and describe the objects. Conversely, field-based representation fills continuous space with attribute measures at all locations. A field offers a conceptual model of spatial variation, and attributes themselves, as opposed to workings of distinct boundaries, define the field.

The object-field dichotomy serves as the underpinning for all methods of spatial representation and analysis, and each perspective offers a characteristic set of abilities and limitations. For instance, objects offer logical concepts like "inside" and "outside," and comparisons such as intersections and buffers. Fields are particularly suited for spatial phenomena considered less bounded, like air, water, temperature, and elevation. Not constrained by distinct edges, fields lend themselves to analyses related to interpolation and global attribute comparison. Analysis and representation of fields often yield areas of peaks, valleys, aspects, and slopes.

Fig. 1.6. Object Example (Cadastral Map)

The representational and analytical potential of both perspectives is mirrored in computer software for modeling reality. Modeling software for analyzing property boundaries, road networks, and individuals, tend to use object-based representation methods. Software for analyzing less-bounded phenomena, like temperature variation, ocean circulation, and hydrologic modeling, typically use field-based models for database storage and analysis.

Ongoing research on the object-field dichotomy has focused on repercussions of accuracy, uncertainty, and usage, the philosophy of spatial reality within the context of such representations, and defining and utilizing object-fields hybrids. Common domains using objects and fields are Geographic Information Systems, spatial ontology, spatial analysis, and database design.

1.3. GEOGRAPHIC COORDINATE SYSTEMS

A geographic coordinate system is a reference system for identifying locations on the curved surface of the earth. Locations on the earth's surface are measured in angular units from the center of the earth relative to two planes: the plane defined by the equator and the plane defined by the prime meridian (which crosses Greenwich England). A location is therefore defined by two values: a latitudinal value and a longitudinal value.

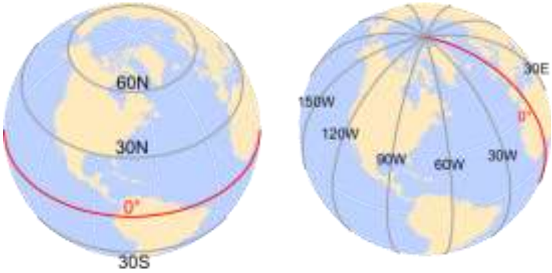


Figure 1.7: Examples of latitudinal lines are shown on the left and examples of longitudinal lines are shown on the right. The 0° degree reference lines for each are shown in red (equator for latitudinal measurements and prime meridian for longitudinal measurements).

A latitude measures the angle from the equatorial plane to the location on the earth's surface. A longitude measures the angle between the prime meridian plane and the north-south plane that intersects the location of interest. For example Colby College is located at around 45.56° North and 69.66° West. In a GIS system, the North-South and East-West directions are encoded as signs. North and East are assigned a positive (+) sign and South and West are assigned a negative (-) sign. Colby College's location is therefore encoded as +45.56° and -69.66°.

Figure 1.8. A slice of earth showing the latitude and longitude measurements.

A GCS is defined by an ellipsoid, geoid and datum. These elements are presented next.

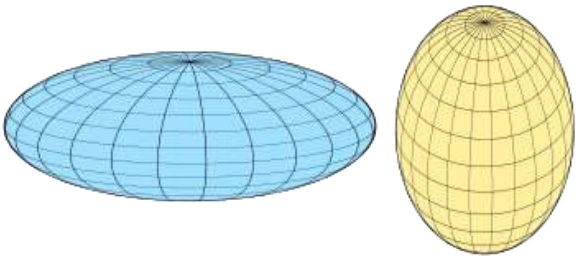
1.3.1. Sphere and Ellipsoid

Assuming that the earth is a perfect sphere greatly simplifies mathematical calculations and works well for small-scale maps (maps that show a large area of the earth). However, when working at larger scales, an ellipsoid representation of earth may be desired if accurate measurements are needed. An ellipsoid is defined by two radii: the semi-major axis (the equatorial radius) and the semi-minor axis (the polar radius).

The reason the earth has a slightly ellipsoidal shape has to do with its rotation which induces a centripetal force along the equator. This results in an equatorial axis that is roughly 21 km longer than the polar axis.

Our estimate of these radii is quite precise thanks to satellite and computational capabilities. The semi-major axis is 6,378,137 meters and the semi-minor axis is 6,356,752 meters.

Differences in distance measurements along the surfaces of an ellipsoid vs. a sphere are small but measurable (the difference can be as high as 20 km) as illustrated in the following lattice plots.



1.9. Fig. Spheroids with vertical rotational axes

1.3.2. Geoid

Representing the earth's true shape, the geoid, as a mathematical model is crucial for a GIS environment. However, the earth's shape is not a perfectly smooth surface. It has undulations resulting from changes in gravitational pull across its surface. These undulations may not be visible with the naked eye, but they are measurable and can influence locational measurements.

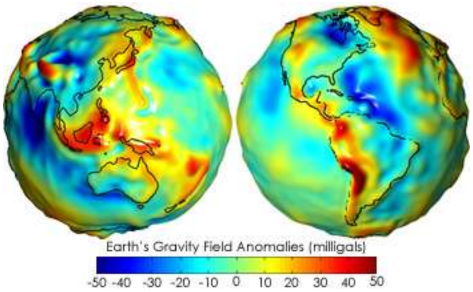


Figure 1.10. Earth's geoid with gravitational field shown in rainbow colors. The ondulations depicted in the graphics are exaggerated for visual effects. (source: NASA)

Note that we are not including mountains and ocean bottoms in our discussion, instead we are focusing solely on the earth's gravitational potential which can be best visualized by imagining the earth's surface completely immersed in water and measuring the sea surface level over the entire earth surface.

The earth's gravitational field is dynamic and is tied to the flow of the earth's hot and fluid core. Hence its geoid is constantly changing, albeit at a large temporal scale.The measurement and representation of the earth's shape is at the heart of geodesy–a branch of applied mathematics.

1.3.3. Datum

So how are we to reconcile our need to work with a (simple) mathematical model of the earth's shape with the on dulating nature of the earth's surface (i.e. its geoid). The solution is to align the geoid with the ellipsoid (or sphere) representation of the earth and to map the earth's surface features onto this ellipsoid/sphere. The alignment can be local where the ellipsoid surface is closely fit to the geoid at a particular location on the earth's surface (such as the state of Kansas) or geocentric where the ellipsoid is aligned with the center of the earth. How one chooses to align the ellipsoid to the geoid defines a datum.

Local Datum

There are many local datums to choose from, some are old while others are more recently defined. The choice of datum is largely driven by the location of interest. For example, when working in the US, a popular local datum to choose from is the North American Datum of 1927 (or NAD27 for short). NAD27 works well for the US but it's not well suited for other parts of the world. For example, a far better local datum for Europe is the European Datum of 1950 (ED50 for short). Examples of common local datums are shown in the following table:

Local datum	Acronym	Best for	Comment
North American Datum of 1927	NAD27	Continental US	This is an old datum but still prevalent because of the wide use of older maps.
European Datum of 1950	ED50	Western Europe	Developed after World War II and still quite popular today. Not used in the UK.
World Geodetic System 1972	WGS72	Global	Developed by the Department of Defense.

Geocentric Datum

Many of the modern datums use a geocentric alignment. These include the popular World Geodetic Survey for 1984 (WGS84) and the North American Datums of 1983 (NAD83). Most of the popular geocentric datums use the WGS84 ellipsoid or the GRS80 ellipsoid. These ellipsoids' semi-major and semi-minor axes are nearly identical: 6,378,137 meters and 6,356,752 meters respectively. Examples of popular geocentric datums are shown in the following table:

Geocentric datum	Acronym	Best for	Comment
North American Datum of 1983	NAD83	Continental US	This is one of the most popular modern datums for the contiguous US.

European Terrestrial Reference System 1989	ETRS89	Western Europe	This is the most popular modern datum for much of Europe.
World Geodetic System 1984	WGS84	Global	Developed by the Department of Defense.

1.4. GIS AS AN INFORMATION SYSTEM

As Definition of GIS indicates GIS as a specialized information system stresses "spatially distributed features (points, lines, areas), activities (physical and human-invoked), and events (time).

GIS as an approach to Geographic Information Science

1) research on GIS (algorithms, analytical methods, visualization tools, user interfaces, human-computer-human interaction)
2) research with GIS: GIS as a tool used by many substantive disciplines in their own ways (anthropology, archeology, forestry, geology, engineering, business and management sciences)

1.5. DEFINING GIS

A GIS is a system of hardware, software, data, people, organizations, and institutional arrangements for collecting, storing, analyzing and disseminating information about areas of the earth. It is also defined as an information system designed to work with data referenced by spatial / geographical coordinates. In other words, GIS is both a database system with specific capabilities for spatially referenced data as well as a set of operations for working with the data. A Geographic Information System is a computer based system which is used to digitally reproduce and analyze the feature present on earth surface and the events that take place on it. In the light of the fact that almost 70% of the data has geographical reference as its denominator, it becomes imperative to underline the importance of a system which can represent the given data geographically.

1.6. HISTORY OF GIS

The idea of portraying different layers of data on a series of base maps, and relating things geographically, has been around much older than computers invention. Thousands years ago, the early man used to draw pictures of the animals they hunted on the walls of caves. These animal drawings are track lines and tallies thought to depict migration routes. While simplistic in comparison to modern technologies, these early records mimic the two-element structure of modern geographic information systems, an image associated with attribute information.

Possibly the earliest use of the geographic method, in 1854 John Snow depicted a cholera outbreak in London using points to represent the locations of some individual cases. His study of the distribution of cholera led to the source of the disease, a contaminated water pump within the heart of the cholera outbreak. While the basic elements of topology and theme existed previously in cartography, the John Snow map was unique, using cartographic methods, not only to depict but also to analyze, clusters of geographically dependent phenomena for the first time.

1.12 GIS

The early 20th century saw the development of "photo lithography" where maps were separated into layers. Computer hardware development spurred by nuclear weapon research led to general-purpose computer "mapping" applications by the early 1960s. In the year 1962, the world's first true operational GIS was developed by the federal Department of Forestry and Rural Development in Ottawa, Canada by Dr. Roger Tomlinson. It was called the "Canada Geographic Information System" (CGIS) and was used to store, analyze, and manipulate data collected for the Canada Land Inventory (CLI). It is an initiative to determine the land capability for rural Canada by mapping information about soils, agriculture, recreation, wildlife, forestry, and land use at a scale of 1:50,000.

CGIS was the world's first "system" and was an improvement over "mapping" applications as it provided capabilities for overlay, measurement, and digitizing or scanning. It supported a national coordinate system that spanned the continent, coded lines as "arcs" having a true embedded topology, and it stored the attribute and location specific information in a separate files. Dr. Tomlinson is known as the "father of GIS," for his use of overlays in promoting the spatial analysis of convergent geographic data.

In 1964, Howard T Fisher formed the Laboratory for Computer Graphics and Spatial Analysis at the Harvard Graduate School of Design, where a number of important theoretical concepts in spatial data handling were developed. This lab had major influence on the development of GIS until early 1980s. Many pioneers of newer GIS "grew up" at the Harvard lab and had distributed seminal software code and systems, such as 'SYMAP', 'GRID', and 'ODYSSEY'.

By the early 1980s, M&S Computing (later Intergraph), Environmental Systems Research Institute (ESRI) and CARIS emerged as commercial vendors of GIS software, successfully incorporating many of the CGIS features, combining the first generation approach to separation of spatial and attribute information with a second generation approach to organizing attribute data into database structures. More functions for user interaction were developed mainly in a graphical way by a user friendly interface (Graphical User Interface), which gave to the user the ability to sort, select, extract, reclassify, reproject and display data on the basis of complex geographical, topological and statistical criteria. During the same time, the development of a public domain GIS begun by the U.S. Army Corp of Engineering Research Laboratory (USA-CERL) in Champaign, Illinois, a branch of the U.S. Army Corps of Engineers to meet the need of the United States military for software for land management and environmental planning.

In the years 1980s and 1990s industry growth were spurred on by the growing use of GIS on Unix workstations and the personal computers. By the end of the 20th century, the rapid growth in various systems had been consolidated and standardized on relatively few platforms and users were beginning to export the concept of viewing GIS4 data over the Internet, requiring uniform data format and transfer standards. More recently, there is a growing number of free, open source GIS packages, which run on a range of operating systems and can be customized to perform specific tasks. As computing power increased and hardware prices slashed down, the GIS became a viable technology for state development planning. It has become a real Management Information System (MIS), and thus able to support decision making processes.

1.7. COMPONENTS OF A GIS

A working GIS integrates five key components: -

i) Hardware ii) Software iii) Data iv) People v) Methods

GIS enables the user to input, manage, manipulate, analyze, and display geographically referenced data using a computerized system. To perform various operations with GIS, the components of GIS such as software, hardware, data, people and methods are essential.

1.7.1. Software

GIS software provides the functions and tools needed to store, analyze, and display geographic information. Key software components are (a) a database management system (DBMS) (b) tools for the input and manipulation of geographic information (c) tools that support geographic query, analysis, and visualization (d) a graphical user interface (GUI) for easy access to tools. GIS software are either commercial software or software developed on Open Source domain, which are available for free. However, the commercial software is copyright protected, can be expensive and is available in terms number of licensees.

Currently available commercial GIS software includes Arc/Info, Intergraph, MapInfo, Gram++ etc. Out of these Arc/Info is the most popular software package. And, the open source software are AMS/MARS etc.

1.7.2. Hardware

Hardware is the computer on which a GIS operates. Today, GIS runs on a wide range of hardware types, from centralized computer servers to desktop computers used in stand-alone or networked configurations. Minimum configuration required to Arc/Info Desktop 9.0 GIS application is as follows:

Product: ArcInfo Desktop 9.0

Platform: PC-Intel

Operating System: Windows XP Professional Edition, Home Edition

Service Packs/Patches: SP 1

SP2 (refer to Limitations)

Shipping/Release Date: May 10, 2004

Hardware Requirements

CPU Speed: 800 MHz minimum, 1.0 GHz recommended or higher

Processor: Pentium or higher

Memory/RAM: 256 MB minumum, 512 MB recommended or higher

Display Properties: Greater than 256 color depth

Swap Space: 300 MB minimum

Disk Space: Typical 605 MB NTFS, Complete 695 MB FAT32 + 50 MB forinstallation

Browser: Internet Explorer 6.0 Requirement:

(Some features of ArcInfo Desktop 9.0 require a minimum installation of Microsoft Internet Explorer Version 6.0.)

1.7.3. Data

The most important component of a GIS is the data. Geographic data or Spatial data and related tabular data can be collected in-house or bought from a commercial data provider. Spatial data can be in the form of a map/remotely-sensed data such as satellite imagery and aerial photography. These data forms must be properly geo-referenced (latitude/longitude). Tabular data can be in the form attribute data that is in some way related to spatial data. Most GIS software comes with inbuilt Database Management Systems (DBMS) to create and maintain a database to help organize and manage data.

1.7.4. Users

GIS technology is of limited value without the users who manage the system and to develop plans for applying it. GIS users range from technical specialists who design and maintain the system to those who use it to help them do their everyday work.

These users are largely interested in the results of the analyses and may have no interest or knowledge of the methods of analysis. The user-friendly interface of the GIS software allows the nontechnical users to have easy access to GIS analytical capabilities without needing to know detailed software commands. A simple User Interface (UI) can consist of menus and pull-down graphic windows so that the user can perform required analysis with a few key presses without needing to learn specific commands in detail.

1.7.5. People

GIS technology has limited value without the people who manage and develop plans for applying it to real world problems. GIS user range from technical specialists who design and maintain the system to those who use it to help them perform their everyday work. The identification of GIS specialist's vs. end users is often critical to the proper implementation of GIS technology. This is what called 'brain ware' which is equally important as the Hardware and software. Brain ware refers to the purpose and objectives, and provides the reason and justification, for using GIS.

This component of GIS includes all those individuals (such as the programmer, database manager, GIS researcher etc.) who are making the GIS work, and also the individuals who are at the user end using the GIS services, applications and tools.

1.7.6. 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.

1.8. PROPRIETARY AND OPEN SOURCE GIS SOFTWARE

1.8.1. AGISMap

AGIS for Windows is a mapping and simple GIS package specifically designed to be easy to use, and distributed as shareware via the world wide web.

Platforms: *Windows*

1.8.2. Autodesk

Autodesk has a series of software applications designed to meet GIS needs in a variety of areas that interfaces with their CAD software.

Platforms: *Windows*

1.8.3. Bentley Systems, Inc.

Bentley provides software for the "Design, construction, and operation of the world's infrastructure". The company's software serves the geospatial, building, plant, and civil vertical markets in the areas of mapping, architecture, engineering, construction (AEC) and operations. Bentley offers a wide range of products for surveying, GPS, photogrammetry, remote sensing, imaging, conversion, mapping, cartography, and other geospatial applications built on MicroStation Products: Bentley Map – Desktop GIS, Bentley Cadastre – Desktop land management GIS, Bentley Descartes – Desktop image editing, analysis, and processing, Bentley Geo Web Publisher – GIS web publishing and viewing, Bentley PowerMap Field – Field-enabled GIS and redlining.

Platforms: *Windows*

1.8.4. Cartographica

Commercial software package for Mac OS featuring: Support for a huge number of import formats, including popular Raster formats, manual editing and georeferencing, automatic geocoding, integration with online mapping, output to large-format printers. Thirty-day demo available for download.

Platforms: *Macintosh*

1.8.5. DeLorme

DeLorme is the producers of XMap, a GIS application "with 80% of the functionality found in a traditional GIS at 15% of the cost". Performs such functions as geocoding, image rectification, 3D visualization and coordinate transformation.

Platforms: *Windows*

1.8.6. Esri

Environmental Systems Research Institue has been creating GIS software for over 30 years. Recognized as the leader in GIS software, it's been estimated that about seventy percent of

1.16 GIS

GIS users use Esri products. Esri overhauled their software packages into an interoperable model called ArcGIS (the desktop GIS is referred to as ArcMap). In addition, Esri has developed plug-ins called extensions which add to the functionality of ArcGIS. Demo and light versions of Esri software are available for downloading. You can also find free data to use with Esri products.

Platforms: *Windows*

Further Resources: *ArcGIS, ArcView 3.x (no longer in production)*

1.8.7. Intergraph

Intergraph makes several GIS applications. Most of the GIS packages are designed with an Open GIS in mind and therefore can work with a variety of other GIS software formats. Intergraph has developed products that help merge GIS with information technology (IT) and business process improvement tools. Intergraph offers the Geo-Media family of solutions and Modular GIS Environment MGE Suite of mapping and GIS applications.

Using an open architecture, the Geo-Media product suite integrates geospatial information throughout the enterprise and provides the tools needed to develop business-to-business and custom client applications using industry standard development tools. Geo-Media offers uninhibited access to all geospatial data formats without the need for data translations. Currently in Version 4.0 the Geo-Media family is made up of Geo-Media, Geo-Media Professional, Geo Media Web Map, and Geo Media Web Enterprise.

- GeoMedia is the universal information integrator, serving as a visualization and analysis tool and as an open platform for custom GIS solution development.
- GeoMedia Professional is a product specifically designed to collect and manage spatial data using standard databases.
- GeoMedia WebMap is a Web-based map visualization tool with real-time links to one or more GIS data warehouses.
- GeoMedia WebEnterprise creates dynamic, custom web-mapping applications that can analyze and manipulate geographic data.
- In addition to these products, Intergraph offers MFworks for GeoMedia which provides users of grid-based software the power of visualization, mapping, and analysis. Intergraph also offers SMMS for GeoMedia which is a desktop tool for geographic metadata creation and geographic data management.

The Modular GIS Environment (MGE) product suite provides production-ready capabilities for automating, managing, analyzing, and presenting GIS data, and is completely interoperable with GeoMedia.

1.8.8. Manifold

Manifold System provides comprehensive, professional grade GIS software for $245 that includes a very wide array of features. Manifold imports data from over 80 different GIS formats, including all formats used by Federal government sites for free Internet downloads, and Manifold allows seamless, simultaneous work with vector drawings, raster images, terrain elevations and raster data sets either as 2D displays or 3D terrain visualizations. Manifold includes exceptional DBMS capabilities, full development facilities and includes a built-in Internet Map Server for fast and easy publication of GIS projects to the web without programming. Options include US Street address geocoding and the Enterprise Edition, for centralized geospatial data storage on enterprise servers that can be used by many GIS operators at once. Manifold was the first GIS to attain "designed for XP" status with Microsoft and the Manifold Internet Map Server works perfectly within ASP.NET servers. For info, see http://www.manifold.net/professional.

Platforms: *Windows*

1.8.9. Ortelius

Ortelius is a "map illustration" software package that adds one more selection to the very limited mapping software options out there for Macintosh users. A free trial download is available.

Platforms: *Macintosh*

1.8.10. MapInfo – Pitney Bowes Business Insight (PBBI)

PBBI's flagship software is MapInfo, a suite of GIS software. MapInfo Professional is their leading GIS product containing the most advanced analytical tools. MapInfo also offers plug-ins called add-ons to enhance the functionality of MapInfo Professional. For the development side, MapInfo offers Map-X. Through an Active X component, developers can embed mapping applications into other applications such as Excel. Although it can be used for a variety of analysis, the makers of MapInfo market the software more towards the business sector. Demo versions are available for downloading for some of MapInfo's products.

Platforms: *Windows OS*

Further Resources: *MapInfo, MapBasic, MapInfo Tutorials*

1.8.11 Maptitude

The Maptitude Mapping Software is a full-featured mapping package for Windows. Designed for ease-of-use, data visualization and geographic analysis, Maptitude comes with comprehensive nationwide and worldwide maps, including complete US street maps, and Census tract and ZIP Code boundaries and demographics. Caliper also produces TransCAD for transportation and logistics. TransCAD is used for solving key analytical problems in transportation planning, management, and operations. TransCAD is used extensively for transportation database development and maintenance, demand forecasting, operations management, and vehicle routing and scheduling.

Platforms: *Windows OS*

1.8.12. MyWorld

My World GIS is a full-featured GIS designed for educational use. My World provides a carefully selected subset of the features of a professional GIS environment. These features include multiple geographic projections, table and map views of data, distance-measurement tools, buffering and query operations, customizable map display. They have been selected to provide the greatest value to students without overwhelming them with complexity The features are accessed through a supportive interface designed with the needs of students and teachers in mind. My

World can import data from the industry-standard shapefile format, as well as from tab and comma-delimited text files.

Platforms: *Windows, Macintosh, Linux, Solaris*

1.8.13 Supergeo Technologies

SuperGIS Desktop is a full-featured GIS platform for Windows OS. It allows users to edit, visualize, manage and analyze geospatial data both in vector and raster, including OGC formats and various geodatabases, such as MSSQL, Oracle spatial, and PostGIS. Capabilities can be powered up via extensions such as Network Analyst, 3D Analyst, Spatial Analyst, Biodiversity Analyst, etc., which enables users to conduct complicated analyses and make smarter decisions. By combing Mobile GIS and Server GIS also from Supergeo, you can have a total geospatial solution from data collection in field to data publishing online with reasonable price. The free trial is available on Supergeo's website. Various product resources and friendly technical support are also provided by Supergeo team.

Platforms: *Windows OS*

1.8.14. TatukGIS Editor

Professional, general-purpose desktop GIS mapping and data editing application with built-in scripting environment for customization and feature extensions. Natively supports most GIS/CAD raster/vector/SQL layer data formats, including enterprise spatial geodatabases (such as Oracle Spatial, PostGIS …) State-of-the-art coordinate system support with nearly 5,000 pre-defined coordinate systems and on-the-fly raster/vector layer reprojection, 3D mapping, raster & vector layer rectification, and compatibility with leading database products. Data grid with advanced query and selection tools.

Platforms: *Windows OS*

1.8.15. Terrain Tools

Terrain Tools, produced by Softree, is a software package for surveying and mapping. It is ideal for the forester, geologist, surveyor or resource scientist who is not a GIS specialist, but who needs to quickly produce working maps and site plans.

Platforms: *Windows*

1.8.16. TerrSet Geospatial Monitoring and Modeling System

Developed by Clark Labs, TerrSet is an integrated geospatial software for monitoring and modeling the earth system. The TerrSet software incorporates the IDRISI GISt Analysis and IDRISI Image Processing tools along with a constellation of modeling environments to analyze land change, image time series, ecosystem services, habitat and biodiversity, climate impacts, and REDD. Learn more at www.clarklabs.org.

Platforms: *Windows*

1.8.17. TNT Products

Created by Microimages, The TNT Products is a suite of GIS applications for fully integrated GIS, image processing, CAD, TIN, desktop cartography, and geospatial database management.

Platforms: *Windows, UNIX, Macintosh*

1.9. OPEN SOURCE GIS SOFTWARE

(i) From the UNH Cooperative Extension course "***GIS on Pennies a Day***," this is a list of freely available, open source GIS software. You can download and use the software.

(ii) QGIS and DIVA-GIS are more widely used than the others. Both have MAC versions available for downloading.

1.9.1. Quantum GIS (QGIS)

Quantum GIS (QGIS) is a user friendly Open Source Geographic Information System (GIS) licensed under the GNU General Public License. QGIS is an official project of the Open Source Geospatial Foundation (OSGeo). It runs on Linux, Unix, Mac OSX, and Windows and supports numerous vector, raster, and database formats and functionalities.

1.9.2. DIVA-GIS

DIVA-GIS are a free computer program for mapping and geographic data analysis (a geographic information system (GIS).

1.9.3. Free QGIS Courses from Geo-Academy

Geo-Academy is an independent consortium promoting the use of open source geospatial software. These free courses are based on years of development of their fee-based coursework.

These are other open source GIS software available via the web.

1.9.4. GeoDa

GeoDa is a free GIS software program primarily used to introduce new users into spatial data analysis. It’s main **functionality is data exploration in statistics**.

One of the nicest things about it is how it comes with sample data for you to give a test-drive. From simple box-plots all the way to regression statistics, GeoDa has **complete arsenal of statistics** to do nearly anything spatially.

It’s user base is strong. For example, Harvard, MIT and Cornell universities have embraced this free GIS software to serve as a gentle introduction to spatial analysis for non-GIS users. From economic development to health and real estate, it’s been used as an exciting analytical in labs as well.

From the Center for Spatial Data Science at the University of Chicago, software for geospatial analysis, geovisualization and other techniques.

1.9.5. gvSIG

In 2004, the gvSIG project emerged as a free, open source GIS software option in Spain. We illustrate in this gvSIG guide and review why we like it SO much:

gvSIG really outperforms QGIS 2 for 3D. It really is the best 3D visualization available in open source GIS.

The NavTable is agile in that it allows you to see records one-by-one vertically.

The CAD tools are impressive on gvSIG. Thanks to the OpenCAD Tools, you can trace geometries, edit vertices, snap and split lines and polygons. If you need GIS on your mobile phone, gvSIG Mobile is perfect for field work because of its interface and GPS tools.

1.9.6. MapWindow GIS

An open source GIS software that is being used in History. For more information on its use in History, a manual and exercises.

1.9.7. SAGA (System for Automated Geoscientific Analyses)

SAGA is a free, open source GIS software used for raster-based analyses and the Earth Sciences.

1.9.8. uDig

uDIG is an acronym to help get a better understanding what this Free GIS software is all about.

- u stands for user-friendly interface
- D stands for desktop (Windows, Mac or Linux). You can run uDIG on a Mac.
- I stand for internet oriented consuming standard (WMS, WFS or WPS)
- G stands for GIS-ready for complex analytical capabilities.

When you start digging into uDig, it's a nice open source GIS software option for basic mapping. uDig's Mapnik lets you import basemaps with the same tune as ArcGIS

Specifically, its easy-to-use, the catalog, symbology and Mac OS functionality are some of the strong points. But it has limited tools and the bugs bog it down to really utilize it as a truly complete free GIS software package.

1.9.9 Environmental Benefits Mapping & Analysis Program (BenMAP)

BenMAP-CE is a open-source software program that calculates the number and economic value of air pollution-related deaths and illnesses. It incorporates a database that includes many of the concentration-response relationships, population files, and health and economic data needed to quantify these impacts. This is the Community Edition.

1.10. DATA TYPES

The basic data type in a GIS reflects traditional data found on a map. Accordingly, GIS technology utilizes two basic types of data. These are:

Spatial data: describes the absolute and relative location of geographic features.

Attribute data: describes characteristics of the spatial features. These characteristics can be quantitative and/or qualitative in nature. Attribute data is often referred to as tabular data.

Attribute Data:

The attributes refer to the properties of spatial entities. They are often referred to as non-spatial data since they do not in themselves represent location information. This type of data describes characteristics of the spatial features. These characteristics can be quantitative and/or qualitative in nature. Attribute data is often referred to as tabular data.

Spatial Data:

Geographic position refers to the fact that each feature has a location that must be specified in a unique way. To specify the position in an absolute way a coordinate system is used. For small areas, the simplest coordinate system is the regular square grid. For larger areas, certain approved cartographic projections are commonly used. Internationally there are many different coordinate systems in use. This locational information is provided in maps by using Points, Lines and Polygons. These geometric descriptions are the basic data elements of a map. Thus spatial data describes the absolute and relative location of geographic features.

The coordinate location of a forest would be spatial data, while the characteristics of that forest, e.g. cover group, dominant species, crown closure, height, etc., would be attribute data. Other data types, in particular image and multimedia data, have become more prevalent with changing technology. Depending on the specific content of the data, image data may be considered either spatial, e.g. photographs, animation, movies, etc., or attribute, e.g. sound, descriptions, narration's, etc.

1.10.1. Spatial Data

- Describes the absolute and relative location of geographic features.
- Represents spatial data, which has a physical dimension on earth.
- Spatial data consist of digital representations of discrete (spatial) objects. The features are shown on a map, e.g. lakes, buildings and contours can be thought of as discrete objects.
- Thus the contents of a map can be captured in a database by turning map features into database objects (entities).
- Components of spatial data.

Location: The spatial mode of information is generally called location.

Spatial relationship: The connections between spatial objects are described as spatial relationships (e.g. A contains B; A is adjacent to B, A is North of B, etc.).

Attributes: Attributes capture the thematic mode by defining different characteristics of objects. Spatial features in the real world are reduced in the form of point, line, area and surface. GIS will store the data either in tabular form, geographical map, digital map or remotely sensed map.

1.10.2. Non-Spatial Data

The non-spatial data or the attribute data, on the other hand, describes the characteristics of the spatial features. These characteristics can be quantitative or qualitative, also called attribute data.

- Describes the characteristics of the spatial features.
- It holds the characteristics of the spatial features and the descriptive information about the geographic features.
- Represented using colors, textures and symbols.
- Eg: Coordinate location of a sanctuary would be spatial data, while the characteristic's like the cover group, dominant species, nature of vegetation would be attribute data.
- These attributes are given in an organized from in a single table or multiple tables.

1.10.3. Spatial Data Models

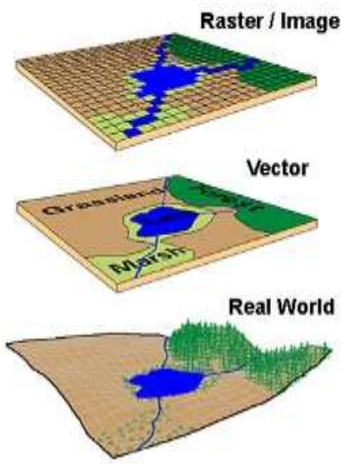
Traditionally spatial data has been stored and presented in the form of a map. Three basic types of spatial data models have evolved for storing geographic data digitally. These are referred to as:

(i) Vector data

(ii) Raster data

(iii) Image data

The following diagram reflects the two primary spatial data encoding techniques. These are vector and raster. Image data utilizes techniques very similar to raster data, however typically lacks the internal formats required for analysis and modeling of the data. Images reflect pictures or photographs of the landscape.



1.10.4. Attribute Data Models

A separate data model is used to store and maintain attribute data for GIS software. These data models may exist internally within the GIS software, or may be reflected in external commercial Database Management Software (DBMS). A variety of different data models exist for the storage and management of attribute data. The most common are:

- Tabular
- Hierarchical
- Network
- Relational
- Object oriented

Tabular Model

The simple tabular model stores attribute data as sequential data files with fixed formats (or comma delimited for ASCII data), for the location of attribute values in a predefined record structure. This type of data model is outdated in the GIS arena. It lacks any method of checking data integrity, as well as being inefficient with respect to data storage, e.g. limited indexing capability for attributes or records, etc.

Hierarchical Model

The hierarchical database organizes data in a *tree* structure. Data is structured downward in a *hierarchy* of tables. Any level in the hierarchy can have unlimited *children*, but any *child* can have only one *parent*. Hierarchical DBMS have not gained any noticeable acceptance for use within GIS.

They are oriented for data sets that are very stable, where primary relationships among the data change infrequently or never at all. Also, the limitation on the number of parents that an element may have is not always conducive to actual geographic phenomenon.

Network Model

The network database organizes data in a network or *plex* structure. Any column in a plex structure can be linked to any other. Like a tree structure, a plex structure can be described in terms of *parents* and *children*.

This model allows for children to have more than one parent. Network DBMS have not found much more acceptance in GIS than the hierarchical DBMS. They have the same flexibility limitations as hierarchical databases; however, the more powerful structure for representing data relationships allows a more realistic modeling of geographic phenomenon. However, network databases tend to become overly complex too easily. In this regard it is easy to lose control and understanding of the relationships between elements.

Relational Model

The relational database organizes data in *tables*. Each table, is identified by a unique table name, and is organized by *rows* and *columns*. Each column within a table also has a unique name.

Columns store the values for a specific attribute, e.g. cover group, tree height. Rows represent one record in the table.

In a GIS each row is usually linked to a separate spatial feature, e.g. a forestry stand. Accordingly, each row would be comprised of several columns, each column containing a specific value for that geographic feature.

Object-Oriented Model

The object-oriented database model manages data through *objects*. An object is a collection of data elements and operations that together are considered a single entity. The object-oriented database is a relatively new model.

This approach has the attraction that querying is very natural, as features can be bundled together with attributes at the database administrator's discretion. To date, only a few GIS packages are promoting the use of this attribute data model.

However, initial impressions indicate that this approach may hold many operational benefits with respect to geographic data processing. Fulfillment of this promise with a commercial GIS product remains to be seen.

1.11. TYPES OF ATTRIBUTE DATA

Attribute data can be store as one of five different field types in a table or database: character, integer, floating, date, and BLOB.

1.11.1. Character Data

The character property (or string) is for text based values such as the name of a street or descriptive values such as the condition of a street. Character attribute data is stored as a series of alphanumeric symbols.

Aside from descriptors, character fields can contain other attribute values such as categories and ranks. For example, a character field may contain the categories for a street: avenue, boulevard, lane, or highway. A character field could also contain the rank, which is a relative ordering of features. For example, a ranking of the traffic load of the street with “1” being the street with the highest traffic.

Character data can be sorted in ascending (A to Z) and descending (Z to A) order. Since numbers are considered text in this field, those numbers will be sorted alphabetically which means that a number sequence of 1, 2, 9, 11, 13, 22 would be sorted in ascending order as 1, 11, 13, 2, 22, 9.

Because character data is not numeric, calculations (sum, average, median, etc.) can’t be performed on this type of field, even if the value stored in the field are numbers (to do that, the field type would need to be converted to a numeric field). Character fields can be summarized to produced counts (e.g. the number of features that have been categorized as “avenue”).

1.11.2. Numeric Data

Integer and floating are numerical values (see: the difference between floating and integer values). Within the integer type, the is a further division between short and long integer values. As would be expected, short integers store numeric values without fractional values for a shorter range than long integers. Floating point attribute values store numeric values with fractional values. Therefore, floating point values are for numeric values with decimal points (i.e numbers to the right of the decimal point as opposed to whole values).

Numeric values will be sorted in sequentially either in ascending (1 to 10) or descending (10 to 1) order. Numerical value fields can have operations performed such as calculating the sum or average value. Numerical field values can be a count (e.g. the total number of students at a school) or be a ratio (e.g. the percentage of students that are girls at a school).

1.11.3. Date/Time Data

Date fields contains date and time values.

1.11.4. BLOB Data

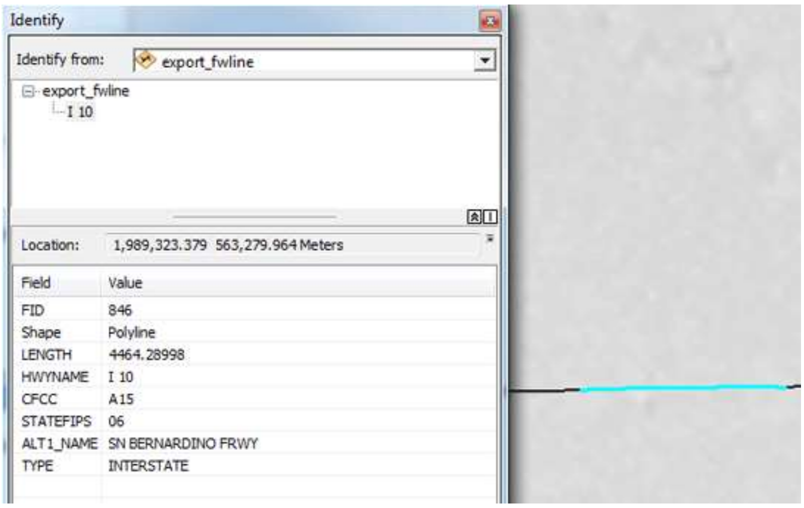


Fig.1.11. Attribute data for a road in gis.

BLOB stands for binary large object and this attribute type is used for storing information such images, multimedia, or bits of code in a field. This field stores object linking and embedding (OLE) which are objects created in other applications such as images and multimedia and linked from the BLOB field Attribute data for a road in gis.

1.12. SCALE OF MEASUREMENT/LEVEL OF MEASUREMENTS

Types of Attribute Data:

Types used in a GIS and in computer programming include character strings, integers, floating points or real numbers, dates and time intervals. Each field in an attribute table is defined with a data type, which applies to the domain of the field.

Another method is to define attribute data by measurement scale. The measurement scale concept groups attribute data into nominal, ordinal, interval and ratio data.

Nominal Data:

Nominal data describe different kinds of different categories of data such as land use types or soil types.

Ordinal Data:

Ordinal data differentiate data by ranking relationship. For example-cities may be grouped into large, medium and small cities by population size.

Interval Data:

Interval data have known intervals between values such as temperature reading. For example- a temperature reading of 700 F is warmer than 600 F by 100 F.

Ratio Data:

Ratio data are the same as interval data except that ratio data are based on a meaningful or absolute zero value. Population densities are an example of ratio data, because a density of 0 is an absolute zero.

Measurement scale (or level) of attribute can be broadly divided into two categories. Some attribute are measured in a numerical scale (such as job accessibility) whereas others are not (such as world language)

Distribution of Indian tribes & language:

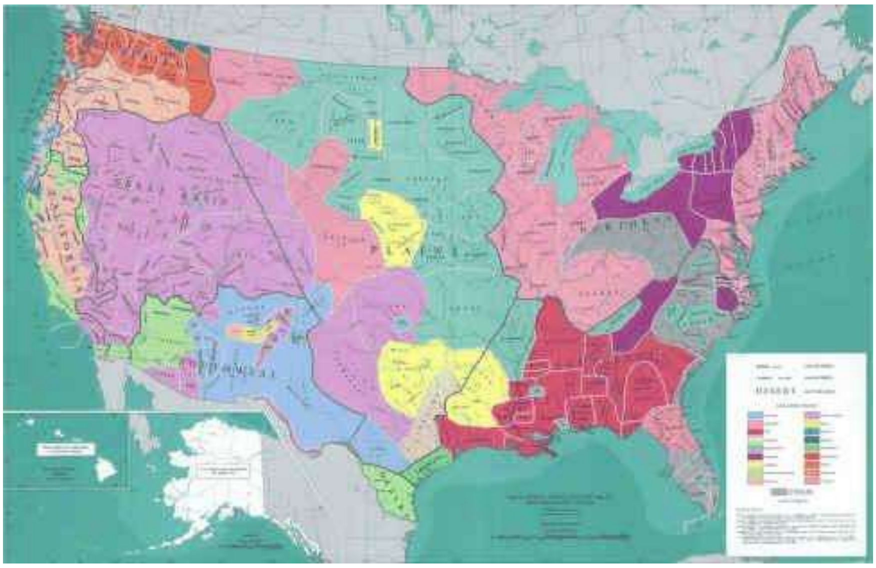


Fig.1.12. National Atlas of United States

This map can be fallen into qualitative thematic map because the measurement level of attribute portrayed (tribe/language) is nominal (not measured in number), and displays one or more particular themes.

Distribution of urban and rural population

Fig.1.13. National Atlas of United States

This map can be fallen into quantitative thematic map because the measurement level of attributed portrayed (population) is countable (measure in number), and displays one particular theme.

More specifically, measurement of attributes is organized into four levels: nominal, ordinal, interval, and ratio, listed in increasing order of sophistication of measurement

1.12.1. Nominal scaling

Only has a value either 0 or 1 (false or true)

Suppose we have two values say region A and region B; we can't determine if A > B or

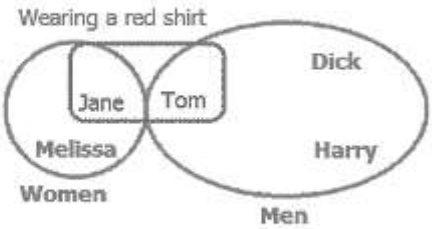
A < B, but we can determine if A=\=B or A = B.

e.g. agricultural region (corn regions, wheat regions, soy-bean regions)

Political party affiliation (Democrat, Republican, Independent)

Sex (male, female)

Response (yes, no)



1.12.2. Ordinal scaling

Value is arranged in a hierarchy of rank

Can determine if A > B or A < B, but can't determine how much they are different

e.g. social power (more, less)

agreement (strongly agree, strongly disagree)

Order of arrival of contestants in footrace

	Women's race	Men's race
First	Jane	Tom
Second	Melissa	Dick
Third	Leila	Harry

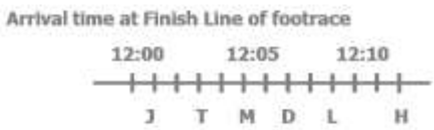
1.12.3. Interval scaling

Ranked

Know the distance between ranks

But it is not measured in an absolute scale; they are relative (has no natural origin)

e.g. Fahrenheit



1.12.4. Ratio scaling

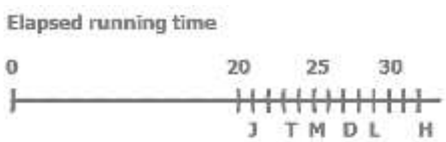
Ranked

Know the distance between ranks

It is measured in an absolute scale (has a natural origin)

e.g. weight, elevation

convey more information and permit more analytical treatment



Level of measurement prescribe the information required for an attribute reference system

Level of measurement	**Information required**
Nominal	Definitions of categories
Ordinal	Definitions of categories plus ordering
Interval	Unit of measure plus zero point
Ratio	Unit of measure

CHAPTER- 2

SPATIAL DATA MODELS

2.1. INTRODUCTION

Spatial data are what drive a GIS. Every functionality that makes a GIS separate from another analytical environment is rooted in the spatially explicit nature of the data.

Spatial data are often referred to as layers, coverage's, or layers. We will use the term layers from this point on, since this is the recognized term used in Arc-GIS. Layers represent, in a special digital storage format, features on, above, or below the surface of the earth. Depending on the type of features they represent, and the purpose to which the data will be applied, layers will be one of two major types.

a) Vector data represent features as discrete points, lines, and polygons.

b) Raster data represent the landscape as a rectangular matrix of square cells.

Depending on the type of problem that needs to be solved, the type of maps that need to be made, and the data source, either raster or vector, or a combination of the two can be used. Each data model has strengths and weaknesses in terms of functionality and representation. As you get more experience with GIS, you will be able to determine which data type to use for a particular application.

2.2. DATABASE STRUCTURES

The two basic data structures in any fully-functional GIS are:

Vector, e.g,

- ArcInfo Coverages
- ArcGIS Shape Files
- CAD (AutoCAD DXF & DWG, or Micro Station DGN files)
- ASCII coordinate data

Raster, e.g,

- ArcInfo Grids
- Images

- Digital Elevation Models (DEMs)
- generic raster datasets

2.3. DATA STRUCTURE MODELS

Data models are the conceptual models that describe the structures of databases. The structure of a database is defined by the data types, the constraints and the relationships for the description or storage of data. Following are the most often used data models:

1) Hierarchical Data Structure Model
2) Network Data Structure Model
3) Relational Data Structure Model
4) Object Oriented Data Structure Model

2.3.1. Hierarchical Data Structure Model

It is the earliest database model that is evolved from file system where records are arranged in a hierarchy or as a tree structure shown in the **figure.2.1.** Records are connected through pointers that store the address of the related record.

Each pointer establishes a parent-child relationship where a parent can have more than one child but a child can only have one parent. There is no connection between the elements at the same level. To locate a particular record, you have to start at the top of the tree with a parent record and trace down the tree to the child.

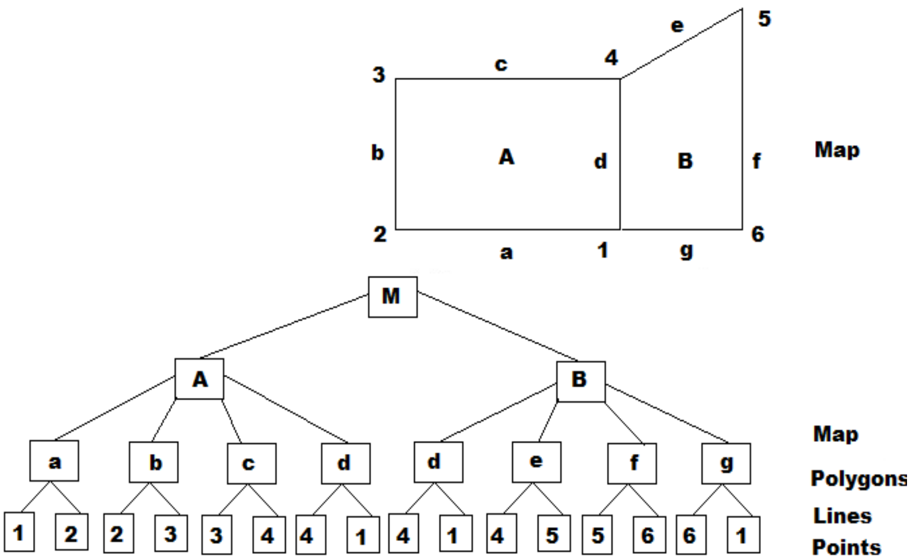


Fig.2.1 Hierarchical Database Structure Based on a Simple Map

Advantages

- Easy to understand: The organization of database parallels a family tree understanding which is quite easy.

- Accessing records or updating records are very fast since the relationships have been predefined.

Disadvantages

- Large index files are to be maintained and certain attribute values are repeated many times which lead to data redundancy and increased storage.
- The rigid structure of this model doesn't allow alteration of tables, therefore to add a new relationship entire database is to be redefined.

2.3.2. Network Data Structure Model

A network is a generalized graph that captures relationships between objects using connectivity shown figure.2.2. A network database consists of a collection of records that are connected to each other through links. A link is an association between two records. It allows each record to have many parents and many children thus allowing a natural model of relationships between entities.

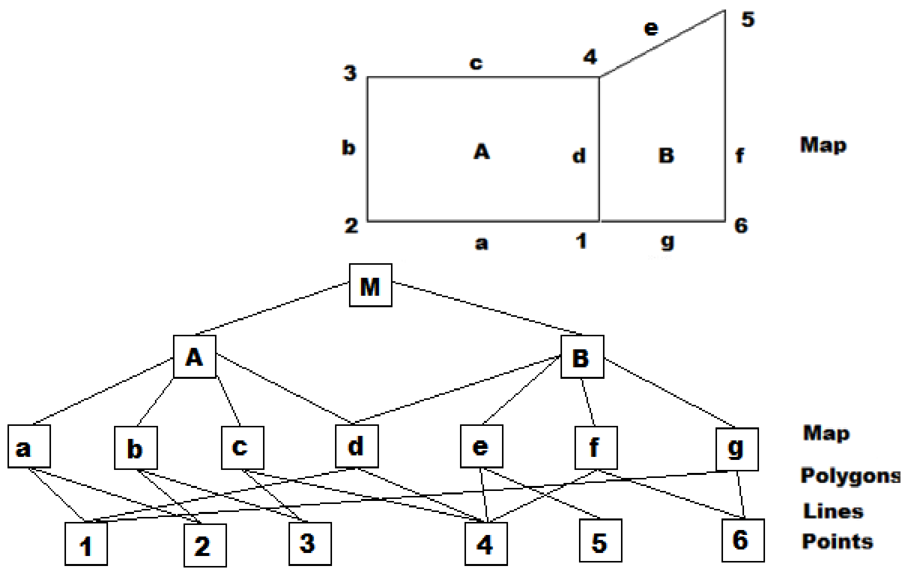


Fig.2.2. Network Data Structure Model

Advantages

- The many too many relationships are easily implemented in a network data model.
- Data access and flexibility in network model is better than that in hierarchical model. An application can access an owner record and the member records within a set.
- It enforces data integrity as a user must first define owner record and then the member records.
- The model eliminated redundancy but at the expense of more complicated relationships.

2.3.3. Relational Data Structure Model

- The relational data model was introduced by Codd in 1970. The relational database relates or connects data in different files through the use of a common field.
- A flat file structure is used with a relational database model. In this arrangement, data is stored in different tables made up of rows and columns as shown in **figure.2.3.**
- The columns of a table are named by attributes. Each row in the table is called a tuple and represents a basic fact.
- No two rows of the same table may have identical values in all columns.

Advantages

- The manager or administrator does not have to be aware of any data structure or data pointer. One can easily add, update, delete or create records using simple logic.

Disadvantages

- A few search commands in a relational database require more time to process compared with other database models.

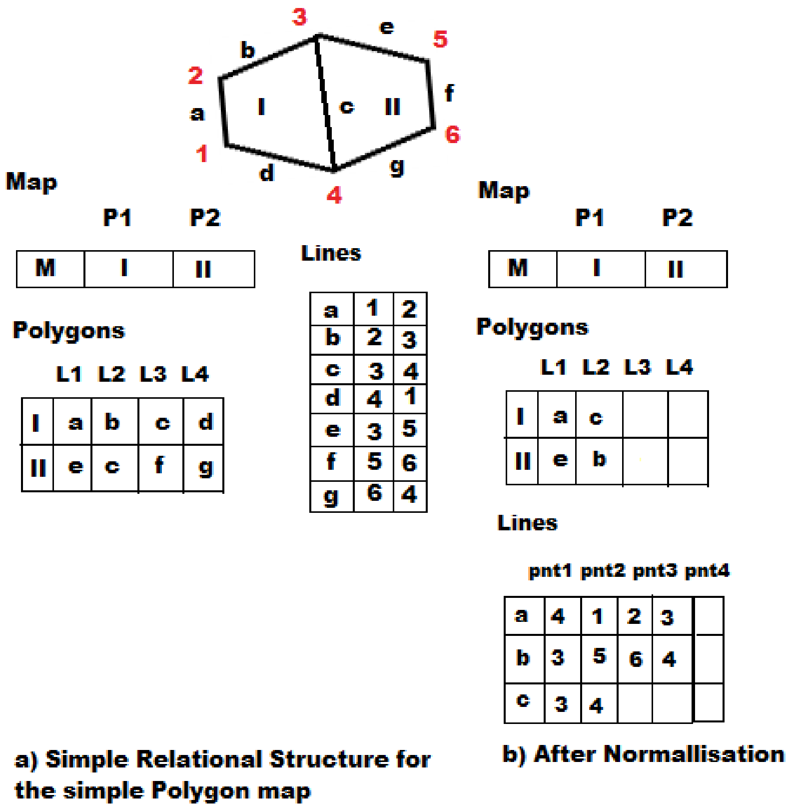


Fig.2.3. Relational Data Structure Model

2.3.4. Object Oriented Database Structure

- An Object Oriented model uses functions to model spatial and non-spatial relationships of geographic objects and the attributes.
- An object is an encapsulated unit which is characterized by attributes, a set of orientations and rules.

An object-oriented model has the following characteristics.

Generic Properties:	there should be an inheritance relationship.
Abtraction:	objects, classes and super classes are to be generated by classification, generalisation, association and aggregation.
Adhoc Queries:	users can order spatial operations to obtain spatial relationships of geographic objects using a special language.

- **For example,** let us try to represent a thought: "Hawaii is an island that is a state of USA" in GIS. In this case, we don't mind the geographic location with latitude and longitude in the conventional GIS model. This is not appropriate to use the layers. In an object-oriented model, we are more careful with spatial relationships for example, "is a" (the island is a land) and "part of" (the state is a part of the country).

- In addition, Hawaii (state) has Honolulu City and also is in Pacific Region. Figure **2.4 (a)** shows "is an" inheritance for the super class of land, while Figure **2.4 (b)** shows the spatial relationships for the object of the state.

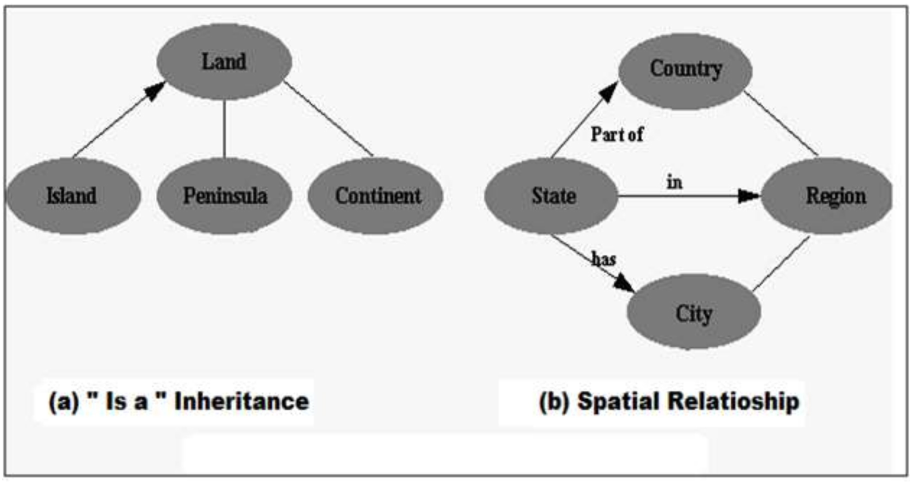


Fig.2.4. Object Oriented Database Structure

An object-oriented database is based on a semantic model as shown in **Figure 2.5.** Which is usually managed by a spatial language although the language has not yet been fully completed.

2.6 GIS

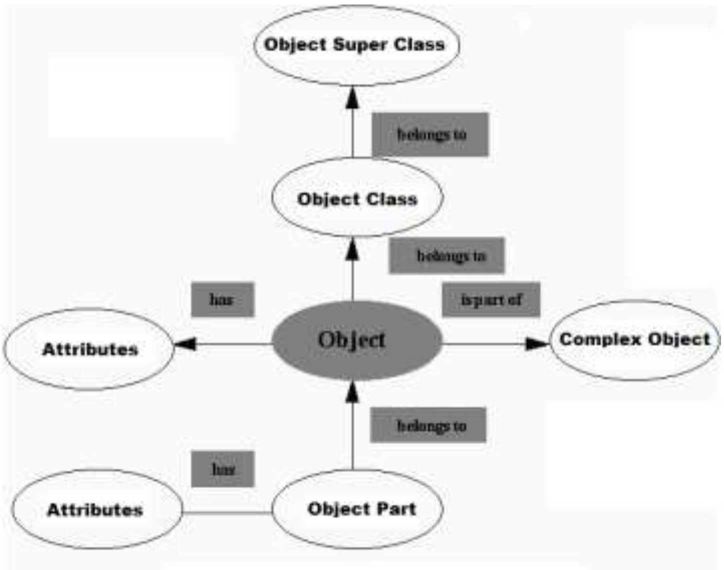


Fig.2.5. Object-Oriented Database is based on a Semantic Model

2.4. DEFINITION - WHAT DOES ENTITY-RELATIONSHIP DIAGRAM (ERD)

An entity-relationship diagram (ERD) is a data modeling technique that graphically illustrates an information system's entities and the relationships between those entities. An ERD is a conceptual and representational model of data used to represent the entity framework infrastructure.

The elements of an ERD are:

- Entities
- Relationships
- Attributes

Steps involved in creating an ERD include:

1) Identifying and defining the entities
2) Determining all interactions between the entities
3) Analyzing the nature of interactions/determining the cardinality of the relationships
4) Creating the ERD

2.4.1. Entity-Relationship Diagram (ERD)

An entity-relationship diagram (ERD) is crucial to creating a good database design. It is used as a high-level logical data model, which is useful in developing a conceptual design for databases.

An entity is a real-world item or concept that exists on its own. Entities are equivalent to database tables in a relational database, with each row of the table representing an instance of that entity.

An attribute of an entity is a particular property that describes the entity. A relationship is the association that describes the interaction between entities. Cardinality, in the context of ERD, is the number of instances of one entity that can, or must, be associated with each instance of another entity. In general, there may be one-to-one, one-to-many, or many-to-many relationships.

For example, let us consider two real-world entities, an employee and his department. An employee has attributes such as an employee number, name, department number, etc. Similarly, department number and name can be defined as attributes of a department. A department can interact with many employees, but an employee can belong to only one department, hence there can be a one-to-many relationship, defined between department and employee.

In the actual database, the employee table will have department number as a foreign key, referencing from department table, to enforce the relationship.

2.4.2. E-R DIAGRAM

E-R Diagram example from Database Management course

As mentioned before, the entity-relationship (E-R) diagram is one of the most commonly implemented conceptual data models used with GIS. Entities, attributes, and relationships are used to represent real-world features, what their properties are, and what the relationships are between these entities. Hardware and software issues are not explored in the E-R Diagram. These are addressed later in logical and physical data models. This first level of data abstraction is used by geospatial analysts as a starting point when analyzing and assessing the data available to them and how it fits together. The example below illustrates an E-R Diagram built during one of my Geospatial Data Structures course assignments.

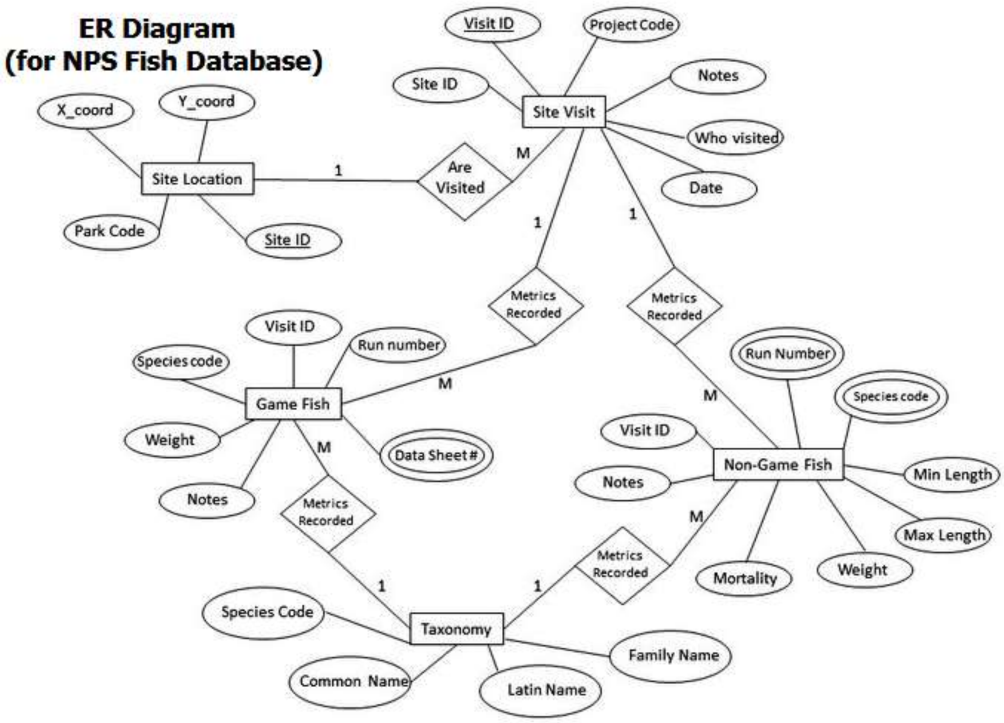
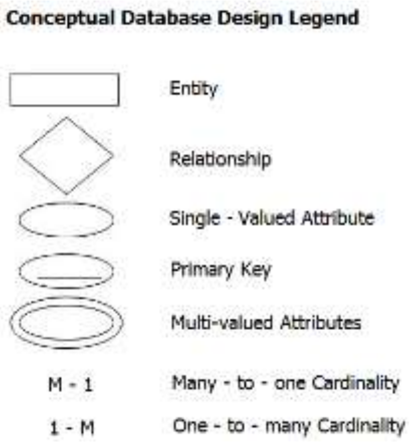


Fig.2.6. E-R Diagram



You could use this database to query how many of a particular species of game fish were examined at a specific park during a data range of interest. This would be non-spatial query because we are just counting an occurrence at one particular location. We are not using the coordinates to perform some type of buffer analysis or other spatial analysis to query the data.

2.5. SPATIAL DATA MODELS

Computers and GIS cannot directly be applied to the real world: a data gathering step comes first. Digital computers operate in numbers and characters held internally as binary digits. The real-world phenomenon of interest must be represented in symbolic form. The abstraction process of representing any property of the earth's surface in a computer accessible form involves the use of symbolic models.

Models are simplification of reality. A map is a symbolic model, because it is a simplified representation of part of the real world. The components of the model are spatial objects, approximating spatial entities of the real world; they are represented on the map by graphical symbols.

- The process of defining and organizing data about the real world into a consistent digital dataset that is useful and reveals information is called data modeling.
- The logical organization of data according to a scheme is known as data models
- **Data** can be defined as verifiable facts.
- **Information** is data organized to reveal patterns, and to facilitate search.
- **Spatial information** is difficult to extract from spatial data, unless the data are organized primarily by spatial attributes.
- **Spatial objects** are characterized by attributes that are both spatial and non-spatial, and the digital description of objects and their attributes comprise spatial datasets.
- **Spatial data** can be organized in different ways, depending on the way they are collected, how they are stored, and the purpose they are put.
- **A database** is a collection of inter-related data and everything that is needed to maintain and use it.

- A Database Management System is a collection of software for storing, editing and retrieving data in a database.

Traditionally spatial data has been stored and presented in the form of a map. Three basic types of spatial data models have evolved for storing geographic data digitally. These are referred to as:

- Vector;
- Raster;
- Image.

The following diagram reflects the two primary spatial data encoding techniques. These are vector and raster. Image data utilizes techniques very similar to raster data, however typically lacks the internal formats required for analysis and modeling of the data. Images reflect pictures or photographs of the landscape.

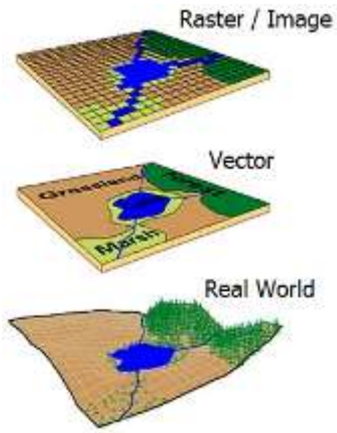


Fig.2.7. Two primary spatial data encoding techniques

2.5.1. Vector Data Formats

All spatial data models are approaches for storing the spatial location of geographic features in a database. Vector storage implies the use of vectors (directional lines) to represent a geographic feature. Vector data is characterized by the use of sequential points or vertices to define a linear segment. Each vertex consists of an X coordinate and a Y coordinate.

Vector lines are often referred to as arcs and consist of a string of vertices terminated by a node. A node is defined as a vertex that starts or ends an arc segment. Point features are defined by one coordinate pair, a vertex. Polygonal features are defined by a set of closed coordinate pairs. In vector representation, the storage of the vertices for each feature is important, as well as the connectivity between features, e.g. the sharing of common vertices where features connect.

Several different vector data models exist, however only two are commonly used in GIS data storage.

2.10 GIS

The most popular method of retaining spatial relationships among features is to explicitly record adjacency information in what is known as the topologic data model. Topology is a mathematical concept that has its basis in the principles of feature adjacency and connectivity.

The topologic data structure is often referred to as an intelligent data structure because spatial relationships between geographic features are easily derived when using them. Primarily for this reason the topologic model is the dominant vector data structure currently used in GIS technology. Many of the complex data analysis functions cannot effectively be undertaken without a topologic vector data structure. Topology is reviewed in greater detail later on in the book.

The secondary vector data structure that is common among GIS software is the computer-aided drafting (CAD) data structure. This structure consists of listing elements, not features, defined by strings of vertices, to define geographic features, e.g. points, lines, or areas. There is considerable redundancy with this data model since the boundary segment between two polygons can be stored twice, once for each feature. The CAD structure emerged from the development of computer graphics systems without specific considerations of processing geographic features. Accordingly, since features, e.g. polygons, are self-contained and independent, questions about the adjacency of features can be difficult to answer. The CAD vector model lacks the definition of spatial relationships between features that is defined by the topologic data model.

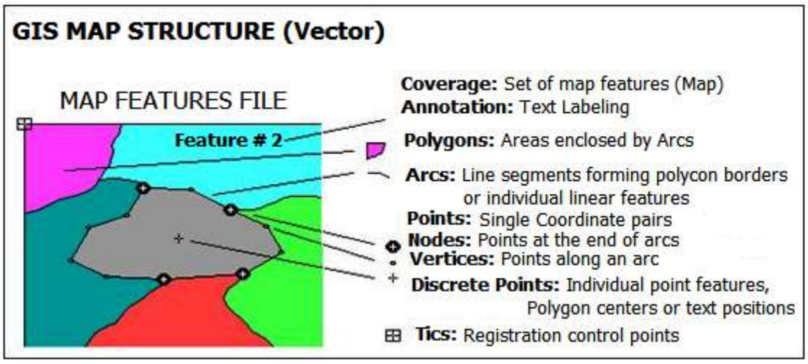


Fig.2.8. GIS MAP Structure - VECTOR systems (Adapted from Berry)

2.5.2. Raster Data Formats

Raster data models incorporate the use of a grid-cell data structure where the geographic area is divided into cells identified by row and column. This data structure is commonly called raster. While the term raster implies a regularly spaced grid other tessellated data structures do exist in grid based GIS systems. In particular, the quadtree data structure has found some acceptance as an alternative raster data model.

The size of cells in a tessellated data structure is selected on the basis of the data accuracy and the resolution needed by the user. There is no explicit coding of geographic coordinates required since that is implicit in the layout of the cells. A raster data structure is in fact a matrix where any coordinate can be quickly calculated if the origin point is known, and the size of the grid cells is known. Since grid-cells can be handled as two-dimensional arrays in computer encoding many analytical operations are easy to program. This makes tessellated data structures a popular choice for many GIS software. Topology is not a relevant concept with tessellated structures since adjacency and connectivity are implicit in the location of a particular cell in the data matrix.

Several tessellated data structures exist, however only two are commonly used in GIS's. The most popular cell structure is the regularly spaced matrix or raster structure. This data structure involves a division of spatial data into regularly spaced cells. Each cell is of the same shape and size. Squares are most commonly utilized.

Since geographic data is rarely distinguished by regularly spaced shapes, cells must be classified as to the most common attribute for the cell. The problem of determining the proper resolution for a particular data layer can be a concern. If one selects too coarse a cell size then data may be overly generalized. If one selects too fine a cell size then too many cells may be created resulting in a large data volume, slower processing times, and a more cumbersome data set. As well, one can imply accuracy greater than that of the original data capture process and this may result in some erroneous results during analysis.

As well, since most data is captured in a vector format, e.g. digitizing, data must be converted to the raster data structure. This is called vector-raster conversion. Most GIS software allows the user to define the raster grid (cell) size for vector-raster conversion. It is imperative that the original scale, e.g. accuracy, of the data be known prior to conversion. The accuracy of the data, often referred to as the resolution, should determine the cell size of the output raster map during conversion.

Most raster based GIS software requires that the raster cell contain only a single discrete value. Accordingly, a data layer, e.g. forest inventory stands, may be broken down into a series of raster maps, each representing an attribute type, e.g. a species map, a height map, a density map, etc. These are often referred to as one attribute maps. This is in contrast to most conventional vector data models that maintain data as multiple attribute maps, e.g. forest inventory polygons linked to a database table containing all attributes as columns. This basic distinction of raster data storage provides the foundation for quantitative analysis techniques. This is often referred to as raster or map algebra. The use of raster data structures allow for sophisticated mathematical modelling processes while vector based systems are often constrained by the capabilities and language of a relational DBMS.

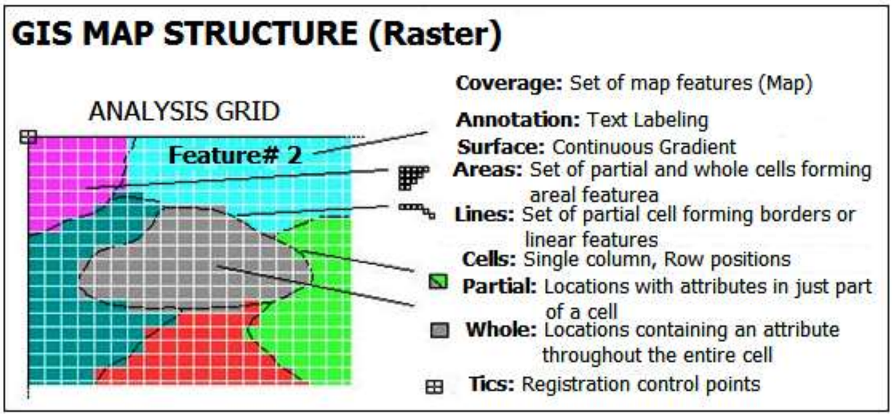


Fig.2.9. GIS MAP Structure - RASTER systems (Adapted from Berry)

This difference is the major distinguishing factor between vector and raster based GIS software. It is also important to understand that the selection of a particular data structure can provide advantages during the analysis stage. For example, the vector data model does not handle continuous data, e.g. elevation, very well while the raster data model is more ideally suited for this type of analysis. Accordingly, the raster structure does not handle linear data analysis, e.g. shortest path, very well while vector systems do. It is important for the user to understand that there are certain advantages and disadvantages to each data model.

The selection of a particular data model, vector or raster, is dependent on the source and type of data, as well as the intended use of the data. Certain analytical procedures require raster data while others are better suited to vector data.

2.5.3. Image Data

Image data is most often used to represent graphic or pictorial data. The term image inherently reflects a graphic representation, and in the GIS world, differs significantly from raster data. Most often, image data is used to store remotely sensed imagery, e.g. satellite scenes or orthophotos, or ancillary graphics such as photographs, scanned plan documents, etc. Image data is typically used in GIS systems as background display data (if the image has been rectified and georeferenced); or as a graphic attribute. Remote sensing software makes use of image data for image classification and processing. Typically, this data must be converted into a raster format (and perhaps vector) to be used analytically with the GIS.

Image data is typically stored in a variety of de facto industry standard proprietary formats. These often reflect the most popular image processing systems. Other graphic image formats, such as TIFF, GIF, PCX, etc., are used to store ancillary image data. Most GIS software will read such formats and allow you to display this data.

Fig.2.10. Image data is most often used for remotely sensed imagery such as satellite imagery or digital orthophotos.

2.5.4. Vector and Raster – Advantages and Disadvantages

There are several advantages and disadvantages for using either the vector or raster data model to store spatial data. These are summarized below.

Vector Data:

Advantages:

- Data can be represented at its original resolution and form without generalization.
- Graphic output is usually more aesthetically pleasing (traditional cartographic representation);
- Since most data, e.g. hard copy maps, is in vector form no data conversion is required.
- Accurate geographic location of data is maintained.
- Allows for efficient encoding of topology, and as a result more efficient operations that require topological information, e.g. proximity, network analysis.

Disadvantages:

- The location of each vertex needs to be stored explicitly. For effective analysis, vector data must be converted into a topological structure. This is often processing intensive and usually requires extensive data cleaning. As well, topology is static, and any updating or editing of the vector data requires re-building of the topology. Algorithms for manipulative and analysis functions are complex and may be processing intensive. Often, this inherently limits the functionality for large data sets, e.g. a large number of features.
- Continuous data, such as elevation data, is not effectively represented in vector form. Usually substantial data generalization or interpolation is required for these data layers.
- Spatial analysis and filtering within polygons is impossible

Raster Data

Advantages:

- The geographic location of each cell is implied by its position in the cell matrix. Accordingly, other than an origin point, e.g. bottom left corner, no geographic coordinates are stored.
- Due to the nature of the data storage technique data analysis is usually easy to program and quick to perform.
- The inherent nature of raster maps, e.g. one attribute maps, is ideally suited for mathematical modeling and quantitative analysis.
- Discrete data, e.g. forestry stands, is accommodated equally well as continuous data, e.g. elevation data, and facilitates the integrating of the two data types.
- Grid-cell systems are very compatible with raster-based output devices, e.g. electrostatic plotters, graphic terminals.

Disadvantages:

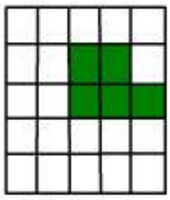
- The location of each vertex needs to be stored explicitly. For effective analysis, vector data must be converted into a topological structure. This is often processing intensive and usually requires extensive data cleaning. As well, topology is static, and any updating or editing of the vector data requires re-building of the topology.
- Algorithms for manipulative and analysis functions are complex and may be processing intensive. Often, this inherently limits the functionality for large data sets, e.g. a large number of features.
- Continuous data; such as elevation data, is not effectively represented in vector form.
- Usually substantial data generalization or interpolation is required for these data layers.
- Spatial analysis and filtering within polygons is impossible.

2.6. RASTER DATA STRUCTURE

In a simple raster data structure the geographical entities are stored in a matrix of rectangular cells. A code is given to each cell which informs users which entity is present in which cell. The simplest way of encoding a raster data into computers can be understood as follows:

(a) Entity model:

It represents the whole raster data. Let us assume that the raster data belongs to an area where land is surrounded by water. Here a particular entity (land) is shown in green color and the area where land is not present is shown by white.



(b) Pixel values:

The pixel value for the full image is shown. Cells having a part of the land are encoded as 1 and others where land is not present are encoded as 0.

0	0	0	0	0
0	0	1	1	0
0	0	1	1	1
0	0	0	0	0
0	0	0	0	0

(c) File structure:

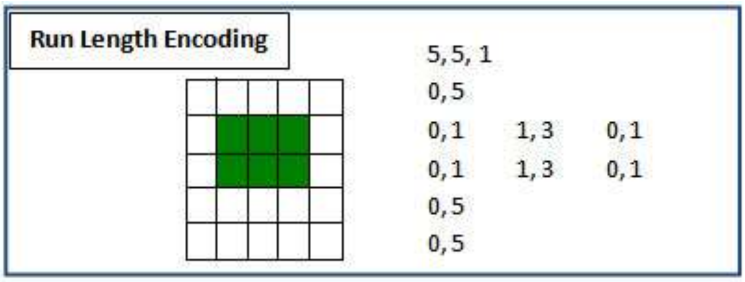
It demonstrates the method of coding raster data. The first row of the file structure data tells that there are 5 rows and 5 columns in the image, and 1 is the maximum pixel value. The subsequent rows have cells with value as either 0 or 1 (similar to pixel values).

5,	5,	1		
0,	0,	0,	0,	0
0,	0,	1,	1,	0
0,	0,	1,	1,	1
0,	0,	0,	0,	0
0,	0,	0,	0,	0

The huge size of the data is a major problem with raster data. An image consisting of twenty different land-use classes takes the same storage space as a similar raster map showing the location of a single forest. To address this problem many data compaction methods have been developed which are discussed below:

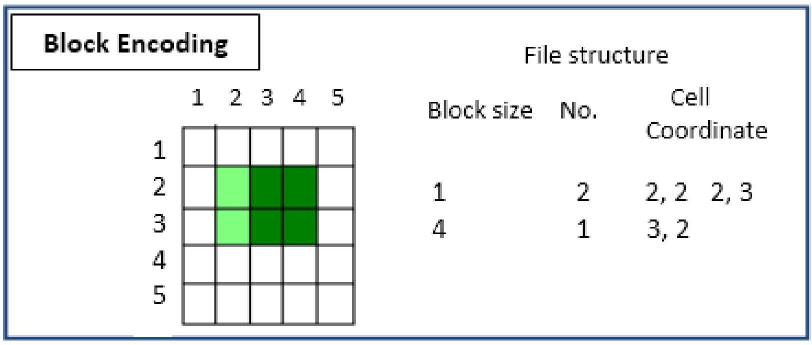
2.6.1. Run length encoding

- Reduction of data on a row by row basis
- Stores a single value for a group of cells rather than storing values for individual cells
- First line represents the dimension of the matrix (5×5) and the number of entities (1) present. In second and subsequent lines, the first number in the pair represents absence (0) or presence (1) of the entity and the second number indicates the number of cells referenced.



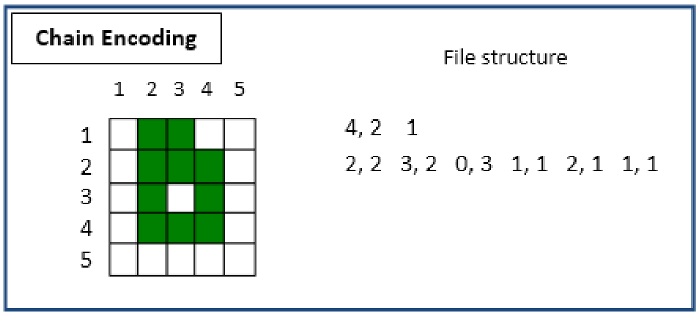
2.6.2. Block encoding

- Data is stored in blocks in the raster matrix.
- The entity is subdivided into hierarchical blocks and the blocks are located using coordinates.
- The first cell at top left hand is used as the origin for locating the blocks



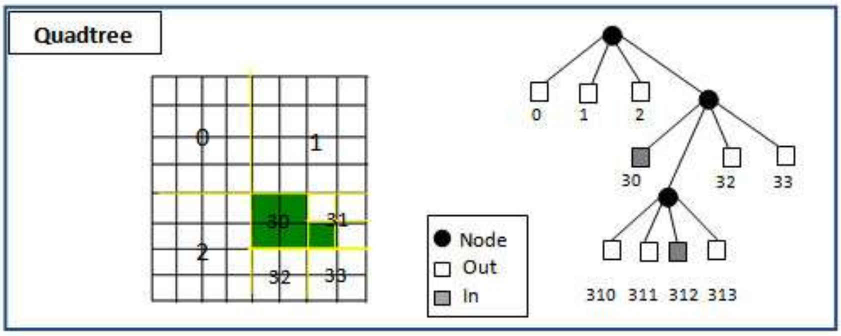
2.6.3. Chain encoding

- Works by defining boundary of the entity i.e. sequence of cells starting from and returning to the given origin
- Direction of travel is specified using numbers. (0 = North, 1 = East, 2 = South, 3 = West)
- The first line tells that the coding started at cell (4, 2) and there is only one chain. In the second line the first number in the pair tells the direction and the second number represents the number of cells lying in this direction.



2.6.4. Quadtree

- A raster is divided into a hierarchy of quadrants that are subdivided based on similar value pixels.
- The division of the raster stops when a quadrant is made entirely from cells of the same value.
- A quadrant that cannot be subdivided is called a leaf node.



A satellite or remote sensing image is a raster data where each cell has some value and together these values create a layer. A raster may have a single layer or multiple layers. In a multi-layer/ multi-band raster each layer is congruent with all other layers, have identical numbers of rows and columns, and have same locations in the plane. Digital elevation model (DEM) is an example of a single-band raster dataset each cell of which contains only one value representing surface elevation.

2.6.5. A single layer raster data can be represented using

(a) Two colors (binary):

The raster is represented as binary image with cell values as either 0 or 1 appearing black and white respectively.

Gray-scale:

Typical remote sensing images are recorded in an 8 bit digital system. A grayscale image is thus represented in 256 shades of gray which range from 0 (black) to 255 (white). However a human eye can't make distinction between the 255 different shades. It can only interpret 8 to 16 shades of gray.

A satellite image can have multiple bands, i.e. the scene/details are captured at different wavelengths (Ultraviolet- visible- infrared portions) of the electromagnetic spectrum. While creating a map we can choose to display a single band of data or form a color composite using multiple bands. A combination of any three of the available bands can be used to create RGB composites. These composites present a greater amount of information as compared to that provided by a single band raster.

2.6.6. Comparison between Vector and Raster Data Models

Data Model	Advantages	Disadvantages
Raster	Simple data structure	Cell size determines the resolution at which the data is represented
	Compatible with remote sensing or scanned data	Requires a lot of storage space
	Spatial analysis is easier	Projection transformations are time consuming
	Simulation is easy because each unit has the same size and shape	Network linkages are difficult to establish

Vector	Data is represented at its original resolution and form without generalization	The location of each vertex is to be stored explicitly
	Require less storage space	Overlay based on criteria is difficult
	Editing is faster and convenient	Spatial analysis is cumbersome
	Network analysis is fast	Simulation is difficult because each unit has a different topological form
	Projection transformations are easier	

2.7. DATA COMPRESSION

[Computing] The process of reducing the size of a file or database. Compression improves data handling, storage, and database performance. Examples of compression methods include quadtrees, run-length encoding, and wavelets.

Compression ratio:

- The compression ratio (that is , the size of the compressed file compared to that of the uncompressed file) of lossy video codec's is nearly always far superior to that of the audio and still-image equivalents Wavelet compression, used by raster formats such as MrSID,JPEG2000,andER Map per's ECW, takes time to decompress before drawing.
- Compression a series of techniques used for the reduction of space, bandwidth, cost, transmission, generating time, and the storage of data.
- It's a computer process using algorithms that reduces the size of electronic documents so they occupy less digital storage space.

2.7.1. Raster Data Compression

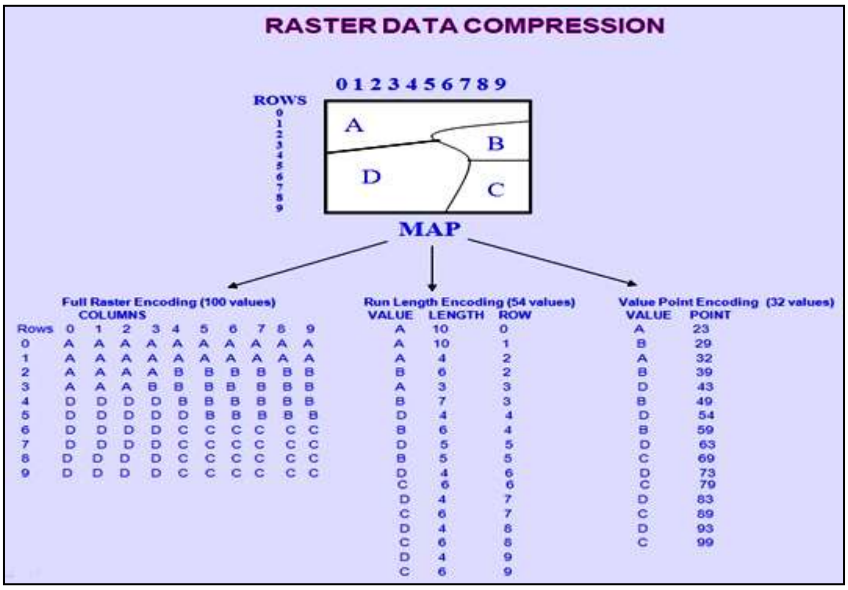


Fig.2.11. Raster Data Compression

Raster Data Compression

- Huge raster data has be to stored, retrieved, manipulated and analyzed.
- Large no. of thematic map layer is involved.
- Many repetitive characters are involved.
- Therefore, for better storage and to preserve highest possible degree of accuracy, we need to go for compact methods of storing.
- Common method is elimination of repetitive characters.

Original Data		Compacted Data	
Northing	**Easting**	**10,000**	**70,000**
10,234	70,565	234	565
10,245	70,599	245	599
10,167	70,423	167	423

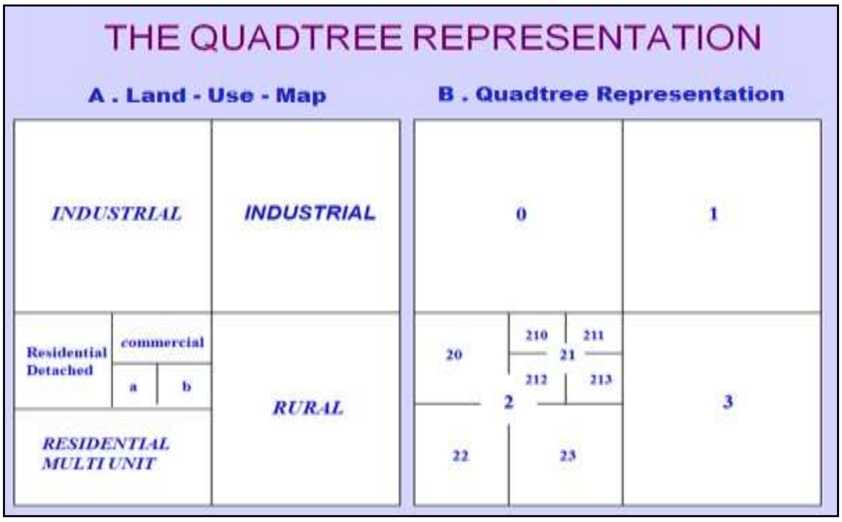


Fig.2.12. Raster Data Compression

2.7.2. Run length Encoding

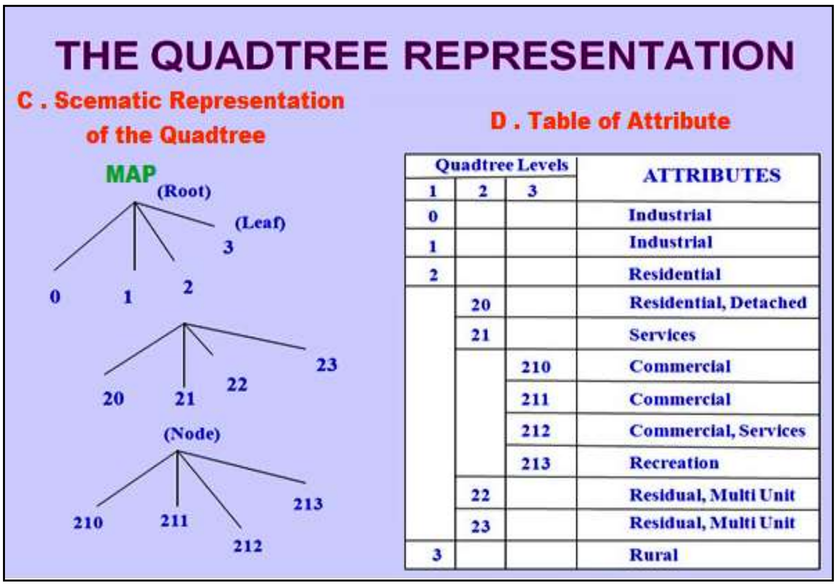
- Value often occur in runs across several cells, i.e., cells of the same value are often neighbours, like same soil type, or similar parameters.
- spatial auto-correlation exists –a tendency for nearby things to be more similar than distant things
- In run length encoding, the cells of the same value in arrow may be compacted by stating the value and their total.
- Thematic maps storage sizes get reduced using runlength encoding.
- Some raster GIS packages have the capability to handle run length encoded files.

Value point encoding

- Cells are assigned position numbers starting in the upper left corner proceeding from left to right and from top to bottom.
- The position no. for end of each run is stored in the point columns. The value for each cell in the run is in the value column.

2.7.3. Quadtree

- Typical type of raster model is dividing area into equal-sized rectangular cells .
- However, many cases, variable sized grid cell size used for more compact raster representation as shown figure.2.13.
- Larger cells used to represent large homogenous areas and smaller cells for finely details.
- Process involves regularly subdividing a map into four equal sized quadrants. Quadrant that has more than one class is again subdivided. Then; it is further subdivided within each quadrant until a square is found to be so homogenous that it is no longer needed to be divided.
- Then a Quadtree is prepared, resembling an inverted tree with "Root", i.e., a point from which all branches expand; Leaf is a lower most point and all other points in the tree are nodes.



Quadtree Levels 1	2	3	ATTRIBUTES
0			Industrial
1			Industrial
2			Residential
	20		Residential, Detached
	21		Services
		210	Commercial
		211	Commercial
		212	Commercial, Services
		213	Recreation
	22		Residual, Multi Unit
	23		Residual, Multi Unit
3			Rural

Fig.2.13. The Quadtree Representation

2.8. VECTOR DATA STRUCTURE

Geographic entities encoded using the vector data model, are often called features. The features can be divided into two classes:

a) Simple features

These are easy to create, store and are rendered on screen very quickly. They lack connectivity relationships and so are inefficient for modeling phenomena conceptualized as fields.

b) Topological features

A topology is a mathematical procedure that describes how features are spatially related and ensures data quality of the spatial relationships. Topological relationships include following three basic elements:

1) Connectivity: Information about linkages among spatial objects
2) Contiguity: Information about neighbouring spatial object
3) Containment: Information about inclusion of one spatial object within another spatial object

2.8.1. Connectivity

Arc node topology defines connectivity - arcs are connected to each other if they share a common node. This is the basis for many network tracing and path finding operations.

Arcs represent linear features and the borders of area features. Every arc has a from-node which is the first vertex in the arc and a to-node which is the last vertex. These two nodes define the direction of the arc. Nodes indicate the endpoints and intersections of arcs. They do not exist independently and therefore cannot be added or deleted except by adding and deleting arcs.

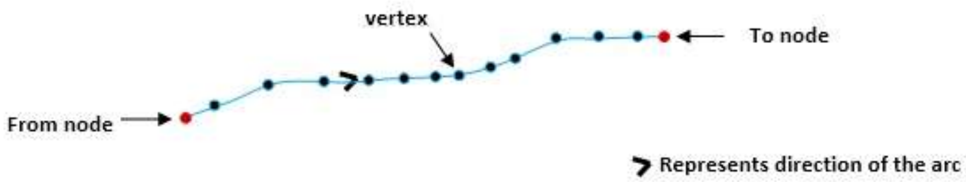


Fig.2.14. Arc-node Topology

Nodes can, however, be used to represent point features which connect segments of a linear feature (e.g., intersections connecting street segments, valves connecting pipe segments).

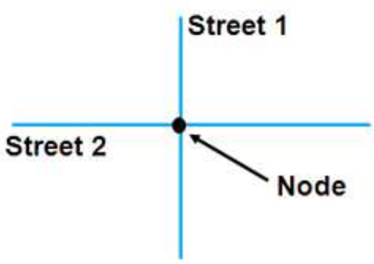
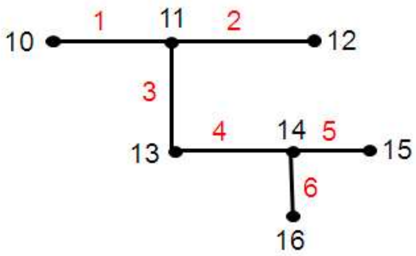


Fig.2.15. Node showing intersection

Arc-node topology is supported through an arc-node list. For each arc in the list there is a from node and a to node. Connected arcs are determined by common node numbers.

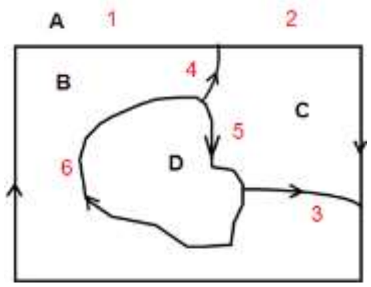


Arc-Node List		
Arc	*From node*	*To node*
1	10	11
2	11	12
3	11	13
4	13	14
5	14	15
6	14	16

Fig.2.16. Arc-Node Topology with list

2.8.2. Contiguity

Polygon topology defines contiguity. The polygons are said to be contiguous if they share a common arc. Contiguity allows the vector data model to determine adjacency.



Left-Right Topology		
Arc	Left Polygon	Right Polygon
1	A	B
2	A	C
3	C	B
4	B	C
5	C	D
6	B	D

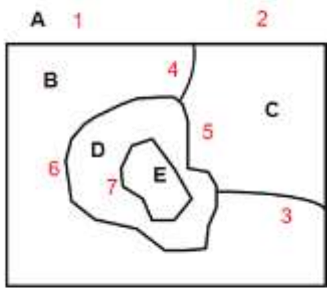
Fig2.17. Polygon Topology

The from node and *to* node of an arc indicate its direction, and it helps determining the polygons on its left and right side. Left-right topology refers to the polygons on the left and right sides of an arc. In the illustration above, polygon B is on the left and polygon C is on the right of the arc 4.

Polygon A is outside the boundary of the area covered by polygons B, C and D. It is called the external or universe polygon, and represents the world outside the study area. The universe polygon ensures that each arc always has a left and right side defined.

2.8.3. Containment

Geographic features cover distinguishable area on the surface of the earth. An area is represented by one or more boundaries defining a polygon.



Polygon arc topology	
Polygon	Arc List
B	1, 4, 6, 3
C	2, 3, 5, 4
D	5, 6, 0, 7
E	7

Fig2.18. Polygon arc topology

The polygons can be simple or they can be complex with a hole or island in the middle. In the illustration given below assume a lake with an island in the middle.

The lake actually has two boundaries, one which defines its outer edge and the other (island) which defines its inner edge. An island defines the inner boundary of a polygon. The polygon D is made up of arc 5, 6 and 7. The 0 before the 7 indicates that the arc 7 creates an island in the polygon.

Polygons are represented as an ordered list of arcs and not in terms of X, Y coordinates. This is called ***Polygon-Arc topology***. Since arcs define the boundary of polygon, arc coordinates are stored only once, thereby reducing the amount of data and ensuring no overlap of boundaries of the adjacent polygons.

2.8.4. Simple Features

Point entities:

These represent all geographical entities that are positioned by a single XY coordinate pair. Along with the XY coordinates the point must store other information such as what does the point represent etc.

Line entities:

Linear features made by tracing two or more XY coordinate pair.

- **Simple line:** It requires a start and an end point.
- **Arc:** A set of XY coordinate pairs describing a continuous complex line. The shorter the line segment and the higher the number of coordinate pairs, the closer the chain approximates a complex curve.

Simple Polygons:

Enclosed structures formed by joining set of XY coordinate pairs. The structure is simple but it carries few disadvantages which are mentioned below:

- Lines between adjacent polygons must be digitized and stored twice, improper digitization give rise to slivers and gaps
- Convey no information about neighbour
- Creating islands is not possible

2.8.5. Topologic Features

Networks:

A network is a topologic feature model which is defined as a line graph composed of links representing linear channels of flow and nodes representing their connections. The topologic relationship between the features is maintained in a connectivity table. By consulting connectivity table, it is possible to trace the information flowing in the network

Polygons with explicit topological structures:

Introducing explicit topological relationships takes care of islands as well as neighbours. The topological structures are built either by creating topological links during data input or using software. Dual Independent Map Encoding (DIME) system of US Bureau of the Census is one of the first attempts to create topology in geographic data.

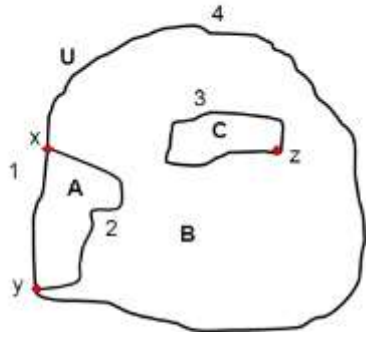


Fig.2.19. Polygon as a topological feature

- Polygons are formed using the lines and their nodes.
- Once formed, polygons are individually identified by a unique identification number.
- The topological information among the polygons is computed and stored using the adjacency information (the nodes of a line, and identifiers of the polygons to the left and right of the line) stored with the lines.

Poly ID	Arcs
A	1, 2
B	2, 3, 4
C	3

Arc ID	From	To
1	x	y
2	x	y
3	z	z
4	x	y

Arc ID	Left Poly	Right Poly
1	A	U
2	B	A
3	C	B
4	U	B

2.8.6. Fully topological polygon network structure

A fully topological polygon network structure is built using boundary chains that are digitized in any direction. It takes care of islands and lakes and allows automatic checks for improper polygons. Neighborhood searches are fully supported. These structures are edited by moving the coordinates of individual points and nodes, by changing polygon attributes and by cutting out or adding sections of lines or whole polygons. Changing coordinates require no modification to the topology but cutting out or adding lines and polygons requires recalculation of topology and rebuilding the database.

2.8.7. Triangular Irregular Network (TIN)

TIN represents surface as contiguous non-overlapping triangles created by performing Delaunay triangulation. These triangles have a unique property that the circum circle that passes through the vertices of a triangle contains no other point inside it. TIN is created from a set of mass points with x, y and z coordinate values. This topologic data structure manages information about the nodes that form each triangle and the neighbors of each triangle.

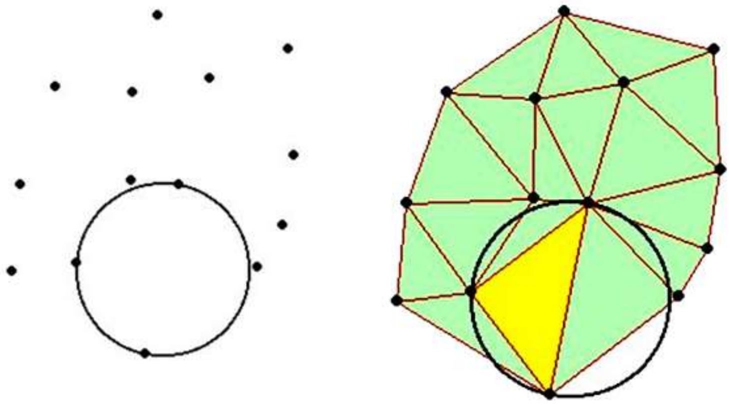
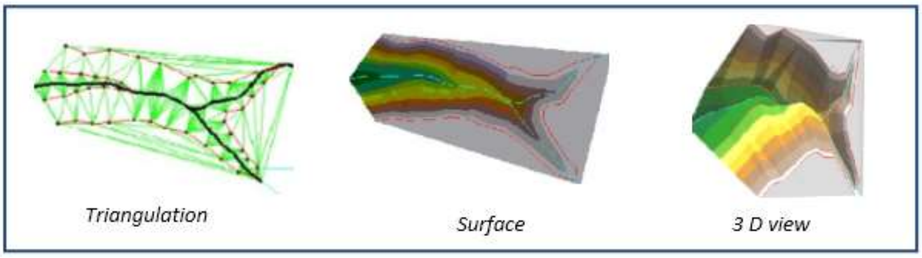


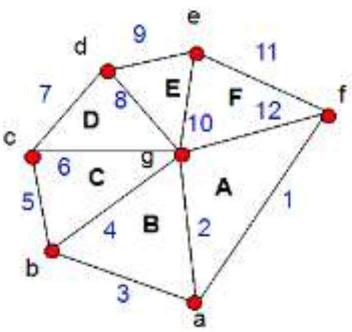
Fig.2.20. Delaunay Triangulation

Advantages of Delaunay triangulation

- The triangles are as equiangular as possible, thus reducing potential numerical precision problems created by long skinny triangles
- The triangulation is independent of the order the points are processed
- Ensures that any point on the surface is as close as possible to a node



Because points can be placed irregularly over a surface a TIN can have higher resolution in areas where surface is highly variable. The model incorporates original sample points providing a check on the accuracy of the model. The information related to TIN is stored in a file or a database table. Calculation of elevation, slope, and aspect is easy with TIN but these are less widely available than raster surface models and more time consuming in term of construction and processing.



Arc attribute table			
Edge ID	Length	From node	To node
1	160	f	a
2	140	a	g
3	130	a	b
4	140	b	g
...			

Arc attribute table			
Edge ID	Length	From node	To node
1	160	f	a
2	140	a	g
3	130	a	b
4	140	b	g
...			

Polygon attribute table					
Triangle ID	Area	Edge1	Edge2	Edge3	Neighbors
A	8200	1	2	12	B, F
B	7040	3	4	2	C, A
C	6000	5	6	4	D, B
D	5440	7	8	6	E, C
...					

The TIN model is a vector data model which is stored using the relational attribute tables. A TIN dataset contains three basic attribute tables: Arc attribute table that contains length, from node and to node of all the edges of all the triangles.

- Node attribute table that contains x, y coordinates and z (elevation) of the vertices
- Polygon attribute table that contains the areas of the triangles, the identification number of the edges and the identifier of the adjacent polygons.

Storing data in this manner eliminated redundancy as all the vertices and edges are stored only once even if they are used for more than one triangle. As TIN stores topological relationships, the datasets can be applied to vector based geo-processing such as automatic contouring, 3D landscape visualization, volumetric design, surface characterization etc.

2.9. RASTER VS VECTOR MODELS

The two primary types of spatial data are vector and raster data in GIS.

Data Model

The data model represents a set of guidelines to convert the real world (called entity) to the digitally and logically represented spatial objects consisting of the attributes and geometry. The attributes are managed by thematic or semantic structure while the geometry is represented by geometric-topological structure.

There are two major types of geometric data model;

1) Vector Data Model
2) Raster Data Model

Vector Data Model: [data models] A representation of the world using points, lines, and polygons (show in the figure 2.21). Vector models are useful for storing data that has discrete boundaries, such as country borders, land parcels, and streets.

Raster Data Model: [data models] A representation of the world as a surface divided into a regular grid of cells. Raster models are useful for storing data that varies continuously, as in an aerial photograph, a satellite image, a surface of chemical concentrations, or an elevation surface.

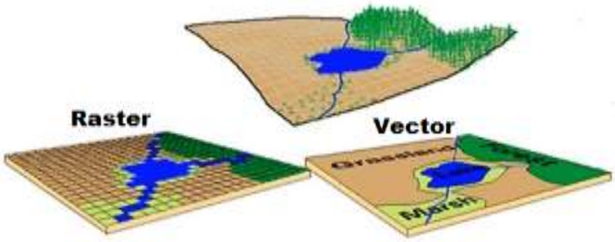


Fig.2.21. Example – Raster Data and Vector Data

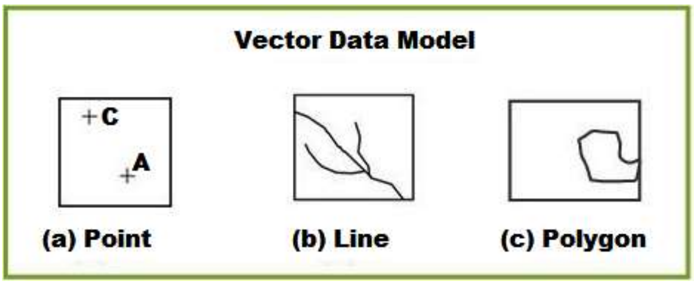


Fig .2.22. Example – Raster Data and Vector Data

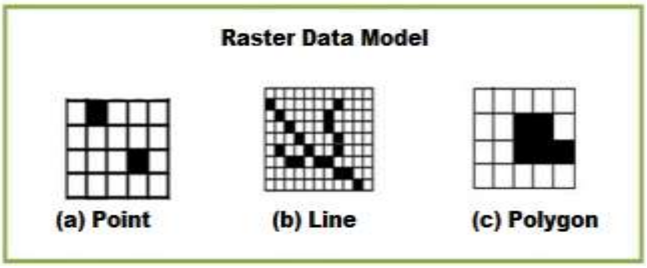


Figure.2.23. Example – Vector Data and Vector Data

2.9.1. Vector Data

Vector data (Show in Fig.2.22) is not made up of a grid of pixels. Instead, vector graphics are comprised of vertices and paths.

The three basic symbol types for vector data are

1) Points
2) Lines And
3) Polygons (areas).

Since the dawn of time, maps have been using symbols to represent real-world features. In GIS terminology, real-world features are called spatial entities.

The cartographer decides how much data needs to be generalized in a map. This depends on scale and how much detail will be displayed in the map. The decision to choose vector points, lines or polygons is governed by the cartographer and scale of the map.

(1) Points

Fig.2.24. Point Vector Data Type

Point Vector Data Type: Simple XY Coordinates

Vector points are simply XY coordinates. When features are too small to be represented as polygons, points are used indicate (fig.2.24.)

For Example: At a regional scale, city extents can be displayed as polygons because this amount of detail can be seen when zoomed in. But at a global scale, cities can be represented as points because the detail of city boundaries cannot be seen.

Vector data are stored as pairs of XY coordinates (latitude and longitude) represented as a point. Complementary information like street name or date of construction could accompany it in a table for its current use.

(2) Lines

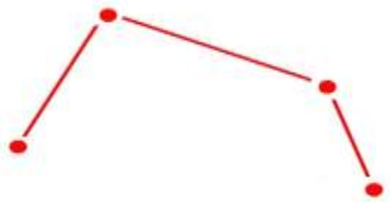


Fig.2.25. Vector Data Type Line

Vector Data Type Line:

Connect the dots and it becomes a line feature. Vector lines connect vertices with paths show in the fig (2.25). If you were to connect the dots in a particular order, you would end up with a vector line feature.

Lines usually represent features that are linear in nature. Cartographers can use a different thickness of line to show size of the feature. For Example, 500 meter Wide River may be thicker than a 50 meter wide river. They can exist in the real-world such as roads or rivers. Or they can also be artificial divisions such as regional borders or administrative boundaries.

Points are simply pairs of XY coordinates (latitude and longitude). When you connect each point or vertex with a line in a particular order, they become a vector line feature. Networks are line data sets but they are often considered to be different. This is because linear networks are topologically connected elements. They consist of junctions and turns with connectivity. If you were to find an optimal route using a traffic line network, it would follow one-way streets and turn restrictions to solve an analysis. Networks are just that smart.

(3) Polygons

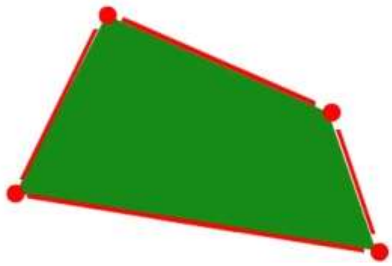


Fig.2.26. Vector Data Type Polygon

Vector Data Type Polygon: Connect the dots and enclose. It becomes a polygon feature when a set of vertices are joined in a particular order and closed; they become a vector Polygon feature shown the (fig.2.26). In order to create a polygon, the first and last coordinate pair is the same and all other pairs must be unique. Polygons represent features that have a two-dimensional area.

Examples of polygons are buildings, agricultural fields and discrete administrative areas. Cartographers use polygons when the map scale is large enough to be represented as polygons.

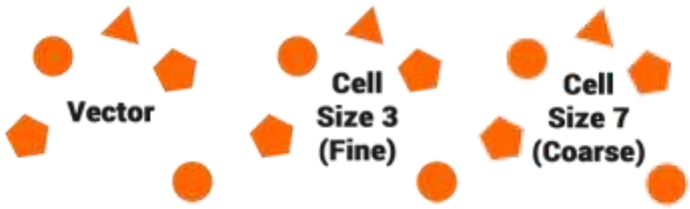
2.9.2. Raster Types: Discrete vs Continuous

Raster data is made up of pixels (also referred to as grid cells). They are usually regularly-spaced and square but they don't have to be. Rasters often look pixelated because each pixel has its own value or class.

For example:

Each pixel value in a satellite image has a red, green and blue value. Alternatively, each value in an elevation map represents a specific height. It could represent anything from rainfall to land cover.

Raster models are useful for storing data that varies continuously. For example, elevation surfaces, temperature and lead contamination.



Raster data models consist of 2 categories – discrete and continuous.

2.9.3. Discrete Raster's have Distinct Values

Discrete raster's have distinct themes or categories. For example, one grid cell represents a land cover class or a soil type.

In a discrete raster land cover/use map, you can distinguish each thematic class. Each class can be discretely defined where it begins and ends. In other words, each land cover cell is definable and it fills the entire area of the cell.

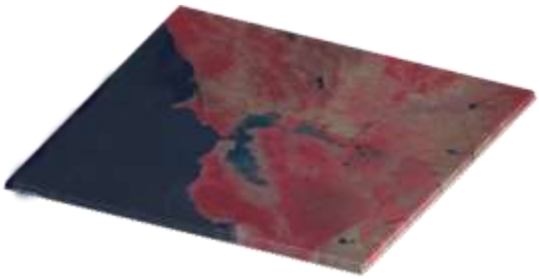
Discrete data usually consists of integers to represent classes. For example, the value 1 might represent urban areas; the value 2 represents forest and so on.

2.9.4. Continuous Rasters have Gradual Change

Continuous rasters (non-discrete) are grid cells with gradual changing data such as elevation, temperature or an aerial photograph.

A continuous raster surface can be derived from a fixed registration point. For example, digital elevation models use sea level as a registration point. Each cell represents a value above or below sea level. As another example, aspect cell values have fixed directions such as north, east, south or west.

Phenomena can gradually vary along a continuous raster from a specific source. In a raster depicting an oil spill, it can show how the fluid moves from high concentration to low concentration. At the source of the oil spill, concentration is higher and diffuses outwards with diminishing values as a function of distance.



2.10. VECTOR VS RASTER: SPATIAL DATA TYPES

It's not always straight-forward which spatial data type you should use for your maps.

In the end, it really comes down to the way in which the cartographer conceptualizes the feature in their map.

- **Do you want to work with pixels or coordinates?** Raster data works with pixels. Vector data consists of coordinates.
- **What is your map scale?** Vectors can scale objects up to the size of a billboard. But you don't get that type of flexibility with raster data
- **Do you have restrictions for file size?** Raster file size can result larger in comparison with vector data sets with the same phenomenon and area.

2.10.1. Vector and Raster – Advantages and Disadvantages

There are several advantages and disadvantages for using either the vector or raster data model to store spatial data. These are summarized below.

2.10.2. Vector Data:

Advantages:

- Data can be represented at its original resolution and form without generalization.
- Graphic output is usually more aesthetically pleasing (traditional cartographic representation);
- Since most data, e.g. hard copy maps, is in vector form no data conversion is required.
- Accurate geographic location of data is maintained.
- Allows for efficient encoding of topology, and as a result more efficient operations that require topological information, e.g. proximity, network analysis.

Disadvantages:

- The location of each vertex needs to be stored explicitly. For effective analysis, vector data must be converted into a topological structure. This is often processing intensive and usually requires extensive data cleaning. As well, topology is static, and any updating or editing of the vector data requires re-building of the topology. Algorithms for manipulative and analysis functions are complex and may be processing intensive. Often, this inherently limits the functionality for large data sets, e.g. a large number of features.
- Continuous data, such as elevation data, is not effectively represented in vector form. Usually substantial data generalization or interpolation is required for these data layers.
- Spatial analysis and filtering within polygons is impossible

2.10.3. Raster Data

Advantages :

- The geographic location of each cell is implied by its position in the cell matrix. Accordingly, other than an origin point, e.g. bottom left corner, no geographic coordinates are stored.
- Due to the nature of the data storage technique data analysis is usually easy to program and quick to perform.

- The inherent nature of raster maps, e.g. one attribute maps, is ideally suited for mathematical modeling and quantitative analysis.
- Discrete data, e.g. forestry stands, is accommodated equally well as continuous data, e.g. elevation data, and facilitates the integrating of the two data types.
- Grid-cell systems are very compatible with raster-based output devices, e.g. electrostatic plotters, graphic terminals.

Disadvantages:

- The location of each vertex needs to be stored explicitly. For effective analysis, vector data must be converted into a topological structure. This is often processing intensive and usually requires extensive data cleaning. As well, topology is static, and any updating or editing of the vector data requires re-building of the topology.
- Algorithms for manipulative and analysis functions are complex and may be processing intensive. Often, this inherently limits the functionality for large data sets, e.g. a large number of features.
- Continuous data; such as elevation data, is not effectively represented in vector form.
- Usually substantial data generalization or interpolation is required for these data layers.
- Spatial analysis and filtering within polygons is impossible.

2.11. TIN AND GRID DATA MODELS

TIN models

TIN stands for Triangular Irregular Network, which is a vector approach to handling a digital elevation model. TIN's are used to interpolate surfaces using multiple triangles. TIN's are able to interpolate surfaces by selecting representative points that are usually data points. TIN's connect these points to form a set of continuous and connected triangles. The data points consist of X, Y and Z values. The final result gives users a TIN surface.

Advantages of TIN models

TIN's give researchers the ability to view 2.5D and 3D at an area that was interpolated from minimal data collection.

- Users can describe a surface at different levels of resolution based on the points that were collected.
- TIN interpolation gives GIS users greater analytical capabilities. TIN models are easy to create and use.
- They provide users a simplified model that represents collected data points.
- Using a TIN surface in conjunction with Arc-Map extensions such as Spatial Analysis and 3D Analyst, TIN users can also derive slope, aspect, elevation, contour lines, hill shades, etc.

2.11.1. Different Types of TIN Methods and Processes

There are many different types of TIN interpolation methods. Some of the most popular TIN methods include:

- Natural Neighbour,
- Krigging,
- Spline,
- Nearest Neighbour and
- Inversed Distance Weighting.

These TIN interpolation methods use mathematical algorithms in order to generate interpolated surfaces. Each of these methods will produce different types of surfaces.

The TIN model (Triangulated Irregular Network):

A triangulated irregular network (TIN) is a digital data structure used in a geographic information system (GIS) for the representation of a surface.

A TIN is a vector based representation of the physical land surface or sea bottom, made up of irregularly distributed nodes and lines with three dimensional coordinates (x,y, and z) that are arranged in a network of non-overlapping triangles. TINs are often derived from the elevation data of a rasterized digital elevation model (DEM).

Structure of TIN Data Model

The TIN model represents a surface as a series of linked triangles, hence the adjective triangulated. Triangles are made from three points, which can occur at any location, giving the adjective, irregular. For each triangle, TIN records:

- The triangle number
- The numbers of each adjacent triangle
- The three nodes defining the triangle
- The x, y coordinates of each node
- The surface z value of each node
- The edge type of each triangle edge (hard or soft)

2.11.2. Components of TIN:

Nodes:

Nodes are the fundamental building blocks of the TIN. The nodes originate from the points and arc vertices contained in the input data sources. Every node is incorporated in the TIN triangulation. Every node in the TIN surface model must have a z value.

Edges:

Every node is joined with its nearest neighbors by edges to form triangles, which satisfy the Delaunay criterion. Each edge has two nodes, but a node may have two or more edges.

Because edges have a node with a z value at each end, it is possible to calculate a slope along the edge from one node to the other.

TIN:

Advantages - ability to describe the surface at different level of resolution, efficiency in storing data.

Disadvantages - in many cases require visual inspection and manual control of the network.

Automated hill shading:

The TIN model of terrain representation lends itself to development of an automated method of hill shading. Slope mapping is possible in TIN.

2.11.3. TIN Data Model

The Triangulated Irregular Network (TIN) data model is an alternative to the raster and vector data models for representing continuous surfaces. It allows surface models to be generated efficiently to analyze and display terrain and other types of surfaces. The TIN model creates a network of triangles by storing the topological relationships of the triangles. The fundamental building block of the TIN data is the node. Nodes are connected to their nearest neighbors by edges, according to a set of rules. Left-right topology is associated with the edges to identify adjacent triangles.

The TIN creates triangles from a set of points called mass points, which always become nodes. The user is not responsible for selecting; all the nodes are added according to a set of rules. Mass points can be located anywhere, the more carefully selected, the more accurate the model of the surface will be. Well-placed mass points occur when there is a major change in the shape of the surface, for example, at the peak of a mountain, the floor of a valley, or at the edge (top and bottom) of cliffs. By connecting points on a valley floor or along the edge of a cliff, a linear break in the surface can be defined. These are called break lines. Break lines can control the shape of the surface model.

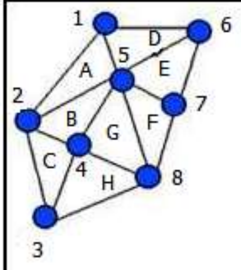
They always form edges of triangles and, generally, cannot be moved. A triangle always has three and only three straight sides, making their representation rather simple. A triangle is assigned a unique identifier that defines by its three nodes and its two or three neighboring triangles.

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 (geographic location) and the (elevation) Z values are encoded.

Four Tables for TIN Model

- Node Table it lists each triangle and the nodes which define it.
- Edge Table it lists three triangles adjacent to each facets. The triangles that border the boundary of the TIN show only two adjacent facets.

- XY Co-ordinate Table it lists the co-ordinate values of each node.
- Z Table it is the altitude value of each node.



X-Y Coordinate Table	
Node	Coordinate
1	X1Y1
2	X2Y2
3	X3Y3
4	X4Y4
5	X5Y5
6	X6Y6
7	X7Y7
8	X8Y8

Z- Coordinate Table	
Node	Elevation or Attribute
1	Z1
2	Z2
3	Z3
4	Z4
5	Z5
6	Z6
7	Z7
8	Z8

EDGES	
▲	Adiacent ▲
A	B, D
B	A, G, C
C	B,H
D	A, E
E	D, F
F	E, G
G	B, F, H
H	C, G

NODES	
▲	Nodes
A	1, 5, 2
B	2, 5, 4
C	2, 4, 3
D	1, 6, 5
E	5, 6, 7
F	5, 7, 8
G	5, 8, 4
H	3, 4, 8

2.12. GRID/LUNR/MAGI

In this model each grid cell is referenced or addressed individually and is associated with identically positioned grid cells in all other coverage's, rather than like a vertical column of grid cells, each dealing with a separate theme. Comparisons between coverage's are therefore performed on a single column at a time. Soil attributes in one coverage can be compared with vegetation attributes in a second coverage. Each soil grid cell in one coverage can be compared with a vegetation grid cell in the second coverage. The advantage of this data structure is that it facilitates the multiple coverage analysis for single cells. However, this limits the examination of spatial relationships between entire groups or themes in different coverage's.

2.12.1. Imgrid GIS

To represent a thematic map of land use that contains four categories: recreation, agriculture, industry and residence, each of these features have to be separated out as an individual layer. In the layer that represents agriculture 1 or 0 will represent the presence or absence of crops respectively. The rest of layer will be represented in the same way, with each variable referenced directly. The major advantage of IMGRID is its two-dimensional array of numbers resembling a map-like structure. The binary character of the information in each coverage simplifies long computations and eliminates the need for complex map legends. Since each coverage feature is uniquely identified, there is no limitation of assigning a single attribute value to a single grid cell. On the other side, the main problem related to information storage in an IMGRID structure is the excessive volume of data stored. Each grid cell will contain more than 1 or 0 values from more than one coverage and a large number of coverages are needed to store different types of information.

2.13. OPEN GEOSPATIAL CONSORTIUM (OGC)

The Open Geospatial Consortium (OGC), an international voluntary consensus standards organization, originated in 1994. In the OGC, more than 500 commercial, governmental, nonprofit and research organizations worldwide collaborate in a consensus process encouraging development and implementation of open standards for geospatial content and services, sensor web and Internet of Things, GIS data processing and data sharing.

2.13.1. Standards

Most of the OGC standards depend on a generalized architecture captured in a set of documents collectively called the Abstract Specification, which describes a basic data model for representing geographic features. Atop the Abstract Specification members have developed and continue to develop a growing number of specifications, or standards to serve specific needs for interoperable location and geospatial technology, including GIS.

The OGC standards baseline comprises more than thirty standards, including:

- CSW – Catalog Service for the Web: access to catalog information
- GML – Geography Mark-up Language: XML-format for geographical information
- Geo-XACML – Geospatial eXtensible Access Control Mark-up Language
- KML – Keyhole Mark-up Language: XML-based language schema for expressing geographic annotation and visualization on existing (or future) Web-based, two-dimensional maps and three-dimensional Earth browsers
- Observations and Measurements
- OGC Reference Model – a complete set of reference models
- OLS – Open Location Service (Open-LS)
- OGC Web Services Context Document defines the application state of an OGC Integrated Client
- OWS – OGC Web Service Common
- SOS – Sensor Observation Service
- SPS – Sensor Planning Service
- Sensor-ML – Sensor Model Language
- Sensor Things API - an open and unified framework to interconnect IoT devices, data, and applications over the Web. Currently a candidate standard waiting for votes.
- SFS – Simple Features – SQL

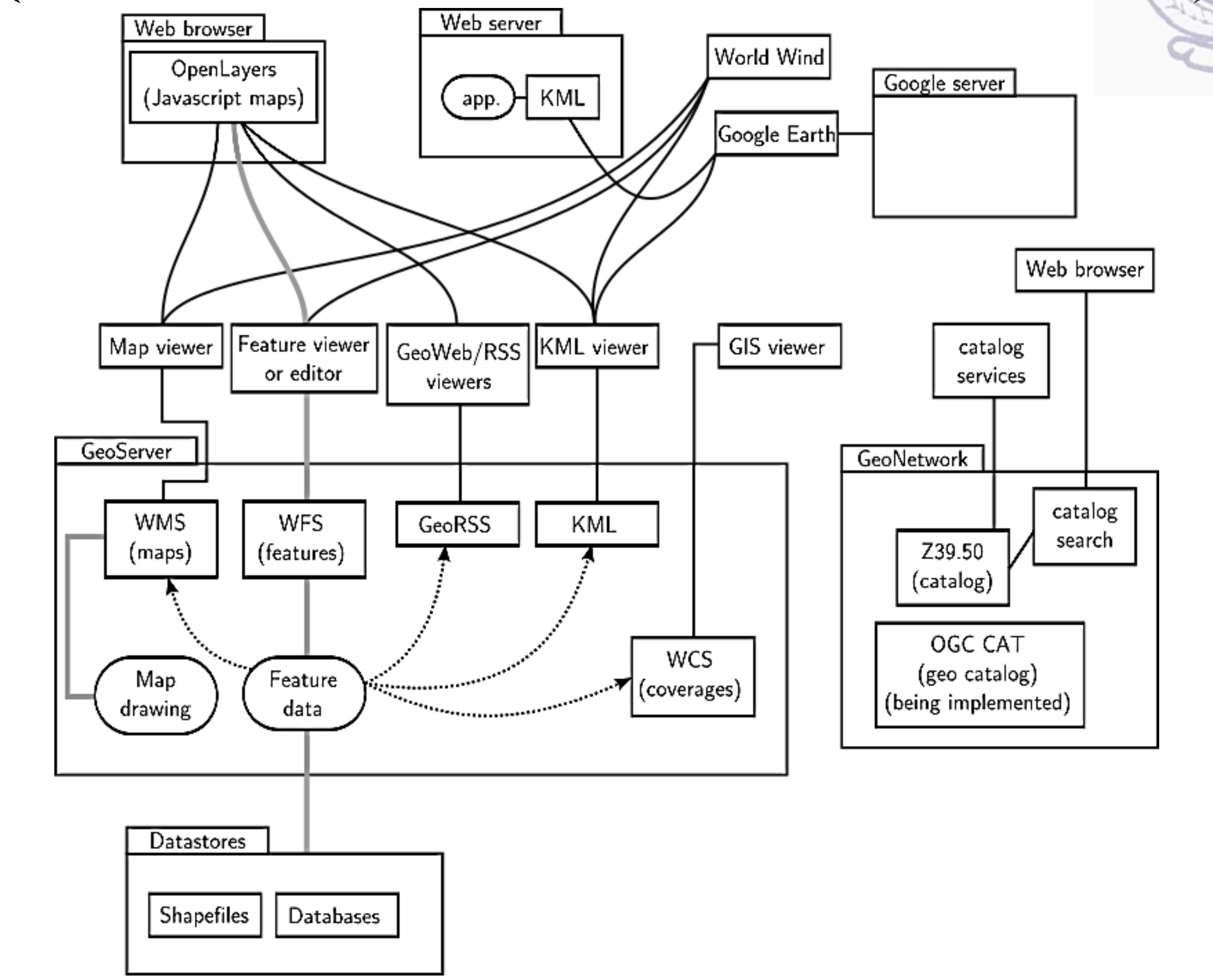


Fig.2.27. Relationship between clients/servers and OGC protocols

- SLD - Styled Layer Descriptor
- SRID, an identification for spatial coordinate systems
- Water-ML – Information model for the representation of hydrological observation data
- WCS – Web Coverage Service: provides access, sub setting, and processing on coverage objects
- WCPS – Web Coverage Processing Service: provides a raster query language for ad-hoc processing and filtering on raster coverage's
- WFS – Web Feature Service: for retrieving or altering feature descriptions
- WMS – Web Map Service: provides map images
- WMTS – Web Map Tile Service: provides map image tiles
- WPS – Web Processing Service: remote processing service
- Geo-SPARQL – Geographic SPARQL Protocol and RDF Query Language: representation and querying of geospatial data for the Semantic Web

- WTS – Web Terrain Service (WTS)

The design of standards were originally built on the HTTP web services paradigm for message-based interactions in web-based systems, but meanwhile has been extended with a common approach for SOAP protocol and WSDL bindings. Considerable progress has been made in defining Representational State Transfer (REST) web services, e.g., OGC Sensor Things API.

2.14. SPATIAL DATA QUALITY

Introduction

GIS developers and users have paid little attention to the problems caused by error, inaccuracy and imprecision in spatial data sets. There was awareness that all data suffers from inaccuracy and imprecision, but effects on GIS problems and solutions were not considered. It is now generally recognized that error, inaccuracy and imprecision can “make or break” GIS projects-making the results of a GIS analysis worthless. Spatial analyses done manually can easily align map boundaries to overlap and be registered. An automated GIS cannot do this, unless it is programmed to recognize the “undershoots, overshoots, and slivers” to connect lines. The level of the data quality must be made clear for the GIS to operate correctly. Assessing the quality of the data, however, may be costly. Data quality generally refers to the relative accuracy and precision of a particular GIS database. Error encompasses both the imprecision of data and its inaccuracies.

Although the term "garbage in, garbage out" certainly applies to GIS data, there are other important data quality issues besides the input data that need to be considered.

2.14.1. Components of Data Quality

There are three main components of data quality.

(i) Micro level components

(ii) Macro level components

(iii) Usage components.

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 and logical consistency given as follows:

a) Position Accuracy

b) Attribute Accuracy

c) Logical Consistency

Position Accuracy

Position accuracy is the expected deviance in the geographical location of an object in the data set (e.g. on a map) from its true ground position. Selecting a specified sample of points in a prescribed manner and comparing the position coordinates with an independent and more accurate source of information usually test it. There are two components to position accuracy: the bias and the precision.

Attribute Accuracy

Attributes may be discrete or continuous variables. A discrete variable can take on only a finite number of values whereas a continuous variable can take on any number of values. Categories like land use class, vegetation type, or administrative area are discrete variables. They are, in effect, ordered categories where the order indicates the hierarchy of the attribute.

Logical Consistency

Logical consistency refers to how well logical relations among data elements are maintained. It also refers to the fidelity of relationships encoded in the database, they may refer to the geometric structure of the data model (e.g. topologic consistency) or to the encoded attribute information e.g. semantic consistency).

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 major macro level components are:

a) Completeness

b) Time

c) Lineage

(a) Completeness

Completeness refers to the exhaustiveness of the information in terms of spatial and attribute properties encoded in the database. It may include information regarding feature selection criteria, definition and mapping rules and the deviations from them. The tests on spatial completeness may be obtained from topological test used for logical consistency whereas the test for attribute completeness is done by comparison of a master list of geo-codes to the codes actually appearing in the database.

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

The completeness of coverage is the proportion of data available for the area of interest.

Completeness of classification is an assessment of how well the chosen classification is able to represent the data. 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.

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.

(b) Time

Time is a critical factor in case of any type of data. Some data will be significantly biased depending on the time period over which they are collected.

Example:

Demographic information is usually very time sensitive. It can change significantly over a year. Land cover will change quickly in an area of rapid urbanization.

(c) Lineage

The lineage of a data set is its history, the source data and processing steps used to produce it. The source data may include transaction records, field notes etc. Ideally, some indication of lineage should be included with the data set since the internal documents are rarely available and usually require considerable expertise to evaluate. Unfortunately, lineage information most often exists as the personal experience of a few staff members and is not readily available to most users.

2.14.2. Usage Components

The usage components of data quality are specific to the resources of the organization. The effect of data cost, for example, depends on the financial resources of the organization. A given data set may be too expensive for one organization and be considered inexpensive by another.

Accessibility refers to the ease of obtaining and using the 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. Census data are usually restricted in this way. 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.

The direct cost of a data set purchased from another organization is usually well known: it is the price paid for the data. However, when the data are generated within the organization, the true cost may be unknown. Assessing the true cost of these data is usually difficult because the services and equipment used in their production support other activities as well.

The indirect costs include all the time and materials used to make use of the data. When data are purchased from another organization, the indirect costs may actually be more significant than the direct ones.

It may take longer for staff to handle data with which they are unfamiliar, or the data may not be compatible with the other data sets to be used.

2.14.3. Causes of Error

In this section, it is examined when and how errors creep into GIS data. The three major causes of GIS data error are problems found in (i) source data, (ii) data entry and (iii) data analysis.

Errors in Source Data

It has become common now to collect GIS data directly in the field. Data collection can be done using field survey instruments that download data directly into GIS or via GPS receivers that directly interface with GIS software on portable PCs. These techniques can eliminate the need for GIS source data.

But during the last many years, GIS data most often have been digitized from several sources, including hard copy maps, rectified aerial photography and satellite imagery. Hard-copy maps (e.g. paper, vellum and plastic film) may contain unintended production errors as well as unavoidable or even intended errors in presentation. The following are "errors" commonly found in maps.

2.14.4. Map Generalization

Cartographers often deliberately misrepresent map features due to limitations encountered when working at given map scales. Complex area features such as industrial buildings may have to be represented as simple shapes. Linear features such as roads may have to be represented by parallel lines that appear wider on a map. Curvilinear features such as streams may have to be represented without their smaller twists and bends.

Indistinct Boundaries

Indistinct boundaries typically include the borders of vegetated areas, soil types, wetlands and land use areas. In the real world, such features are characterized by gradual change, but cartographers represent these boundaries with a distinct line. Some compromise is inevitable.

Map Scale

Cartographers and photogrammetrists work to accepted levels of accuracy for a given map scale as per National Map Accuracy Standards. Locations of map features may disagree with actual ground locations, although the error likely will fall within specified tolerances. Of course, the problem is compounded by limitations in linear map measurements-typically about 1/100th of an inch on a map scale.

Map Symbology

It is impossible to perfectly depict the real world using lines, colors, symbols and patterns. Cartographers work with certain accepted conventions. As a result, facts and features represented on maps often must be interpreted or interpolated, which can produce errors. For example, terrain elevations typically are depicted using topographic contour lines and spot elevations. Elevations of the ground between the lines and spots must be interpolated. Also, areas symbolized as "forest" may not depict all open areas among the trees.

Errors during Data Entry

GIS data typically are created from hard copy source data. The process often is called "digitization", because the source data are converted to a computerized (digital) format. Human digitization can compound errors in source data as well as introduce new errors. The following are the primary methods of digitizing hard copy source data:

Manual Digitizing

Although manual digitizing is used less often today, it was the predominant digitizing method in the 1980s. Maps are affixed to digitizing tables, registered to a GIS coordinate system and "traced" into a GIS. A digitizing table has embedded in its surface a fine grid of wires that sense the position of a cross hair on a hand held cursor. When a cursor button is pressed, the system records a point at that location in the GIS database. The operator also identifies the type of feature being digitized as well as its attributes.

Photogrammetric mapping also is a manual digitizing process. Through an exacting and rigorous technical process of aerotriangulation, overlapping pairs of aerial photographs are registered to one another and viewed as a 3-D image in a stereoplotter or via special 3-D viewers. In a process called "stereocompilation," a photogrammetrist traces map features that are encoded directly into a database.

Scanning and Keyed Data Entry

In scanning, source data are mechanically read by a device that resembles a large format copy machine. Sensors encode the image as a large array of dots, much like a fax machine scans a letter. High resolution scanners can capture data at 2,000 dots per inch (dpi), but maps and drawings typically are scanned at 100 dpi to 400 dpi. The resulting raster image then is processed and displayed on a computer screen. Further onscreen manual digitizing (i.e. "heads-up digitizing") usually is needed to complete the data entry process. If the source data contain coordinate values for points or the bearings and distances of lines (e.g. parcel lines), then map features can be keyed into a GIS with great precision.

General Data Entry

Accurate digitizing is not easy. It requires certain basic physical and visual skills as well as training, patience and concentration. There also are many opportunities for error, because the process is subject to visual and mental mistakes, fatigue, distraction and involuntary muscle movements. In addition, the "set up" of a map on a digitizing table or a scanned raster image can produce errors. Cell size of a scanned raster image also can affect the accuracy of heads- up digitizing.

A digitizer must accurately discern the centre of a line or point as well as accurately trace it with a cursor. This task is especially prone to error if the map scale is small and the lines or symbols are relatively thick or large. The method of digitizing curvilinear lines also affects accuracy. "Point-mode" digitizing, for example, places sample points at selected locations along a line to best represent it in a GIS. The process is subject to judgment of the digitizer who selects the number and placement of data points. "Stream-mode" digitizing collects data points at a pre-set frequency, usually specified as the distance or time between data points. Every time an operator strays from an intended line, a point digitized at that moment would be inaccurate. This method also collects more data points than may be needed to faithfully represent a map feature. Therefore, post-processing techniques often are used to "weed out" unneeded data points.

Heads-up digitizing often is preferred over table digitizing, because it typically yields better results more efficiently. Keyed data entry of land parcel data is the most precise method. Moreover, most errors are fairly obvious, because the source data usually are carefully computed and thoroughly checked. Most keyed data entry errors show as obvious mismatches in the parcel "fabric."

GIS software usually includes functions that detect several types of database errors. These error-checking routines can find mistakes in data topology, including gaps, overshoots, dangling lines and unclosed polygons. An operator sets tolerances that the routine uses to search for errors, and system effectiveness depends on setting correct tolerances. For example, tolerances too small may pass over unintentional gaps, and tolerances too large may improperly remove short dangling lines or small polygons that were intentionally digitized.

Errors during Data Analysis

Even if "accurate," the manipulation and analysis of GIS data can create errors introduced within the data or produced when the data are displayed on screen or plotted in hard copy format.

2.14.5. Sources of Possible Errors

Obvious Sources of Error

1) Age of data. With the exception of geological data the reliability decreases with age.
2) Areal coverage: partial or complete. Many countries still have fragmentary coverage of maps at scales of 1:25000 to 1:50000. Moreover, during the last 30 to 40 years, concepts and definitions of map units, the way they should be mapped have changed.
3) Map scale.
4) Density of observation. How dense should observations be "to support a map".
5) Relevance. Not all data used are directly relevant for the purpose for which they are used. Prime example: remotely sensed data.
6) Accessibility. Not all data are equally accessible (e.g. military secrecy). Sometimes data is not available because of inter-department secrecy.
7) Cost.

Errors Resulting from Natural Variations or from Original Measurement

a) **Positional accuracy.** Topographical data often available with a high degree of positional accuracy. But position of vegetation boundaries etc. often influenced by the subjective judgment of surveyor or by interpretation of remotely sensed data.
b) **Accuracy of content:** qualitative and quantitative. The problem of whether the attributes attached to points, lines or polygons are correct and free from bias. Sometimes systematic errors occur because of instrument. If pixel is too large then it is not clear that it should be classified as forest, road or camp.
c) **Sources of variations in data:** data entry, observer bias, natural variation. Data entry error. Field data very much influenced by surveyor (elevations or census takers).
d) **Errors arising through processing.** Numerical errors in the computer (e.g. joining, matching of a field in GIS software), rounding off errors, truncation etc.

Numerical Errors in Computers

1) Faults arising through topological analyses. Problems associated with map overlay. Digitizing considered infallible. Boundary data is assumed to be sharply defined. All algorithms are assumed to operate in a fully deterministic way.

2) Classification and generalization problems: class intervals, interpolation.

The concepts of complete, compatible, consistent and applicable GIS data previously defined particularly apply to data analysis. Users must consider whether a selected GIS dataset is complete, consistent and applicable for an intended use, and whether it is compatible with other datasets used in the analysis.

The phrasing of spatial and attribute queries also may lead to errors. In addition, the use of Boolean operators can be complicated, and results can be decidedly different, depending on how a data query is structured or a series of queries are executed. For example, the query, "Find all structures within the 100 year flood zone," yields a different result than, "Find all structures touching the 100 year flood zone." The former question will find only those structures entirely within the flood zone, whereas the latter also will include structures that are partially within the zone.

Dataset overlay is a powerful and commonly used GIS tool, but it can yield inaccurate results. To determine areas suitable for a specific type of land development project, one may overlay several data layers, including natural resources, wetlands, flood zones, land uses, land ownership and zoning. The result usually will narrow the possible choices down to a few parcels that would be investigated more carefully to make a final choice. The final result of the analysis will reflect any errors in the original GIS data. Its accuracy only will be as good as the least accurate GIS dataset used in the analysis.

It is also common to overlay and merge GIS data to form new layers. In certain circumstances, this process introduces a new type of error: the polygon "sliver." Slivers often appear when two GIS datasets with common boundary lines are merged. If the common elements have been digitized separately, the usual result will be sliver polygons. Most GIS software products offer routines that can find and fix such errors, but users must be careful in setting search and correction tolerances.

2.14.6. Controlling Errors

GIS data errors are almost inevitable, but their negative effects can be kept to a minimum. Knowing the types and causes of GIS data errors is half the battle; the other half is employing proven techniques for quality control at key stages in the GIS "work flow."

Many errors can be avoided through proper selection and "scrubbing" of source data before they are digitized. Data scrubbing includes organizing, reviewing and preparing the source materials to be digitized. The data should be clean, legible and free of ambiguity. "Owners" of source data should be consulted as needed to clear up questions that arise.

Data entry procedures should be thoroughly planned, organized and managed to produce consistent, repeatable results. Nonetheless, a thorough, disciplined quality review and revision process also is needed to catch and eliminate data entry errors. All production and quality control procedures should be documented, and all personnel should be trained in these procedures. Moreover, the work itself should be documented, including a record of what was done, who did it, when was it done, who checked it, what errors were found and how they were corrected.

To avoid misusing GIS data and the misapplication of analytical software, GIS analysts including casual users need proper training. Moreover, GIS data should not be provided without metadata indicating the source, accuracy and specifics of how the data were entered.

CHAPTER- 3

DATA INPUT AND TOPOLOGY

3.1. INTRODUCTION TO SCANNERS FOR RASTER DATA INPUT

Scanners are used to convert from analog maps or photographs to digital image data in raster format. Digital image data are usually integer-based with one byte gray scale (256 gray tones from 0 to 255) for black and white image and a set of three gray scales of red (R), green (G) and blue(B) for color image. The following four types of scanner are commonly used in GIS and remote sensing.

(a) Mechanical Scanner

It is called drum scanner since a map or an image placed on a drum is digitized mechanically with rotation of the drum and shift of the sensor as shown in Figure 3.1 (a). It is accurate but slow.

(b) Video Camera

Video camera with CRT (cathode ray tube) is often used to digitize a small part of map of firm. This is not very accurate but cheap. (See Figure 3.1 (b).

(c) CCD Camera

Area CCD camera (called digital still camera) instead of video camera will be also convenient to acquire digital image data (see Figure 3.1 (c). It is more stable and accurate than video camera.

(d) CCD Scanner

Flat bed type or roll feed type scanner with linear CCD (charge coupled device) is now commonly used to digitize analog maps in raster format; either in mono-tone or color mode (see Figure **3.1 (d).** It is accurate but expensive.

Comparison of Scanners			
Type	**Resolution**	**Accuracy**	**Cost**
Mechanical Scanner	very high (25~100 μm)	Very high	high
Vedeo Camera	low (500x500 pixels)	low	cheap
CCD Camera	medium (500 x 500 ~ 4,000x4,000)	medium	cheap (low resolution) high (high resolution)
CCD Scanner	very high (300~600 dpi)	high	high

Table.3.1. Shows Comparison of scanners

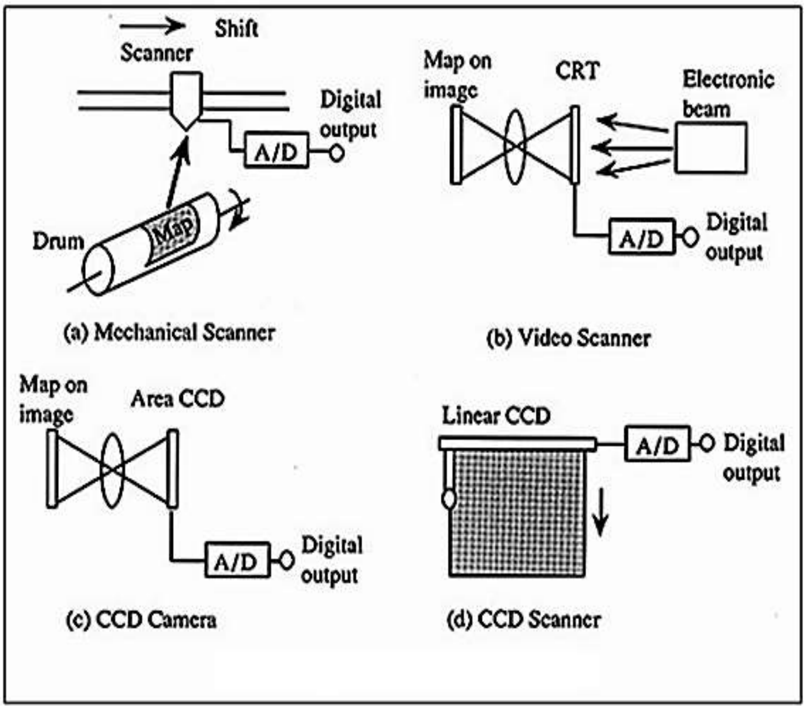


Fig.3.1. Scanners for Raster Data Input

3.1.1. Scanned data

- A scanner is used to convert analog source map or document into digital images by scanning successive lines across a map or document and recording the amount of light reflected from the data source.
- Documents such as building plans, CAD drawings, images and maps are scanned prior to vectorization.
- Scanning helps in reducing wear and tear; improves access and provides integrated storage.
- There are three different types of scanner that are widely used as shown in Figure **3.2**.

3.1.2. Types of Scanner

1) Flat bed scanner
2) Rotating drum scanner
3) Large format feed scanner

(a) Flat bed Scanner (b) Rotating Drum Scanner (c) Large Format Feed Scanner

Fig.3.2. Types of Scanner

Flat bed scanner is a PC peripheral which is small and comparatively inaccurate. The rotating drum scanners are accurate but they tend to be slow and expensive. Large format feed scanner are the most suitable type for inputting GIS data as they are cheap, quick and accurate.

3.2. RASTER DATA INPUT

(A) Introduction

- Need to have tools to transform spatial data of various types into digital format
- data input is a major bottleneck in application of GIS technology
  - costs of input often consume 80% or more of project costs
  - data input is labor intensive, tedious, error-prone
  - there is a danger that construction of the database may become an end in itself and the project may not move on to analysis of the data collected
  - essential to find ways to reduce costs, maximize accuracy
- need to automate the input process as much as possible, but:
  - automated input often creates bigger editing problems later
  - source documents (maps) may often have to be redrafted to meet rigid quality requirements of automated input
- because of the costs involved, much research has gone into devising better input methods - however, few reductions in cost have been realized
- sharing of digital data is one way around the input bottleneck
  - more and more spatial data is becoming available in digital form
- data input to a GIS involves encoding both the locational and attribute data
- the locational data is encoded as coordinates on a particular Cartesian coordinate system

- source maps may have different projections, scales
- several stages of data transformation may be needed to bring all data to a common coordinate system

- attribute data is often obtained and stored in tables

3.2.1. Modes of data input

- Keyboard entry for non-spatial attributes and occasionally locational data
- Manual locating devices
    - user directly manipulates a device whose location is recognized by the computer
    - e.g. digitizing
- Automated devices
    - automatically extract spatial data from maps and photography
    - e.g. scanning
- Conversion directly from other digital sources
- Voice input has been tried, particularly for controlling digitizer operations
    - Not very successful - machine needs to be recalibrated for each operator, after coffee breaks, etc.

3.2.2. Data Input Techniques

Since the input of attribute data is usually quite simple, the discussion of data input techniques will be limited to spatial data only. There is no single method of entering the spatial data into a GIS. Rather, there are several, mutually compatible methods that can be used singly or in combination.

The choice of data input method is governed largely by the application, the available budget, and the type and the complexity of data being input.

There are at least four basic procedures for inputting spatial data into a GIS. These are:

- Manual digitizing;
- Automatic scanning;
- Entry of coordinates using coordinate geometry; and the
- Conversion of existing digital data.

Digitizing

While considerable work has been done with newer technologies, the overwhelming majority of GIS spatial data entry is done by manual digitizing. A digitizer is an electronic device consisting of a table upon which the map or drawing is placed. The user traces the spatial features with a hand-held magnetic pen, often called a mouse or cursor. While tracing the features the coordinates of selected points, e.g. vertices, are sent to the computer and stored. All points that are recorded are registered against positional control points, usually the map corners that are keyed in by the user at the beginning of the digitizing session. The coordinates are recorded in a user defined coordinate system or map projection. Latitude and longitude and UTM is most often used. The ability to adjust or transform data during digitizing from one projection to another is a desirable function of the GIS software. Numerous functional techniques exist to aid the operator in the digitizing process.

Digitizing can be done in a point mode, where single points are recorded one at a time, or in a stream mode, where a point is collected on regular intervals of time or distance, measured by an X and Y movement, e.g. every 3 metres. Digitizing can also be done blindly or with a graphics terminal. Blind digitizing infers that the graphic result is not immediately viewable to the person digitizing. Most systems display the digitized linework as it is being digitized on an accompanying graphics terminal.

Most GIS's use a spaghetti mode of digitizing. This allows the user to simply digitize lines by indicating a start point and an end point. Data can be captured in point or stream mode. However, some systems do allow the user to capture the data in an arc/node topological data structure. The arc/node data structure requires that the digitizer identify nodes.

Data capture in an arc/node approach helps to build a topologic data structure immediately. This lessens the amount of post processing required to clean and build the topological definitions. However, most often digitizing with an arc/node approach does not negate the requirement for editing and cleaning of the digitized line work before a complete topological structure can be obtained.

The building of topology is primarily a post-digitizing process that is commonly executed in batch mode after data has been cleaned. To date, only a few commercial vector GIS software offerings have successfully exhibited the capability to build topology interactively while the user digitizes.

Manual digitizing has many advantages. These include:

- Low capital cost, e.g. digitizing tables are cheap;
- Low cost of labour;
- Flexibility and adaptability to different data types and sources;
- Easily taught in a short amount of time - an easily mastered skill
- Generally the quality of data is high;
- Digitizing devices are very reliable and most often offer a greater precision that the data warrants; and
- Ability to easily register and update existing data.

For raster based GIS software data is still commonly digitized in a vector format and converted to a raster structure after the building of a clean topological structure. The procedure usually differs minimally from vector based software digitizing, other than some raster systems allow the user to define the resolution size of the grid-cell. Conversion to the raster structure may occur on-the-fly or afterwards as a separate conversion process.

Automatic Scanning

A variety of scanning devices exist for the automatic capture of spatial data. While several different technical approaches exist in scanning technology, all have the advantage of being able to capture spatial features from a map at a rapid rate of speed. However, as of yet, scanning has not proven to be a viable alternative for most GIS implementation. Scanners are generally expensive to acquire and operate. As well, most scanning devices have limitations with respect to the capture of selected features, e.g. text and symbol recognition. Experience has shown that most scanned data requires a substantial amount of manual editing to create a clean data layer. Given these basic constraints some other practical limitations of scanners should be identified. These include:

- Hard copy maps are often unable to be removed to where a scanning device is available, e.g. most companies or agencies cannot afford their own scanning device and therefore must send their maps to a private firm for scanning;
- Hard copy data may not be in a form that is viable for effective scanning, e.g. maps are of poor quality, or are in poor condition;
- Geographic features may be too few on a single map to make it practical, cost-justifiable, to scan;
- Often on busy maps a scanner may be unable to distinguish the features to be captured from the surrounding graphic information, e.g. dense contours with labels;
- With raster scanning there it is difficult to read unique labels (text) for a geographic feature effectively; and
- Scanning is much more expensive than manual digitizing, considering all the cost/performance issues.

Consensus within the GIS community indicates that scanners work best when the information on a map is kept very clean, very simple, and uncluttered with graphic symbology.

The sheer cost of scanning usually eliminates the possibility of using scanning methods for data capture in most GIS implementations. Large data capture shops and government agencies are those most likely to be using scanning technology.

Currently, general consensus is that the quality of data captured from scanning devices is not substantial enough to justify the cost of using scanning technology. However, major breakthroughs are being made in the field, with scanning techniques and with capabilities to automatically clean and prepare scanned data for topological encoding. These include a variety of line following and text recognition techniques. Users should be aware that this technology has great potential in the years to come, particularly for larger GIS installations.

Coordinate Geometry

A third technique for the input of spatial data involves the calculation and entry of coordinates using coordinate geometry (COGO) procedures. This involves entering, from survey data, the explicit measurement of features from some known monument. This input technique is obviously very costly and labour intensive. In fact, it is rarely used for natural resource applications in GIS. This method is useful for creating very precise cartographic definitions of property, and accordingly is more appropriate for land records management at the cadastral or municipal scale.

Conversion of Existing Digital Data

A fourth technique that is becoming increasingly popular for data input is the conversion of existing digital data. A variety of spatial data, including digital maps, are openly available from a wide range of government and private sources. The most common digital data to be used in a GIS is data from CAD systems. A number of data conversion programs exist, mostly from GIS software vendors, to transform data from CAD formats to a raster or topological GIS data format. Several ad hoc standards for data exchange have been established in the market place. These are supplemented by a number of government distribution formats that have been developed. Given the wide variety of data formats that exist, most GIS vendors have developed and provide data exchange/conversion software to go from their format to those considered common in the market place.

Most GIS software vendors also provide an ASCII data exchange format specific to their product, and a programming subroutine library that will allow users to write their own data conversion routines to fulfill their own specific needs. As digital data becomes more readily available this capability becomes a necessity for any GIS. Data conversion from existing digital data is not a problem for most technical persons in the GIS field. However, for smaller GIS installations who have limited access to a GIS analyst this can be a major stumbling block in getting a GIS operational. Government agencies are usually a good source for technical information on data conversion requirements.

Some of the data formats common to the GIS marketplace are listed below. Please note that most formats are only utilized for graphic data. Attribute data is usually handled as ASCII text files. Vendor names are supplied where appropriate.

IGDS - Interactive Graphics Design Software (Intergraph / Micro station)	This binary format is a standard in the turnkey CAD market and has become a de facto standard in Canada's mapping industry. It is a proprietary format, however most GIS software vendors provide DGN translators.
DLG - Digital Line Graph (US Geological Survey)	This ASCII format is used by the USGS as a distribution standard and consequently is well utilized in the United States. It is not used very much in Canada even though most software vendors provide two way conversions to DLG.
DXF - Drawing Exchange Format (Auto-cad)	This ASCII format is used primarily to convert to/from the Auto-cad drawing format and is a standard in the engineering discipline. Most GIS software vendors provide a DXF translator.
GENERATE - ARC/INFO Graphic Exchange Format	A generic ASCII format for spatial data used by the ARC/INFO software to accommodate generic spatial data.
EXPORT - ARC/INFO Export Format.	An exchange format that includes both graphic and attribute data. This format is intended for transferring ARC/INFO data from one hardware platform, or site, to another. It is also often used for archiving.

	ARC/INFO data. This is not a published data format, however some GIS and desktop mapping vendors provide translators. EXPORT format can come in either uncompressed, partially compressed, or fully compressed format

A wide variety of other vendor specific data formats exist within the mapping and GIS industry. In particular, most GIS software vendors have their own proprietary formats. However, almost all provide data conversion to/from the above formats. As well, most GIS software vendors will develop data conversion programs dependant on specific requests by customers. Potential purchasers of commercial GIS packages should determine and clearly identify their data conversion needs, prior to purchase, to the software vendor.

3.3 RASTER DATA FILE FORMATS

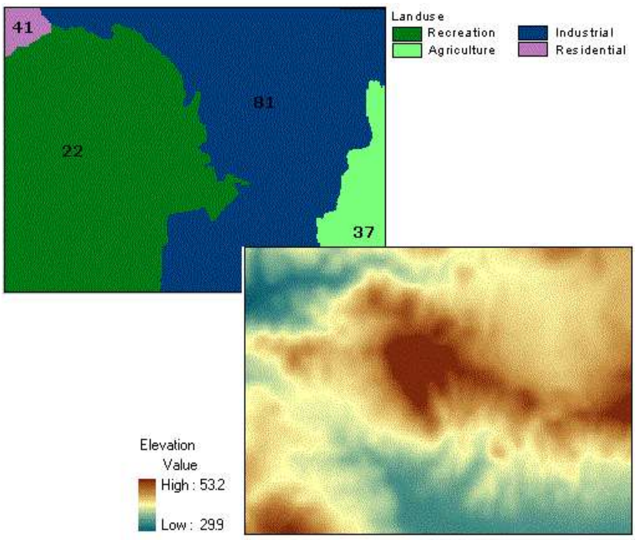
Raster data formats

Two common data formats based on the raster data model are grids and images.

3.3.1. Grids

Grids are an ESRI file format used to store both discrete features such as buildings, roads, and parcels, and continuous phenomena such as elevation, temperature, and precipitation. Recall that the basic unit of the raster data model is the cell. Cells store information about what things are like at a particular location on the earth's surface. Depending on the type of data being stored, cell values can be either integers (whole numbers) or floating points (numbers with decimals). There are two types of grids: one store integers and the other stores floating points.

A discrete grid contains cells whose values are integers, often code numbers for a particular category. Cells can have the same value in a discrete grid. For example, in a discrete grid of land use, each land use type is coded by a different integer, but many cells may have the same code. Discrete grids have an attribute table that stores the cell values and their associated attributes.



Continuous grid is used to represent continuous phenomena; its cell values are floating points. Each cell in a continuous grid can have a different floating point value. For example, in a continuous grid representing elevation, one cell might store an elevation value of 564.3 meters, while the cell to the left might store an elevation value of 565.1 meters. Unlike discrete grids, continuous grids don't have an attribute table.

Discrete grids represent discrete features such as land use categories with integer values. Continuous grids represent continuous phenomena such as elevation with floating point values.

The attribute tables of discrete grids are INFO format, the same format in which coverage feature class attribute tables are stored. As with coverage attribute tables, the INFO table of a discrete grid is stored within an info folder, which is stored at the same level as the grid in a workspace folder. Again like coverages, there is one info folder for all the grids in a workspace folder. To avoid breaking or corrupting the connection between grid files and the info folder, always use ArcCatalog to move, copy, rename, and delete grids.

The Grids workspace folder contains two grids: soils and vegetation. The attribute tables for both grids are stored in the info folder. Auxiliary files called soils.aux and vegetation.aux link the grids and their attribute tables.

3.3.2. Images

The term "image" is a collective term for raster's whose cells, or pixels, store brightness values of reflected visible light or other types of electromagnetic radiation, such as emitted heat (infrared) or ultraviolet (UV). Aerial photos, satellite images, and scanned paper maps are examples of images commonly used in a GIS.

Images can be displayed as layers in a map or they can be used as attributes for vector features. For example, a real estate company might include photos of available houses as an attribute of a home's layer. To be displayed as a layer, however, images must be referenced to real-world locations.

For example, an aerial photo as it comes from the camera is just a static picture, like a picture of a house. There's no information about what part of the world the photo has captured, and the photo may contain distortion and scale variations caused by the angle of the camera. To display properly with other map layers, the aerial photo must be assigned a coordinate system and some of its pixels must be linked to known geographic coordinates.

Raster images, such as aerial photographs and scanned maps, can be referenced to real-world locations, then displayed as a layer in a GIS map.

There are many image file formats, which differ in the type of compression used to reduce the file size. Some of the image formats supported by ArcGIS software.

3.4. VECTOR DATA INPUT

Digitization:

Digitizing is the process of interpreting and converting paper map or image data to vector digital data.

3.4.1. Heads down digitization

Digitizers are used to capture data from hardcopy maps shown in the (fig.**3.3**). Heads down digitization is done on a digitizing table using a magnetic pen known as Puck. The position of a cursor or puck is detected when passed over a table inlaid with a fine mesh of wires. The function of a digitizer is to input correctly the coordinates of the points and the lines. Digitization can be done in two modes:

Fig.3.3. Heads down digitization

- Point mode: In this mode, digitization is started by placing a point that marks the beginning of the feature to be digitized and after that more points are added to trace the particular feature (line or a polygon). The number of points to be added to trace the feature and the space interval between two consecutive points are decided by the operator.
- Stream mode: In stream digitizing, the cursor is placed at the beginning of the feature, a command is then sent to the computer to place the points at either equal or unequal intervals as per the position of the cursor moving over the image of the feature.

3.4.2. Heads-up digitization

This method uses scanned copy of the map or image and digitization is done on the screen of the computer monitor (fig.3.4). The scanned map lays vertical which can be viewed without bending the head down and therefore is called as heads up digitization. Semi-automatic and automatic methods of digitizing requires post processing but saves lot of time and resources compared to manual method and is described in the following section.

Fig.3.4. Screenshot of On-screen/Heads up digitization

3.4.3. Digitizers for Vector Data Input

Tablet digitizers with a free cursor connected with a personal computer are the most common device for digitizing spatial features with the plan metric coordinates from analog maps.

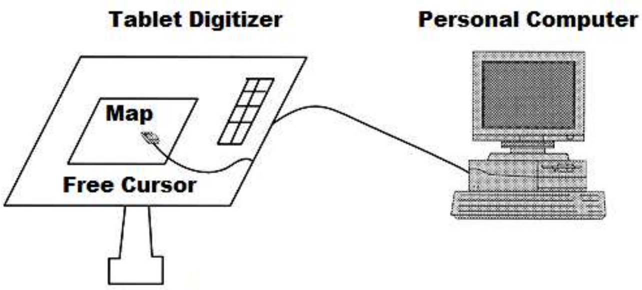


Fig.3.5. Digitizing Tablet

The analog map is placed on the surface of the digitizing tablet as shown in Figure 3.5. The size of digitizer usually ranges from A3 to A0 size.

The digitizing operation is as follows.

Step 1: a map is affixed to a digitizing table.

Step 2: control points or tics at four corners of this map sheet should be digitized by the digitizer and input to PC together with the map coordinates of the four corners.

Step 3: map contents are digitized according to the map layers and map code system in either point mode or stream mode at short time interval.

Step 4: editing errors such as small gaps at line junctions, overshoots, duplicates etc. should be made for a clean dataset without errors.

Step 5: conversion from digitizer coordinates to map coordinates to store in a spatial database.

3.4.4. Major problems of Map Digitization:

- The map will stretch or shrink day by day which makes the newly digitized points
- slightly off from the previous points.
- The map itself has errors discrepancies across neighboring map sheets will produce dis-connectivity.
- Operators will make a lot of errors and mistakes while digitizing as shown in Figure.3.6.

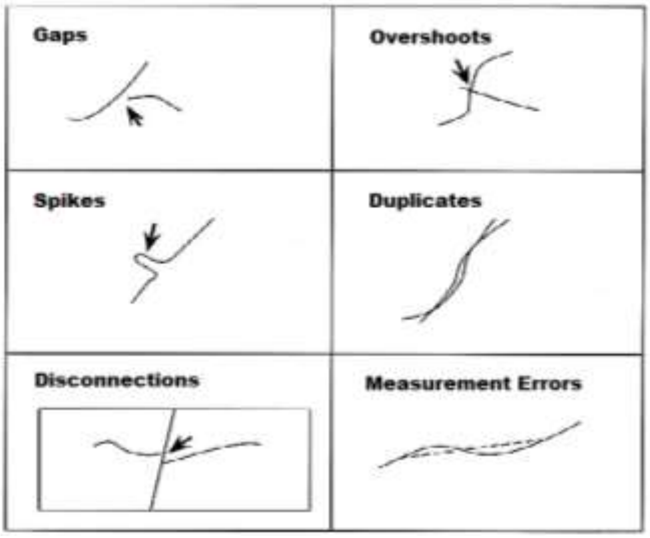


Fig.3.6. Digitizing Typical Errors

3.5. DIGITIZERS

- Digitizers are the most common device for extracting spatial information from maps and photographs
  - the map, photo, or other document is placed on the flat surface of the digitizing tablet

3.5.1. Hardware

- The position of an indicator as it is moved over the surface of the digitizing tablet is detected by the computer and interpreted as pairs of x,y coordinates
  - the indicator may be a pen-like stylus or a cursor (a small flat plate the size of a hockey puck with a cross-hair)
- frequently, there are control buttons on the cursor which permit control of the system without having to turn attention from the digitizing tablet to a computer terminal
- digitizing tablets can be purchased in sizes from 25x25 cm to 200x150 cm, at approximate costs from $500 to $5,000
- early digitizers (ca. 1965) were backlit glass tables

- a magnetic field generated by the cursor was tracked mechanically by an arm located behind the table
- the arm's motion was encoded, coordinates computed and sent to a host processor
- some early low-cost systems had mechanically linked cursors - the free-cursor digitizer was initially much more expensive

- the first solid-state systems used a spark generated by the cursor and detected by linear microphones
  - problems with errors generated by ambient noise
- contemporary tablets use a grid of wires embedded in the tablet to generate a magnetic field which is detected by the cursor
  - accuracies are typically better than 0.1 mm
  - this is better than the accuracy with which the average operator can position the cursor
  - functions for transforming coordinates are sometimes built into the tablet and used to process data before it is sent to the host

The digitizing operation

- The map is affixed to a digitizing table
- Three or more control points ("reference points", "tics", etc.) are digitized for each map sheet
  - these will be easily identified points (intersections of major streets, major peaks, points on coastline)
  - the coordinates of these points will be known in the coordinate system to be used in the final database, e.g. lat/long, State Plane Coordinates, military grid
  - the control points are used by the system to calculate the necessary mathematical transformations to convert all coordinates to the final system
  - the more control points, the better
- digitizing the map contents can be done in two different modes:
  - in point mode, the operator identifies the points to be captured explicitly by pressing a button
  - in stream mode points are captured at set time intervals (typically 10 per second) or on movement of the cursor by a fixed amount
- advantages and disadvantages:
  - in point mode the operator selects points subjectively
    - two point mode operators will not code a line in the same way
  - stream mode generates large numbers of points, many of which may be redundant
  - stream mode is more demanding on the user while point mode requires some judgement about how to represent the line
- most digitizing is currently done in point mode

3.5.2. Problems with digitizing maps

- arise since most maps were not drafted for the purpose of digitizing
  - paper maps are unstable: each time the map is removed from the digitizing table, the reference points must be re-entered when the map is affixed to the table again
  - if the map has stretched or shrunk in the interim, the newly digitized points will be slightly off in their location when compared to previously digitized points
  - errors occur on these maps, and these errors are entered into the GIS database as well
  - the level of error in the GIS database is directly related to the error level of the source maps
- maps are meant to display information, and do not always accurately record locational information
  - for example, when a railroad, stream and road all go through a narrow mountain pass, the pass may actually be depicted wider than its actual size to allow for the three symbols to be drafted in the pass
- discrepancies across map sheet boundaries can cause discrepancies in the total GIS database
  - e.g. roads or streams that do not meet exactly when two map sheets are placed next to each other
- user error causes overshoots, undershoots (gaps) and spikes at intersection of lines
- diagram
- user fatigue and boredom
- for a complete discussion on the manual digitizing process, see Marble et al, 1984

Editing errors from digitizing

- Some errors can be corrected automatically
  - small gaps at line junctions
  - overshoots and sudden spikes in lines
- Error rates depend on the complexity of the map, are high for small scale, complex maps
- These topics are explored in greater detail in later Units
  - Unit 13 looks at the process of editing digitized data
  - Units 45 and 46 discuss digitizing error

3.5.3. Digitizing costs

- A common rule of thumb in the industry is one digitized boundary per minute

- e.g. it would take 99/60 = 1.65 hours to digitize the boundaries of the 99 counties of Iowa

3.6 TOPOLOGY IN GIS

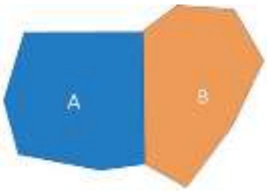
In geo-databases, a topology is a set of rules that defines how point, line, and polygon features share coincident geometry. Topology describes the means whereby lines, borders, and points meet up, intersect, and cross. This includes how street centrelines and census blocks share common geometry, and adjacent soil polygons share their common boundaries. Another example could be how two counties that have a common boundary between them will share an edge, creating a spatial relationship.

Common terms used when referring to topology include: dimensionality, adjacency, connectivity, and containment, with all but dimensional dealing directly with the spatial relationships of features.

Dimensionality - the distinction between point, line, area, and volume, which are said to have topological dimensions of 0, 1, 2, and 3 respectively.

3.6.1. Adjacency

Adjacency including the touching of land parcels, counties, and nation-states (They share a common border).

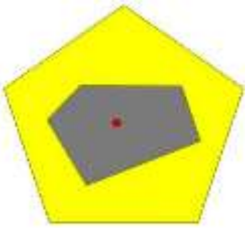


3.6.2. Connectivity

Connectivity including junctions between streets, roads, railroads, and rivers (Very common topological error. See diagrams about "Overshoot" below).

3.6.3. Containment

Containment when a point lies inside rather than outside an area.



Topology defines and enforces data integrity rules (there should be no gaps between polygons). It supports topological relationship queries and navigation (navigating feature adjacency or connectivity), sophisticated editing tools, and allows feature construction from unstructured geometry (constructing polygons from lines).

Addressing topology is more than providing a data storage mechanism. In GIS, topology is maintained by using some of the following aspects:

- The geo-database includes a topological data model using an open storage format for simple features (i.e., feature classes of points, lines, and polygons), topology rules, and topologically integrated coordinates among features with shared geometry. The data model includes the ability to define the integrity rules and topological behaviour of the feature classes that participate in a topology.
- Most GIS programs include a set of tools for query, editing, validation, and error correction of topology.
- GIS software can navigate topological relationships, work with adjacency and connectivity, and assemble features from these elements. It can identify the polygons that share a specific common edge; list the edges that connect at a certain node; navigate along connected edges from the current location; add a new line and "burn" it into the topological graph; split lines at intersections; and create resulting edges, faces, and nodes.

3.6.4. Topological Consistency rules

Topological consistency describes the trustworthiness of the topological and logical relationships between the dataset segments (Joksic and Bajat, 2004). These relations typically involve spatial data inconsistencies such as incorrect line intersections, polygons not properly closed, duplicate lines or boundaries, or gaps in lines. It deals with the structural integrity of a given data set based on a formal framework for modelling of spatial data and relationships among objects. These types of errors must be corrected to avoid incomplete features and to ensure data integrity. Topological errors, which occur during digitizing and data exploration processes, are also known as semantic errors (Ubeda and Egenhofer, 1997). Topological errors exist due to violation of predefined topology rules. The most common topology errors in map data are shown in Figure including: - Duplicate Lines - Overshoots - Undershoots - Micro Segments - Pseudo Nodes - Merge Adjacent Endpoints - Self Intersection.

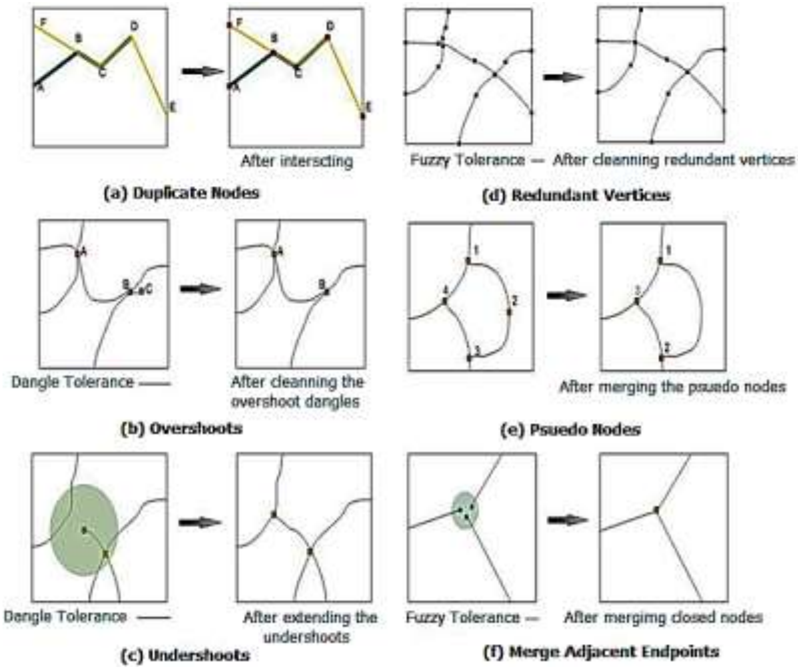


Fig.3.7. Various Topological Errors

3.6.5. Non Topological File Formats

Raster formats

- **ADRG** – National Geospatial-Intelligence Agency (NGA)'s ARC Digitized Raster Graphics
- **Binary file** – An unformatted file consisting of raster data written in one of several data types, where multiple band are stored in BSQ (band sequential), BIP (band interleaved by pixel) or BIL (band interleaved by line). Geo-referencing and other metadata are stored one or more sidecar files.
- **Digital raster graphic** (DRG) – digital scan of a paper USGS topographic map
- **ECRG** – National Geospatial-Intelligence Agency (NGA)'s Enhanced Compressed ARC Raster Graphics (Better resolution than CADRG and no color loss)
- **ECW** – Enhanced Compressed Wavelet (from ERDAS). A compressed wavelet format, often lossy.
- **Esri grid** – proprietary binary and meta-dataless ASCII raster formats used by Esri
- **GeoTIFF** – TIFF variant enriched with GIS relevant metadata
- IMG – ERDAS IMAGINE image file format
- **JPEG2000** – Open-source raster format. A compressed format, allows both lossy and lossless compression.
- **MrSID** – Multi-Resolution Seamless Image Database (by Lizardtech). A compressed wavelet format, allows both lossy and lossless compression.
- **netCDF**-CF – netCDF file format with CF medata conventions for earth science data. Binary storage in open format with optional compression. Allows for direct web-access of subsets/aggregations of maps through OPeNDAP protocol.
- **RPF** – Raster Product Format, military file format specified in MIL-STD-2411
- **CADRG** – Compressed ADRG, developed by NGA, nominal compression of 55:1 over ADRG (type of Raster Product Format)
- **CIB** – Controlled Image Base, developed by NGA (type of Raster Product Format)

Vector formats

- **AutoCAD DXF** – contour elevation plots in AutoCAD DXF format (by Autodesk)
- **Cartesian coordinate system** (XYZ) – simple point cloud
- **Digital line graph** (DLG) – a USGS format for vector data
- **Esri TIN -** proprietary binary format for triangulated irregular network data used by Esri
- **Geography Markup Language** (GML) – XML based open standard (by OpenGIS) for GIS data exchange
- **GeoJSON** – a lightweight format based on JSON, used by many open source GIS packages

- **GeoMedia** – Intergraph's Microsoft Access based format for spatial vector storage
- **ISFC** – Intergraph's MicroStation based CAD solution attaching vector elements to a relational Microsoft Access database
- **Keyhole Markup Language** (KML) – XML based open standard (by OpenGIS) for GIS data exchange
- **MapInfo TAB format** – MapInfo's vector data format using TAB, DAT, ID and MAP files
- **National Transfer Format** (NTF) – National Transfer Format (mostly used by the UK Ordnance Survey)
- **Spatialite** – is a spatial extension to SQLite, providing vector geo-database functionality. It is similar to Post-GIS, Oracle Spatial, and SQL Server with spatial extensions
- **Shapefile** – a popular vector data GIS format, developed by Esri
- **Simple Features** – Open Geospatial Consortium specification for vector data
- **SOSI** – a spatial data format used for all public exchange of spatial data in Norway
- **Spatial Data File** – Autodesk's high-performance geo-database format, native to MapGuide
- **TIGER** – Topologically Integrated Geographic Encoding and Referencing
- **Vector Product Format** (VPF) – National Geospatial-Intelligence Agency (NGA)'s format of vectored data for large geographic databases

Grid formats

- **USGS DEM –** The USGS' Digital Elevation Model
- **GTOPO30** – Large complete Earth elevation model at 30 arc seconds, delivered in the USGS DEM format
- **DTED** – National Geospatial-Intelligence Agency (NGA)'s Digital Terrain Elevation Data, the military standard for elevation data
- **GeoTIFF** – TIFF variant enriched with GIS relevant metadata
- **SDTS** – The USGS' successor to DEM

3.7. LINKING THE ATTRIBUTE DATA TO THE SPATIAL DATA

Before you can use your spatial data as a basis for exploring your attribute data, you must link the attribute data to the spatial data. One way to use the attribute data after you have linked it to the spatial data is by creating a theme to control the appearance of features in the spatial data. See Overview of SAS/GIS Software for more information.

In the layer bar, right-click the COUNTY layer name to open the pop-up menu for the COUNTY layer. Select Edit to open the GIS Layer window. In the definition for the COUNTY layer, select Thematic. The GIS Attribute Data Sets window appears for you to define the link to the theme data set.

In the GIS Attribute Data Sets window, select New to define a new link. In the resulting select a Member window, select MAPS.USAAC. You must next specify the values that are common to both the attribute and spatial data, because the common values provide the connection between the spatial data and the attribute data.

The spatial database and the MAPS.USAAC data set share compatible state and county codes, so first select STATE in both the Data Set Vars and Compositeslists, and then select COUNTY in both lists. Select Save to save the link definition to the Links list. Finally, select Continue to close the GIS Attribute Data Setswindow.

After the GIS Attribute Data Sets window closes, the Var window automatically opens for you. Select which variable in the attribute data provides the theme data for your theme. Select the CHANGE variable to have the counties colored according to the level of change in the county population. Select OK to close the Var window.

The counties in the spatial data are colored according to the demographic values in the attribute data set, as shown in the following display.

Linking the Attribute Data as a Theme

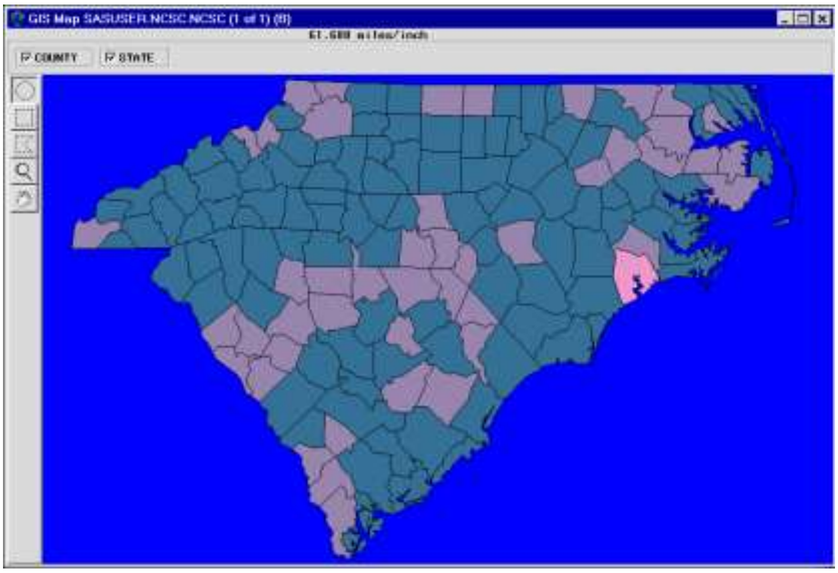


Fig.3.8. Linking the Attribute Data as a Theme

3.7.1. Linking External Databases

The ArcGIS Maps Connect workflow supports external content from Microsoft SQL Server 2008 R2, 2012, 2012 R2, and 2014, including the SQL Server Express editions. The external content must contain data that can be geo-coded, such as an address, U.S. city, U.S. state, ZIP code, or world city. The external content must also contain a primary key column. Alternatively, the table can contain an existing SQL server spatial data type (geography or geometry) column that is then converted by the Arc-GIS Maps Connect workflow for use in ArcGIS Maps for SharePoint.

If the external table has an existing spatial column that contains no data, the ArcGIS Maps Connect workflow populates the column based on other location information in the table (for example, address). If no spatial column exists, the ArcGIS Maps Connect workflow creates a geography spatial type column named EsriShape with a Spatial Reference Identifier (SRID) of 4326 (WGS 84). The EsriShape field supports all geometries including points, lines, and polygons. In all scenarios, the external content can be enriched with additional geographic data variables from ArcGIS.

3.7.2. Note

If the ArcGIS Maps Connect workflow fails, ensure the appropriate permissions for Microsoft SQL Server have been set. You can view the error messages in the SharePoint site workflow history to view exact details on the settings that need to be corrected.

When the ArcGIS Maps Connect workflow completes, the result is a regular SharePoint list, not an external list. That said, the fields created from the SQL Server database are of an external type, and edits made to these fields in SharePoint cannot be passed back to the database. SharePoint can only pass back the fields it has created, such as for the ArcGIS Maps Locate workflow and geoenrichment.

3.8. OPEN DATABASE CONNECTIVITY (ODBC)

In computing, Open Database Connectivity (ODBC) is a standard application programming interface (API) for accessing database management systems (DBMS). The designers of ODBC aimed to make it independent of database systems and operating systems. An application written using ODBC can be ported to other platforms, both on the client and server side, with few changes to the data access code.

ODBC accomplishes DBMS independence by using an ODBC driver as a translation layer between the application and the DBMS. The application uses ODBC functions through an ODBC driver manager with which it is linked, and the driver passes the query to the DBMS. An ODBC driver can be thought of as analogous to a printer driver or other driver, providing a standard set of functions for the application to use, and implementing DBMS-specific functionality. An application that can use ODBC is referred to as "ODBC-compliant". Any ODBC-compliant application can access any DBMS for which a driver is installed. Drivers exist for all major DBMSs, many other data sources like address book systems and Microsoft Excel, and even for text or comma-separated values (CSV) files.

ODBC was originally developed by Microsoft and Simba Technologies during the early 1990s, and became the basis for the Call Level Interface (CLI) standardized by SQL Access Group in the UNIX and mainframe field. ODBC retained several features that were removed as part of the CLI effort. Full ODBC was later ported back to those platforms, and became a de facto standard considerably better known than CLI. The CLI remains similar to ODBC, and applications can be ported from one platform to the other with few changes.

3.8.1. History of Before ODBC

The introduction of the mainframe-based relational database during the 1970s led to a proliferation of data access methods. Generally these systems operated together with a simple command processor that allowed users to type in English-like commands, and receive output. The best-known examples are SQL from IBM and QUEL from the Ingres project. These systems may or may not allow other applications to access the data directly, and those that did use a wide variety of methodologies. The introduction of SQL aimed to solve the problem of language standardization, although substantial differences in implementation remained.

Also, since the SQL language had only rudimentary programming features, users often wanted to use SQL within a program written in another language, say Fortran or C. This led to the concept of Embedded SQL, which allowed SQL code to be embedded within another language. For instance, a SQL statement like SELECT * FROM city could be inserted as text within C source code, and during compiling it would be converted into a custom format that directly called a function within a library that would pass the statement into the SQL system. Results returned from the statements would be interpreted back into C data formats like char * using similar library code.

There were several problems with the Embedded SQL approach. Like the different varieties of SQL, the Embedded SQLs that used them varied widely, not only from platform to platform, but even across languages on one platform – a system that allowed calls into IBM's DB2 would look very different from one that called into their own SQL/DS. Another key problem to the Embedded SQL concept was that the SQL code could only be changed in the program's source code, so that even small changes to the query required considerable programmer effort to modify. The SQL market referred to this as static SQL, versus dynamic SQL which could be changed at any time, like the command-line interfaces that shipped with almost all SQL systems, or a programming interface that left the SQL as plain text until it was called. Dynamic SQL systems became a major focus for SQL vendors during the 1980s.

Older mainframe databases, and the newer microcomputer based systems that were based on them, generally did not have a SQL-like command processor between the user and the database engine. Instead, the data was accessed directly by the program – a programming library in the case of large mainframe systems, or a command line interface or interactive forms system in the case of dBASE and similar applications. Data from dBASE could not generally be accessed directly by other programs running on the machine. Those programs may be given a way to access this data, often through libraries, but it would not work with any other database engine, or even different databases in the same engine. In effect, all such systems were static, which presented considerable problems.

3.8.2. Early efforts

By the mid-1980s the rapid improvement in micro-computers and especially the introduction of the graphical user interface and data-rich application programs like Lotus 1-2-3 led to an increasing interest in using personal computers as the client-side platform of choice in client-server computing. Under this model, large mainframes and minicomputers would be used primarily to serve up data over local area networks to microcomputers that would interpret, display and manipulate that data. For this model to work, a data access standard was a requirement – in the mainframe field it was highly likely that all of the computers in a shop were from one vendor and clients were computer terminals talking directly to them, but in the micro field there was no such standardization and any client might access any server using any networking system.

By the late 1980s there were several efforts underway to provide an abstraction layer for this purpose. Some of these were mainframe related, designed to allow programs running on those machines to translate between the variety of SQL's and provide a single common interface which could then be called by other mainframe or microcomputer programs. These solutions included IBM's Distributed Relational Database Architecture (DRDA) and Apple Computer's Data Access Language. Much more common, however, were systems that ran entirely on microcomputers, including a complete protocol stack that included any required networking or file translation support.

One of the early examples of such a system was Lotus Development's DataLens, initially known as Blueprint. Blueprint, developed for 1-2-3, supported a variety of data sources, including SQL/DS, DB2, FOCUS and a variety of similar mainframe systems, as well as microcomputer systems like dBase and the early Microsoft/Ashton-Tate efforts that would eventually develop into Microsoft SQL Server. Unlike the later ODBC, Blueprint was a purely code-based system, lacking anything approximating a command language like SQL. Instead, programmers used data structures to store the query information, constructing a query by linking many of these structures together. Lotus referred to these compound structures as query trees.

Around the same time, an industry team including members from Sybase (Tom Haggin), Tandem Computers (Jim Gray & Rao Yendluri) and Microsoft (Kyle G) were working on a standardized dynamic SQL concept. Much of the system was based on Sybase's DB-Library system, with the Sybase-specific sections removed and several additions to support other platforms. DB-Library was aided by an industry-wide move from library systems that were tightly linked to a specific language, to library systems that were provided by the operating system and required the languages on that platform to conform to its standards. This meant that a single library could be used with (potentially) any programming language on a given platform.

The first draft of the Microsoft Data Access API was published in April 1989, about the same time as Lotus' announcement of Blueprint. In spite of Blueprint's great lead – it was running when MSDA was still a paper project – Lotus eventually joined the MSDA efforts as it became clear that SQL would become the de facto database standard. After considerable industry input, in the summer of 1989 the standard became SQL Connectivity (SQLC).

3.8.3. SAG and CLI

In 1988 several vendors, mostly from the UNIX and database communities, formed the SQL Access Group (SAG) in an effort to produce a single basic standard for the SQL language. At the first meeting there was considerable debate over whether or not the effort should work solely on the SQL language itself, or attempt a wider standardization which included a dynamic SQL language-embedding system as well, what they called a Call Level Interface (CLI). While attending the meeting with an early draft of what was then still known as MS Data Access, Kyle Geiger of Microsoft invited Jeff Balboni and Larry Barnes of Digital Equipment Corporation (DEC) to join the SQLC meetings as well. SQLC was a potential solution to the call for the CLI, which was being led by DEC.

The new SQLC "gang of four", MS, Tandem, DEC and Sybase, brought an updated version of SQLC to the next SAG meeting in June 1990. The SAG responded by opening the standard effort to any competing design, but of the many proposals, only Oracle Corp had a system that presented serious competition. In the end, SQLC won the votes and became the draft standard, but only after large portions of the API were removed – the standards document was trimmed from 120 pages to 50 during this time. It was also during this period that the name Call Level Interface was formally adopted. In 1995 SQL/CLI became part of the international SQL standard, ISO/IEC 9075-3.[8] The SAG itself was taken over by the X/Open group in 1996, and, over time, became part of The Open Group's Common Application Environment.

MS continued working with the original SQLC standard, retaining many of the advanced features that were removed from the CLI version. These included features like scrollable cursors, and metadata information queries. The commands in the API were split into groups; the Core group was identical to the CLI, the Level 1 extensions were commands that would be easy to implement in drivers, while Level 2 commands contained the more advanced features like cursors. A proposed standard was released in December 1991, and industry input was gathered and worked into the system through 1992, resulting in yet another name change to ODBC.

3.8.4. JET and ODBC

During this time, Microsoft was in the midst of developing their Jet database system. Jet combined three primary subsystems; an ISAM-based database engine (also named Jet, confusingly), a C-based interface allowing applications to access that data, and a selection of driver dynamic-link libraries (DLL) that allowed the same C interface to redirect input and output to other ISAM-based databases, like Paradox and xBase. Jet allowed using one set of calls to access common microcomputer databases in a fashion similar to Blueprint, by then renamed DataLens. However, Jet did not use SQL; like DataLens, the interface was in C and consisted of data structures and function calls.

The SAG standardization efforts presented an opportunity for Microsoft to adapt their Jet system to the new CLI standard. This would not only make Windows a premier platform for CLI development, but also allow users to use SQL to access both Jet and other databases as well. What was missing was the SQL parser that could convert those calls from their text form into the C-interface used in Jet. To solve this, MS partnered with PageAhead Software to use their existing query processor, SIMBA. SIMBA was used as a parser above Jet's C library, turning Jet into an SQL database. And because Jet could forward those C-based calls to other databases, this also allowed SIMBA to query other systems. Microsoft included drivers for Excel to turn its spreadsheet documents into SQL-accessible database tables.

3.8.5. Release and continued development

ODBC 1.0 was released in September 1992. At the time, there was little direct support for SQL databases (versus ISAM), and early drivers were noted for poor performance. Some of this was unavoidable due to the path that the calls took through the Jet-based stack; ODBC calls to SQL databases were first converted from Simba Technologies's SQL dialect to Jet's internal C-based format, then passed to a driver for conversion back into SQL calls for the database. Digital Equipment and Oracle both contracted Simba Technologies to develop drivers for their databases as well.

3.24 GIS

Circa 1993, OpenLink Software shipped one of the first independently developed third-party ODBC drivers, for the PROGRESS DBMS, and soon followed with their UDBC (a cross-platform API equivalent of ODBC and the SAG/CLI) SDK and associated drivers for PROGRESS, Sybase, Oracle, and other DBMS, for use on Unix-like OS (AIX, HP-UX, Solaris, Linux, etc.), VMS, Windows NT, OS/2, and other OS.

Meanwhile, the CLI standard effort dragged on, and it was not until March 1995 that the definitive version was finalized. By then, Microsoft had already granted Visigenic Software a source code license to develop ODBC on non-Windows platforms. Visigenic ported ODBC to a wide variety of Unix platforms, where ODBC quickly became the de facto standard. "Real" CLI is rare today. The two systems remain similar, and many applications can be ported from ODBC to CLI with few or no changes.

Over time, database vendors took over the driver interfaces and provided direct links to their products. Skipping the intermediate conversions to and from Jet or similar wrappers often resulted in higher performance. However, by then Microsoft had changed focus to their OLE DB concept (recently reinstated), which provided direct access to a wider variety of data sources from address books to text files. Several new systems followed which further turned their attention from ODBC, including ActiveX Data Objects (ADO) and ADO.net, which interacted more or less with ODBC over their lifetimes.

As Microsoft turned its attention away from working directly on ODBC, the UNIX field was increasingly embracing it. This was propelled by two changes within the market, the introduction of graphical user interfaces (GUIs) like GNOME that provided a need to access these sources in non-text form, and the emergence of open software database systems like PostgreSQL and MySQL, initially under Unix. The later adoption of ODBC by Apple for using the standard Unix-side iODBC package Mac OS X 10.2 (Jaguar) (which OpenLink Software had been independently providing for Mac OS X 10.0 and even Mac OS 9 since 2001) further cemented ODBC as the standard for cross-platform data access.

Sun Microsystems used the ODBC system as the basis for their own open standard, Java Database Connectivity (JDBC). In most ways, JDBC can be considered a version of ODBC for the programming language Java instead of C. JDBC-to-ODBC bridges allow Java-based programs to access data sources through ODBC drivers on platforms lacking a native JDBC driver, although these are now relatively rare. Inversely, ODBC-to-JDBC bridges allow C-based programs to access data sources through JDBC drivers on platforms or from databases lacking suitable ODBC drivers.

3.8.6. ODBC today

ODBC remains in wide use today, with drivers available for most platforms and most databases. It is not uncommon to find ODBC drivers for database engines that are meant to be embedded, like SQLite, as a way to allow existing tools to act as front-ends to these engines for testing and debugging.

However, the rise of thin client computing using HTML as an intermediate format has reduced the need for ODBC. Many web development platforms contain direct links to target databases – MySQL being very common. In these scenarios, there is no direct client-side access nor multiple client software systems to support; everything goes through the programmer-supplied HTML application. The virtualization that ODBC offers is no longer a strong requirement, and development of ODBC is no longer as active as it once was.[citation needed]

3.9. GPS

Stands for "Global Positioning System." GPS is a satellite navigation system used to determine the ground position of an object. GPS technology was first used by the United States military in the 1960s and expanded into civilian use over the next few decades. Today, GPS receivers are included in many commercial products, such as automobiles, Smartphone, exercise watches, and GIS devices.

The GPS system includes 24 satellites deployed in space about 12,000 miles (19,300 kilometers) above the earth's surface. They orbit the earth once every 12 hours at an extremely fast pace of roughly 7,000 miles per hour (11,200 kilometers per hour). The satellites are evenly spread out so that four satellites are accessible via direct line-of-sight from anywhere on the globe.

Each GPS satellite broadcasts a message that includes the satellite's current position, orbit, and exact time. A GPS receiver combines the broadcasts from multiple satellites to calculate its exact position using a process called triangulation. Three satellites are required in order to determine a receiver's location, though a connection to four satellites is ideal since it provides greater accuracy.

In order for a GPS device to work correctly, it must first establish a connection to the required number of satellites. This process can take anywhere from a few seconds to a few minutes, depending on the strength of the receiver. For example, a car's GPS unit will typically establish a GPS connection faster than the receiver in a watch or Smartphone. Most GPS devices also use some type of location caching to speed up GPS detection. By memorizing its previous location, a GPS device can quickly determine what satellites will be available the next time it scans for a GPS signal.

3.9.1. Uses of GPS

GPS has many uses, for example;

- **Clock synchronization:** The GPS time signals use highly accurate atomic clocks. This technology can be used for things like automatic updates of daylight saving times on cell phones
- Disaster relief and emergency services: Depend upon GPS for location
- Tracking a vehicle, person, pet or aircraft: Receivers provide continuous tracking and can provide an alert if the receiver leaves a set area. Pets can be chipped so they can be found if they become lost
- **Geotagging:** Applying location coordinates to digital objects such as photographs and other documents for purposes such as creating map overlays.
- **Bus tour commentary:** your location will determine what information is displayed about approaching points of interest
- **Bus stops:** to show how long the bus will take to arrive at a bus stop

- **Navigation:** eg Navman. The device uses voice activation to describe a preferred route based on the position of the receiver, the position of the destination and a street map
- **Personal Locator Beacons (PLB):** used to inform search and rescue authorities of your exact location in the event of an emergency
- **Recreation:** For example, geo-caching and way-marking
- **Surveying:** Surveyors use absolute locations to make maps and determine property boundaries
- **Tectonics:** enables fault motion measurement in earthquakes

Maps have come a long way since people first began drawings to show where they were. Modern maps are created using special software that combines lots of different sorts of information. This system of modern mapping is called GIS – Geographic Information Systems. GIS is used by organizations, such as city councils, that need access to data and need to be able to combine different data sets together. GIS gives people in these organisations graphical representations of data that allows them to:

- Analyze situations
- Write reports
- Track changes
- Make decisions
- Plan for the future, for example which parts of the high country have undergone tenure review

GIS requires four things:

People: people who use GIS are professionals who have been educated to use GIS and have made a career out of working with GIS

Data: geospatial information (where things are located) and the details of objects such as services, roads, buildings etc. are collected and entered into the GIS software

Software: GIS software analyses data and presents it in different combinations for the user

Hardware: includes hand held devices for collecting data and computers with GIS software

3.9.2. Basic structure of GPS

Three-block configuration

GPS consists of the following three segments.

Space segment (GPS satellites)

A number of GPS satellites are deployed on six orbits around the earth at the altitude of approximately 20,000 km (four GPS satellites per one orbit), and move around the earth at 12-hour-intervals.

Control segment (Ground control stations)

Ground control stations play roles of monitoring, controlling and maintaining satellite orbit to make sure that the deviation of the satellites from the orbit as well as GPS timing are within the tolerance level.

User segment (GPS receivers)

User segment (GPS receivers)

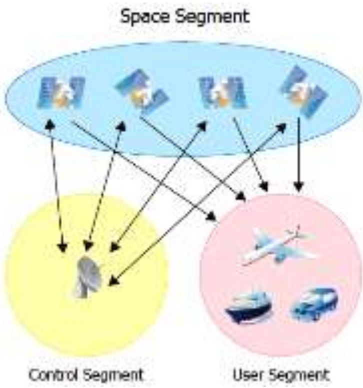


Fig.3.9. Three elements of GPS

3.9.3. GPS positioning

Firstly, the signal of time is sent from a GPS satellite at a given point. Subsequently, the time difference between GPS time and the point of time clock which GPS receiver receives the time signal will be calculated to generate the distance from the receiver to the satellite. The same process will be done with three other available satellites. It is possible to calculate the position of the GPS receiver from distance from the GPS receiver to three satellites. However, the position generated by means of this method is not accurate, for there is an error in calculated distance between satellites and a GPS receiver, which arises from a time error on the clock incorporated into a GPS receiver. For a satellite, an atomic clock is incorporated to generate on-the-spot time information, but the time generated by clocks incorporated into GPS receivers is not as precise as the time generated by atomic clocks on satellites. Here, the fourth satellite comes to play its role: the distance from the fourth satellite to the receiver can be used to compute the position in relations to the position data generated by distance between three satellites and the receiver, hence reducing the margin of error in position accuracy.

The Fig--- below illustrates an example of positioning by two dimensions (position acquisition by using two given points). We can compute where we are at by calculating distance from two given points, and the GPS is the system that can be illustrated by multiplying given points and replacing them with GPS satellites on this figure.

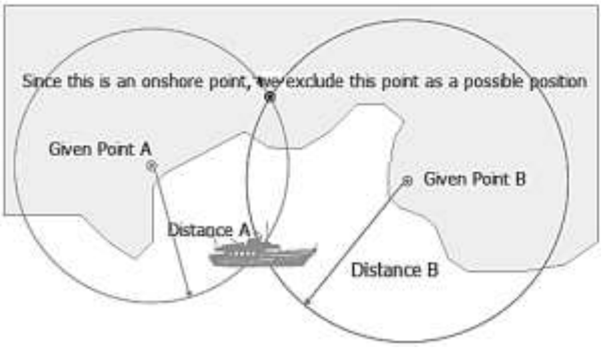


Fig.3.10. position acquisition by using two given points

3.9.4. GPS signals

GPS satellites broadcast beams in two carrier frequencies; L1 (1,575.42 MHz) and L2 (1,227.60 MHz). Beams that can be accessible to the general public are encoded in C/A (Coarse/Acquisition) code, and the beams that can be used only by the US military force are encoded in P (Precise) code. C/A code consists of identification codes of each satellite and is broadcast together with navigation messages. The data of the orbit of each satellite is called the ephemeris*, and the data of orbit of all satellite is called the almanac**. The navigation messages are broadcast at a rate of 50 bits per second. Utilizing this collection of data, GPS receiver calculates distance between satellites and the receiver in order to generate position data. In the Fig 1-4, the details of C/A code is described, and in the Fig 1-5, navigation messages are described.

**The ephemeris provides the precise orbit for the satellite itself, which can be used to generate precise location of the satellite, necessary information for calculating position information. It is the indigenous data that is used only by each of the GPS satellites with specific identification number.*

***The almanac can be regarded as simplified ephemeris data and contains coarse orbit and status information for all satellites in the network. It is used to locate available satellites in order a GPS receiver to generate current position and time. It takes 12.5 minutes to receive all the almanac data.*

3.9.5. C/A code:

L1 signal from the GPS satellites is phase-modulated in C/A code, which is the pseudorandom code. The pseudorandom code is also called pseudorandom noise code, which is known as a Gold code. As the Fig. 1-4 illustrates, C/A code is a sequence of digital signals "1" and "0". In GPS, 1,023 consecutive patterns comprise a sequence, and subsequently, this sequence will continually repeat one after another.

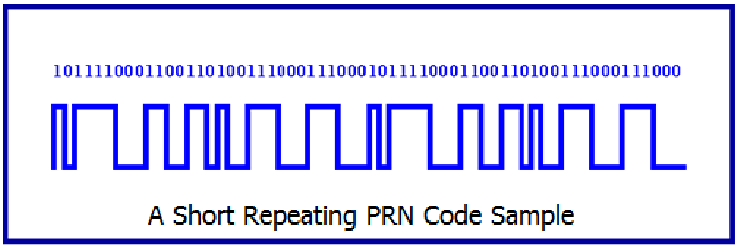


Fig.3.11. C/A code used identify the GPS

3.9.6. Navigation message

Navigation message consists of 25 frames, each of which includes 5 sub-frames of 300 bits each. The data length of 1 bit is 20 ms, and thus, the length of each sub-frame is 6 seconds, and each frame is a grouping of 1,500 bits of information with the frame length of 30 seconds. Since navigation message consists of 25 frames, this would add up to the message length of 12.5 minutes (30 seconds x 25=12.5 minutes). The GPS receiver requires 12.5 minutes to receive all the necessary set of data, necessary condition for positioning, when initial power activation takes place. The GPS receiver is capable of storing this set of data gained in the past internal backup battery, and it reads out the set of data when power reactivation takes place, hence instantaneously starting to receive GPS position.

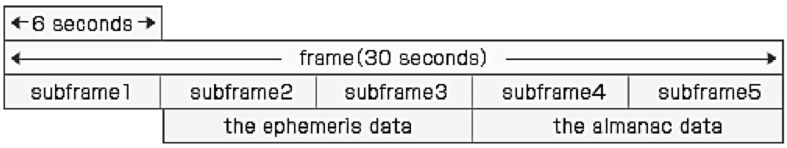


Fig.3.12. Navigation message

Positioning accuracy

Factors that trigger GPS position errors

Ionosphere

The ionosphere is a portion of the upper atmosphere, between the thermosphere and the exosphere. When GPS signals pass through this layer, the propagation velocity of the GPS signal goes slower, hence causing propagation error.

Troposphere

The troposphere is the lowest portion of Earth's atmosphere. Radio reflections caused by dry atmosphere and water vapor within provoke GPS position error.

Multipath propagation

GPS signal is not immune to reflection when it hits on the ground, structures and many others. This phenomenon is called multipath propagation, one of the causes of GPS position errors.

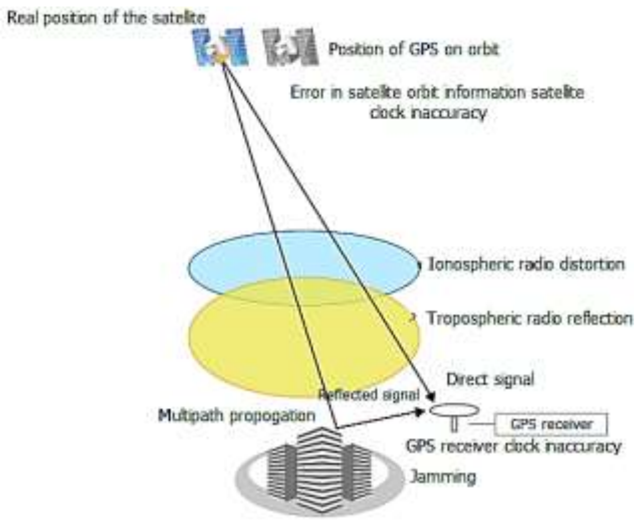


Fig.3.13. GPS position error

3.9.7. DOP (Dilution of Precision)

DOP is a value that shows the degree of degradation of the GPS positioning accuracy. The smaller the value is, the higher the positioning accuracy is. This value depends upon the positions of the GPS satellites tracked for positioning. If the tracked satellites spread evenly over the earth, the positioning accuracy would become higher, and if the positions of tracked satellites are disproportionate, the positioning accuracy would become lower.

Fig.3.14. DOP value is smaller *Fig.3.15. DOP value greater*

3.9.8. Signal strength

State of reception of GPS depends upon the strength of GPS signals. The greater the signal strength is, the more stable the reception status is. Whereas the reception status would become unstable when the GPS signal became weaker, due to obstacles or noise sources in the vicinity of a GPS receiver.

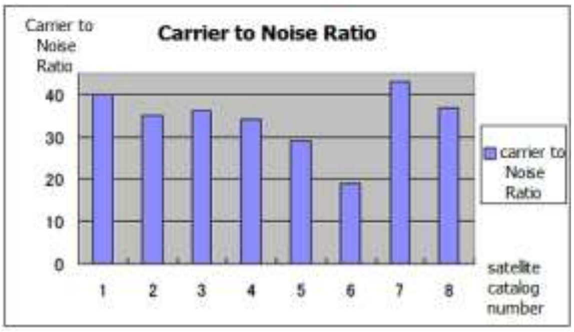


Fig.3.16. Example of stable GPS positioning

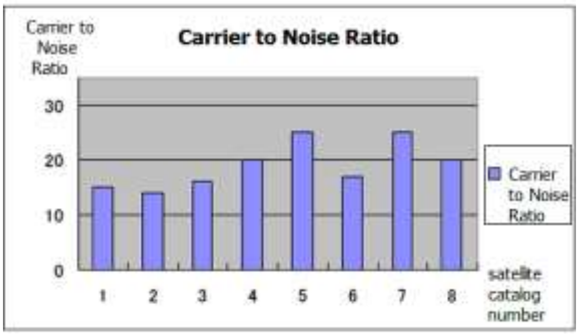


Fig.3.17. Example of unstable GPS positioning

3.9.9. Number of satellites tracked for positioning

State of reception of GPS depends upon the number of satellites tracked for positioning.

If the number of the tracked satellites is great, GPS positioning becomes greater, but if there were a fewer satellites tracked for positioning, it would be difficult to generate GPS position. The Fig. 1-11 illustrates the occasion where the GPS receiver tracks a greater number of satellites for positioning. The Fig. 1-12 illustrates the occasion where the GPS receiver tracks only a few number of satellites for positioning.

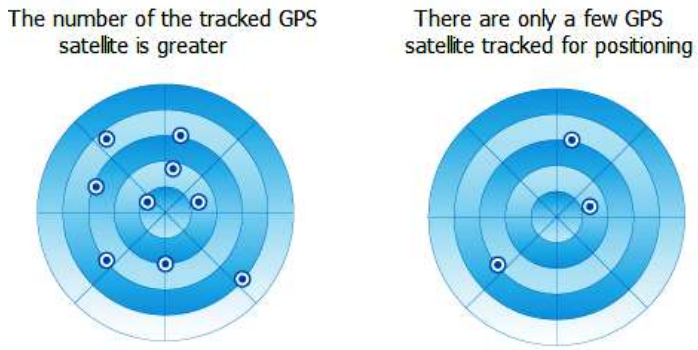


Fig.3.18. GPS satellite are portrayed as blue circles above

3.10. CONCEPT OF GPS BASED MAPPING

GPS consists of a constellation of radio navigation satellite and a ground control segment. It manages satellite operation and users with specialized receivers who use the satellite data to satisfy a broad range of positioning requirements.

In brief, following are the key features of GPS:-

- The basis of GPS is „triangulation‟ more precisely trilateration from satellites
- A GPS receiver measures distance using the travel time of radio signals.
- To measure travel time GPS needs very accurate timing that is achieved with some techniques.
- Along with distance, one needs to know exactly where the satellites are in space.
- Finally one must correct for any delays, the signal experience as it travels through the atmosphere.

The whole idea behind GPS is to use satellites in space as reference points for location here on earth. By very accurately measuring the distances from at least three satellites, we can „triangulate‟ our position anywhere on the earth by resection method.

3.10.1. GPS Elements

GPS has 3 parts: the space segment, the user segment, and the control segment, Figure-1.2 illustrates the same. The space segment consists of a constellation of 24 satellites, each in its own orbit, which is 11,000 nautical miles above the Earth. The user segment consists of receivers, which can be held in hand or mount in the vehicle. The control segment consists of ground stations (six of them, located around the world) that make sure the satellites are working properly. More details on each of these elements can be referred from any standard book or online literature on GPS.

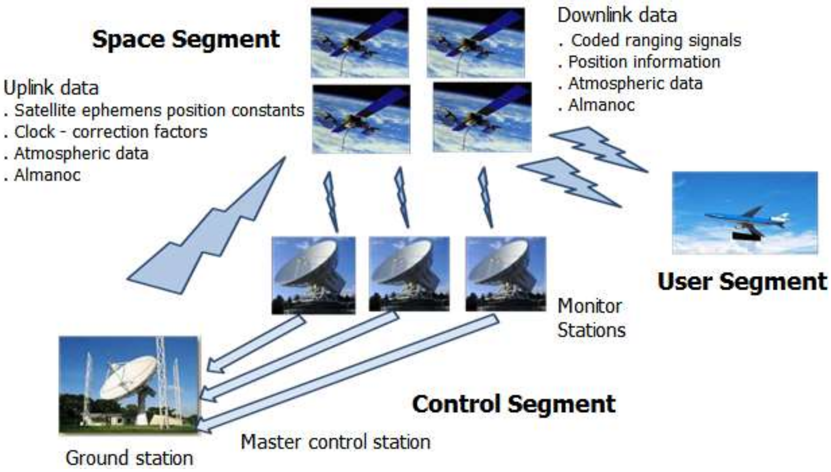


Fig.3.19. GPS segments

3.10.2. GPS Satellite Navigation System

GPS is funded and controlled by the U. S. Department of Defense (DOD). While there are many thousands of civil users of GPS worldwide, the system was designed for and is operated by the U. S. military. It provides specially coded satellite signals that can be processed in a GPS receiver, enabling the receiver to compute position, velocity and time. Four GPS satellite signals are used to compute positions in three dimensions and the time offset in the receiver clock.

3.10.3. GPS Positioning Techniques

GPS positioning techniques may be categorized as being predominantly based on code or carrier measurements. Code techniques are generally simple and produce low accuracy, while carrier techniques are more complex and produce higher accuracy. There exist a variety of positioning methods for both code and carrier measurements. The suitability of each for a specific application is dependent on the desired accuracy, logistical constraints and costs. Many variables affect this accuracy, such as the baseline lengths, ionospheric conditions, magnitude of selective availability, receiver types used, and processing strategies adopted.

3.10.4. Differential GPS (DGPS)

The technique used to augment GPS is known as “differential”. The basic idea is to locate one or more reference GPS receivers at known locations in users‟ vicinities and calibrate ranging errors as they occur. These errors are transmitted to the users in near real time. The errors are highly correlated across tens of kilometers and across many minutes. Use of such corrections can greatly improve the accuracy and integrity.

To increase the accuracy of positioning, Differential-GPS (D-GPS) was introduced. The idea is as follows: a reference station is located at a known and accurately surveyed point. The

GPS reference station determines its GPS position using four or more satellites. Given that the position of the GPS reference station is exactly known, the deviation of the measured position to the actual position and more importantly the measured pseudo range to each of the individual satellites can be calculated. The differences are either transmitted immediately by radio or used afterwards for correction after carrying out the measurements. The man made error like "Selective Availability" can be corrected using this.

3.10.5. GPS applications in Transportation

Due to the high accuracy, usability, ease and economy of operations in all weather, offered by GPS, it has found numerous applications in many fields ranging from accuracy level of mm for the high precision geodesy to several meters for navigational positioning. Some of the applications in urban and transportation field are:

- Establishment of ground control points for imageries / map registration,
- Determination of a precise geo ID using GPS data,
- Survey control for topographical and cadastral surveys,
- Air, road, rail, and marine navigation,
- Intelligent traffic management system,
- Vehicle tracking system etc.

CHAPTER- 4

DATA QUALITY AND STANDARDS

4.1. VECTOR DATA ANALYSIS TOOLS

A systematic examination of a problem or complex entity in order to provide new information from what is already known (ESRI – GIS Dictionary)

Spatial analysis is the process that turns raw input spatial data into value-added useful data or information that ultimately supports decision making and reveals hidden or unknown patterns

Common vector analysis tools are:

- Buffering
- Overlay
- Other feature manipulation tools

Buffering

- Buffer – a region that is less than or equal to a specified distance
- Can buffer points, lines and polygons used to examine proximity constraints?

Example:

- Identify potential customers within 3km of store
- Identify parks within 10km of Islamabad Highway
- Identify schools within 5 km of industrial zone

4.1.1. Buffer Analysis:

- A buffer is a zone with a width created around a spatial feature and is measured in units of distance from the feature. The generated buffer takes the shape of the feature.
- In case of a point the buffer is a circle (refer Figure 4.1 (a), with a radius equal to the buffer distance. In case of a line (refer Figure 4.1 (b), it is a band and for a polygon it is a belt of a specific buffer distance from the edge of polygon, surrounding the polygon.
- The inward buffer for a polygon is called setback (refer Figure 4.1 (c), the polygon on the right hand side).

- Buffering is used for neighborhood analysis which aims to evaluate the characteristics of the area surrounding the spatial feature.

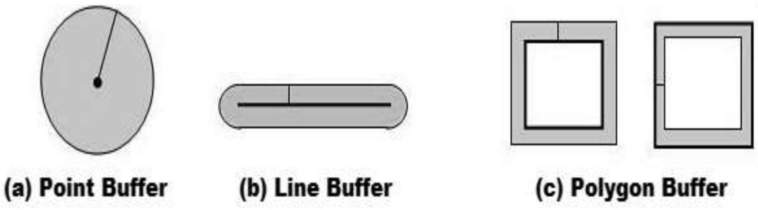


Figure.4.1. Buffer Analysis Diagram

- Common examples of buffering include the identification of properties within a certain distance of an object, delineation of areas around natural features where human activities are restricted, determination of areas affected by location etc.
- **Clip** is used to subset a point, line or a polygon theme using another polygon theme as the boundary of the area of interest.
- In the illustration above, the input, point feature shows the location of drinking water wells in three villages as shown in Figure 4.2.
- To know how many wells fall in village 1, the input feature class is clipped using the boundary of the village 1. The output feature class shows that five wells are present in village 1(One).

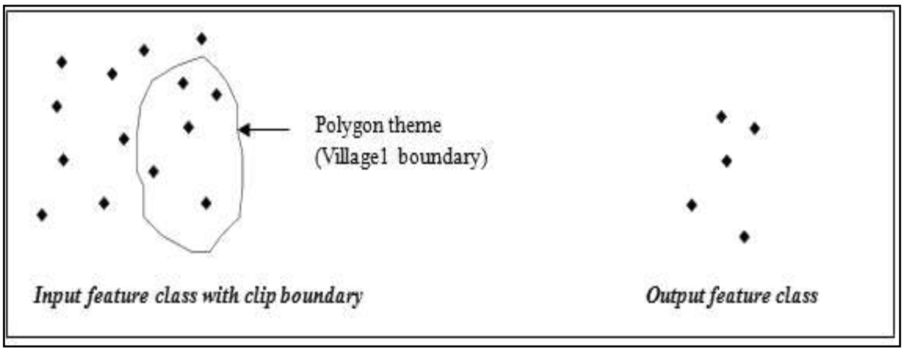


Figure.4.2. Drinking Water Wells in Three Villages

- **Split** causes the input features to form subset of multiple output feature classes. The split field's unique values form the names of the output feature classes.

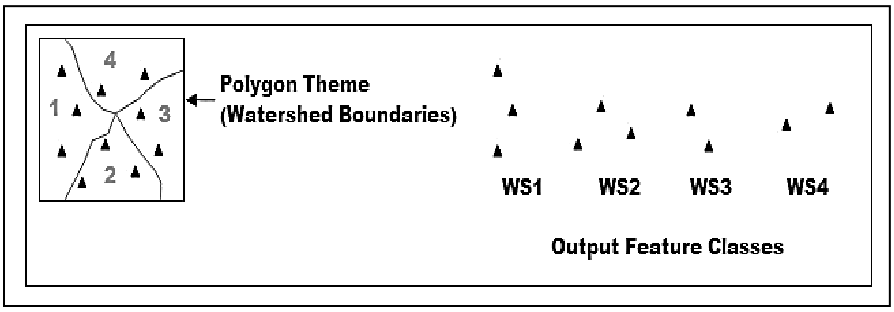


Fig.4.3. Polygon theme of Watershed Boundaries

- In the illustration above, a point theme of wells is split using the polygon theme of watershed boundaries as shown in Figure 4.3.
- The output of this operation contains multiple feature classes which are named on the unique value of watershed boundaries (in this case, the unique value is the watershed number WS1, WS2 etc.).
- Each output class represents the number of wells present in a particular watershed i.e. WS1 or watershed 1 has three wells. Similarly, WS2, WS3 and WS4 have 3, 2, and 2 wells respectively.

4.1.2. Overlay Analysis

- **Union** creates a new theme by overlaying two polygon themes. It is same as 'or' Boolean operator. The output theme contains the combined polygons and attributes of both themes. Only polygon themes can be combined using union as shown in Figure **4.4.**

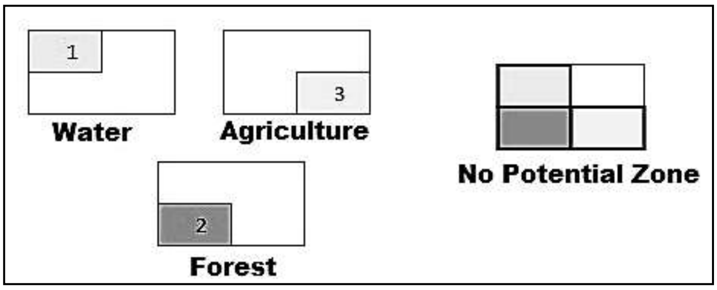


Fig.4.4. Overlay Analysis – Union

- Let's say we are interested in knowing no potential zone for urban development. It is clear that no construction can be done on a water body or land covered by agriculture or forest. So, we can say union of areas under water, agriculture and forest would present us the area having no potential for urban development.
- **Intersect** creates a new theme by overlaying a point, line or polygon theme with an intersecting polygon theme. It is same as 'and' Boolean operator. The output theme contains only the feature inside the intersecting polygons.

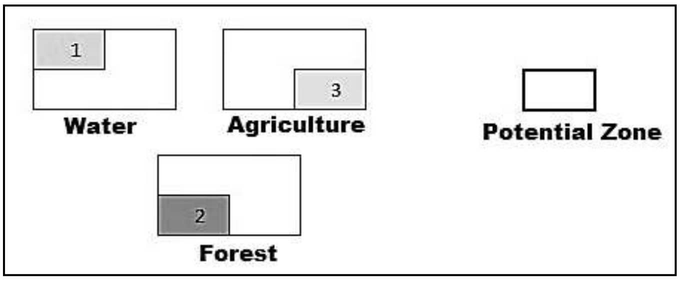


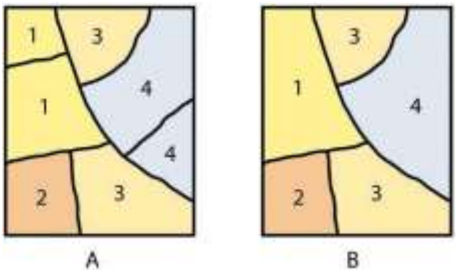
Fig.4.5. Overlay Analysis – Intersect

- From the same example given above, if we try to know the area having potential for
- Urban development we need to intersect the polygon themes to get a common area which is not under water, agriculture or forest as shown in Figure 4.5.

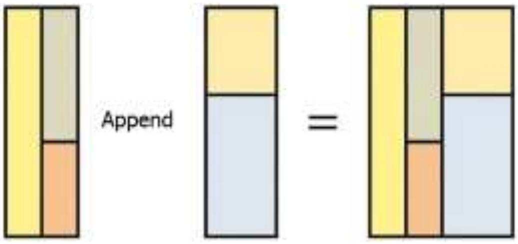
- From the same example given above, if we try to know the area having potential for urban development we need to intersect the polygon themes to get a common area which is not under water, agriculture or forest.

4.1.3. Feature Manipulation

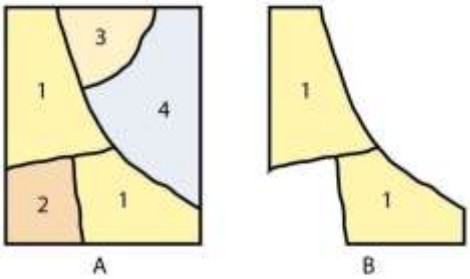
Dissolve



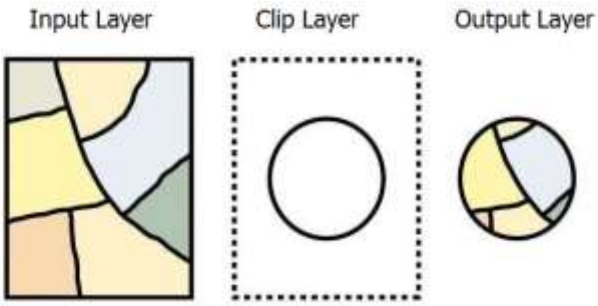
Append



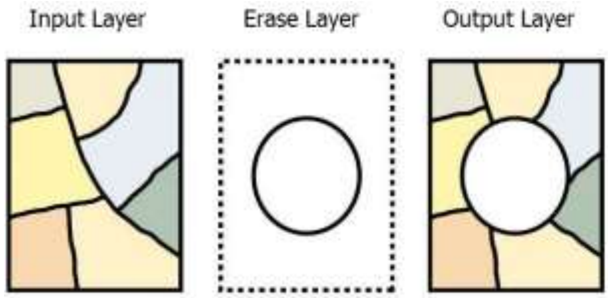
Select



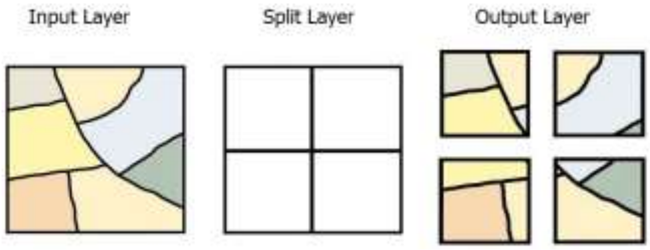
Clip



Erase



Split



4.2. DATA ANALYSIS TOOLS

- The Analysis toolbox contains a powerful set of tools that perform the most fundamental GIS operations. With the tools in this toolbox, you can perform overlays, create buffers, calculate statistics, perform proximity analysis, and much more. Whenever you need to solve a spatial or statistical problem, you should always look in the Analysis toolbox.
- The Analysis toolbox has four toolsets. Each toolset performs specific GIS analysis of feature data.

Toolsets	Description
Extract	GIS datasets often contain more data than you need. The Extract tools let you select features and attributes in a feature class or table based on a query (SQL expression) or spatial extraction. The output features and attributes are stored in a feature class or table.
Overlay	The Overlay toolset contains tools to overlay multiple feature classes to combine, erase, modify, or update spatial features, resulting in a new feature class. New information is created when overlaying one set of features with another. There are six types of overlay operations; all involve joining two existing sets of features into a single set of features to identify spatial relationships between the input features.
Proximity	The Proximity toolset contains tools that are used to determine the proximity of features within one or more feature classes or between two feature classes. These tools can identify features that are closest to one another or calculate the distances between or around them.

4.6 GIS

Statistics	The Statistics toolset contains tools that perform standard statistical analysis (such as mean, minimum, maximum, and standard deviation) on attribute data as well as tools that calculate area, length, and count statistics for overlapping and neighboring features.

4.3. NETWORK ANALYSIS

It is a type of line analysis which involves set of interconnected lines. Railways, highways, transportation routes, rivers etc are examples of networks. Network analysis is used to find the shortest alternated routes between origins to destination Network Analyst provides network based spatial analysis tools for solving complex routing problems.

Modeling in GIS Highway alignment studies:

4.3.1. Highway Alignment

The position or the layout of the centre line of the highway on the ground is called the alignment. The Horizontal Alignment includes the straight path, the horizontal deviations and curves. Changes in gradient curves are covered under vertical alignment of roads.

A new road should be aligned very carefully as improper alignment would result in one or more of the following disadvantages:

1) Increase in construction cost
2) Increase in maintenance cost
3) Increase in vehicles operation cost
4) Increase in accident rate.

The basic requirements of an ideal alignment between two terminal stations are that it should be:

1) Short 2) Easy 3) Safe 4) Economical

4.3.2. Factor affecting Highway Alignment:

The various factors which control the highway alignment in general may be listed as:

1) Obligatory points
2) Traffic
3) Geometric design
4) Economics
5) Other considerations In hill roads additional care has to be given for: Stability-
6) Drainage

4.3.3. Stages of New Highway Project:

1) Selection of route, finalization of highway alignment and geometric design details.
2) Collection of materials and testing of subgrade soil and other construction material, mix design of pavement materials and design details of pavement layer.
3) Construction stages including quality control.

4.3.4. Steps Involved in a New Highway Project:

1) Map study
2) Reconnaissance Survey
3) Preliminary survey
4) Location of Final Alignment
5) Detailed survey
6) Material survey
7) Design
8) Earth work
9) Pavement Construction
10) Construction

4.3.5. Need of Study

The conventional method of highway alignment is a tedious and time consuming process

- The conventional highway alignment needs a lot of manual work and expensive
- Remote sensing and Geographical information system makes the highway alignment easier. It needs less man power,
- less time consuming and economic.

4.3.6. Objectives of study

The objectives of the present study are as follows,

- To identify the factors that influence on highway alignment studies
- To prepare the thematic layers based on the identified factors
- To analyses the traffic volume and future expansion.
- To identify the favorable route for highway alignment.

4.3.7. Methodology

- The base (study area) map, Drainage, Slope and Contour maps were prepared with the help of SOI Topo-sheet (on 1:50,000 scale).

4.8 GIS

- IRS LISS III satellite data was used and by using Digital Image Processing techniques the following thematic maps such as geomorphology, Land use/ Land Cover were generated as shown in the figure.5.22.
- The DEM is used in order to understand the terrain condition, environmental factors and social economic status in this study area.
- The factors considered are mainly related to the land use, geology, land value and soil. The weights and ranks are assigned to each of the above themes, according to expert opinions, for GIS analysis. After assigning weights and ranks these themes are overlaid to get an overlaid map.
- Finally, possible/feasible route was identified based on various physical and cultural parameters and their inherent properties.
- The cost reduction analysis was also done for substantiating the formation of highway. Finally, possible/feasible route was identified based on various physical and cultural parameters and their inherent properties. The cost reduction analysis was also done for substantiating the formation of highway.

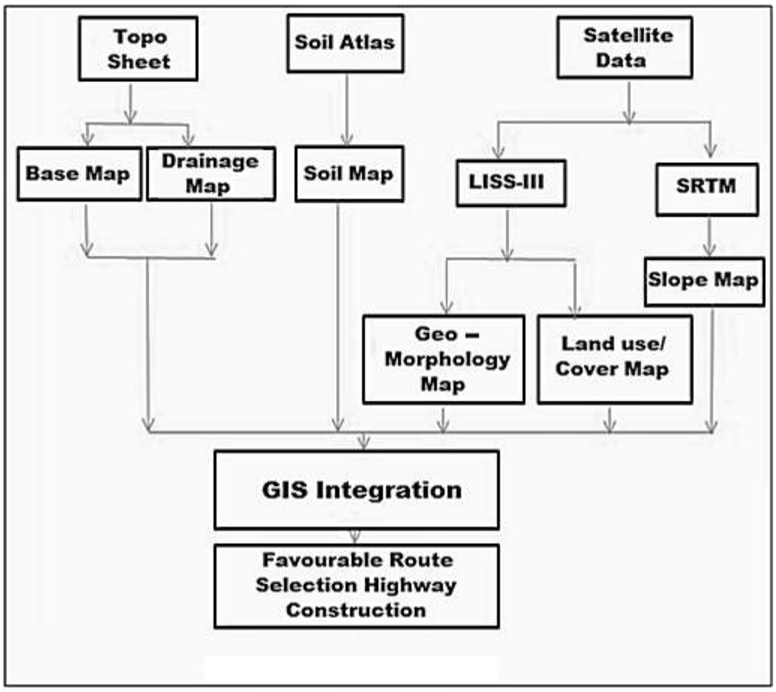


Figure.5.22. Methodology

- The main purpose of traffic survey are traffic monitoring, traffic control and management, traffic enforcement, traffic forecasting, model calibration and validating etc.
- The purpose of caring out traffic volume count are designing, improving traffic system, planning, management.
- The traffic volume count study is carried out to get following useful information. Magnitudes, classifications and the time and directional split of vehicular flows

- Proportion of vehicles in traffic stream - Hourly, daily, yearly and seasonal variation of vehicular flows
- Flow fluctuation on different approaches at a junction or different parts of a road network system.
- Network analysis is used to find the shortest alternated routes between origins to destination Network Analyst provides network-based spatial analysis tools for solving complex routing problems.

4.3.8. Conclusion

- The purpose of this study was to develop a tool to locate a suitable less time consuming, shortest route between two points.
- The GIS approach using ground parameters and spatial analysis provided to achieve this goal. Raster based map analysis provide a wealth of capabilities for incorporating terrain information surrounding linear infrastructure.
- Costs resulting from terrain, geomorphology, land use, drainage and elevation resulting the shortest routes for the study area.
- Results indicate that the route which was designed applying GIS method is avoid traffic problems, less time consuming more environmentally effective, and cheaper.
- This proposed shortest route provides traffic free, pollution free, risk free, operating for movement of vehicle passing from chettikullam to kottar.
- Time and consumption of fuel will also be reduced considerably. GIS method can also be used for route determination for irrigation, drainage channels, power lines and railways.

4.4. TWO MODELS OF DIGITAL EDUCATION

From the introduction of the World Wide Web in 1993 the young of the world have experienced two models of digital education, that outside the school walls and that within.

Outside the young and the digitally connected families of the world employed – unseen – the naturally evolving laissez faire model. Within the school the young worked within the traditional, highly structured model.

4.10 GIS

It is time the difference is understood, the global success and benefits of the laissez faire recognised and lauded, and the serious shortcomings of the highly structured understood and addressed.

For much of the period the two models ran in parallel, with most schools showing little or no interest in the out of school digital education.

Around 2010 – 2012 the scene began to change when a handful of digitally mature schools began genuinely collaborating with their families in the 24/7/365 digital education of the children. Those schools had reached the evolutionary stage where their teaching model and culture closely mirrored that of the families. They revealed what was possible with collaboration.

That said it took time for that collaboration to take hold more widely and for the most part the parallel models continue in operation today, with the difference between the in and out of school teaching growing at pace.

It is surely time for schools and government to question the retention of the parallel modes and to ask if taxpayers are getting value for the millions upon millions spent solely on schools when the digitally connected families receive no support.

Might it be time to employ a more collaborative approach where the schools complement and add value to the contribution of the families?

Without going into detail, it bears reflecting on the distinguishing features of the learning environment and digital education model, of both the digitally connected family and the school, and asking what is the best way forward,

4.4.1. The learning environments.

Digitally connected families

That of the families we know well. It has been built around the home’s warmth and support, and the priority the parents attached to their children having a digital education that would improve their education and life chances. The focus has always been on the child – the individual learner – with the children from the outset being provided the current technology by their family and empowered to use that technology largely unfettered.

Importantly the family as a small regulating unit, with direct responsibility for a small number of children could readily trust each, and monitor, guide and value their learning from birth onwards, assisting ensure each child had use of the current technology and that the use was wise and balanced.

The learning occurred within a freewheeling, dynamic, market driven, naturally evolving environment, anywhere, anytime, just in time and invariably in context. Those interested could operate at the cutting edge and the depth desired.

Very early on the young’s use of the digital was normalized, with the learning occurring as a natural part of life, totally integrated, with no regard for boundaries

The time available to the digitally connected family was – and continues to be – at least four/five times greater than that in the school.

It was too many seemingly chaotic, but also naturally evolving.

Very quickly the family learning environment became collaborative, socially networked, global in its outlook, highly enjoyable and creative where the young believed anything was possible.

By the latter 2000's most families had created – largely unwittingly – their own increasingly integrated and sophisticated digital ecosystem, operating in the main on the personal mobile devices that connected all in the family to all manner of other ecosystems globally.

4.4.2. Digital learning in the school.

The general feature of the school digital learning environment has been invariably one of unilateral control, where the ICT experts controlled every facet of the technology and its teaching.

They chose, configured and controlled the use of both the hardware and software, invariably opting for one device, one operating system and a standard suite of applications.

The students were taught within class groups, using highly structured, sequential, teacher directed, regularly assessed instructional programs.

The school knew best. The clients – the parents and students – were expected to acquiesce. There was little or no recognition of the out of school learning or technology or desire to collaborate with the digitally connected families.

The teaching was insular, inward looking, highly site fixated.

In reflecting on school's teaching with the digital between 1993 and 2016 there was an all-pervasive sense of constancy, continuity, with no real rush to change. There was little sense that the schools were readying the total student body to thrive within in a rapidly evolving digitally based world.

Significantly by 2016 only a relatively small proportion of schools globally were operating as mature digital organizations, growing increasingly integrated, powerful higher order digitally based ecosystems.

The reality was that while the learning environment of the digitally connected families evolved naturally at pace that of most schools changed only little, with most schools struggling to accommodate rapid digital evolution and transformation.

4.4.3. The teaching models

With the advantage of hindsight, it is quite remarkable how hidden the laissez faire model has remained for twenty plus years, bearing in mind it has been employed globally since the advent of the WWW.

4.12 GIS

For years, it was seen simply as a different, largely chaotic approach used by the kids – with the focus being on the technological breakthroughs and the changing practices rather than on the underlying model of learning that was being employed.

It wasn't until the authors identified and documented the lead role of the digitally connected families of the world did we appreciate all were using basically the same learning approach. The pre-primary developments of the last few years affirmed the global application of the model.

We saw at play a natural model that was embraced by the diverse families of the world.

All were using the same model – a naturally evolving model where the parents were 'letting things take their own course '(OED).

The learning was highly individualized, with no controls other than the occasional parent nudge. That said the learning was simultaneously highly collegial, with the young calling upon and collaborating with their siblings, family members, peers and social networks when desired.

Interestingly from early on the young found themselves often knowing more about the technology in some areas than their elders – experiencing what Tapscott (1998) termed an 'inverted authority' – being able to assist them use the technology.

Each child was free to learn how to use, and apply those aspects of the desired technologies they wanted, and to draw upon any resources or people if needed.

In the process the children worldwide – from as young as two – directed their own learning, opting usually for a discovery based approach, where the learning occurred anytime, anywhere 24/7/365. Most of the learning was just in time, done in context and was current, relevant, highly appealing and intrinsically motivating. Invariably it was highly integrated, with no thought given to old boundaries – like was it educational, entertainment, communication, social science or history.

In contrast the school digital teaching model has always been highly structured and focused on what the school or education authority 'experts' believed to be appropriate.

Throughout the period the teaching has been unilaterally controlled, directed by the classroom teacher, with the students disempowered, distrusted and obliged to do as told.

The teaching built upon linear, sequential instructional programs where the digital education was invariably treated like all other subjects, shoehorned into an already crowded curriculum and continually assessed. Some authorities made the 'subject' compulsory, others made it optional.

The focus – in keeping with the other 'subjects' in the curriculum – was academic. There was little interest in providing the young the digital understanding for everyday life.

The teaching took place within a cyber walled community, at the time determined by the teaching program.

Increasingly the course taught and assessed became dated and irrelevant.

In considering why the young and the digitally connected families of the world have embraced the laissez faire model of digital education aside from the young's innate curiosity and desire to learn we might do well to examine the model of digital learning we have used over the last twenty plus years and reflect on how closely it approximates that adopted by the young.

Might they be following that ancient practice of modelling the behavior of their parents?

4.4.4. The way forward

Near a quarter of a century on since the introduction of the WWW and an era of profound technological and social change it is surely time for governments and educators globally to

- Publicly recognise the remarkable success of the digitally connected families and the laissez faire teaching model in the 24/7/365 digital education of both the children and the wider family
- Understand the digitally connected families are on trend to play an even greater lead role
- identify how best to support the family's efforts without damaging the very successful teaching model employed
- Consider how best to enhance the educational contribution of all the digitally connected families in the nation, including the educationally disadvantaged
- Rethink the existing, somewhat questionable contribution of most schools and the concept of schools as the sole provider of digital education for the young
- Examine where scarce taxpayer monies can best be used to improve the digital education in the networked world.

Let us all finally recognize the core qualities and the remarkable global success of the laissez faire digital education model and build upon its achievements.

4.5. 3D DATA COLLECTION AND UTILIZATION

Geographic Information Science (GIS) offers powerful tools for performing detailed analysis of spatial information and solving complex problems. Traditional GIS data is based on mapping in two dimensions, an x and y-value, which can be limiting in some applications. Utilizing 3D GIS software lets users engage with data from a whole new perspective that results in more nuanced insights and detailed visualizations.

- 3D GIS bring enhanced depth into data collection and analysis by incorporating a z-value into mapping. Most commonly, that means including elevation data, but users have many options for adding layers of information. For instance, a map might include a dimension based on the concentrations of certain chemicals and minerals or which parcels of land are best suited for development. Working with three dimensions, GIS professionals can often apply their findings to address real-world issues with greater accuracy.
- While 3D models are more difficult to create and maintain than 2D ones, there are myriad 3D GIS applications where this technology is greatly beneficial. These four

examples demonstrate how an investment in 3D GIS modeling can generate added value:

4.5.1. City Planning

Cities have a way of growing to encompass previously under- or undeveloped areas in a process often called urbanization or urban sprawl. There are many reasons behind urban sprawl, including a desire to build improved infrastructure, affordable land or tax rates, or overcrowding inside the city. Urban sprawl can have a major impact on people who decide to leave the city as well as those who remain. For example, as residents move farther away from the city center, infrastructure such as roads or public transportation systems must accommodate their commutes, and traffic can lead to higher rates of air pollution.

To minimize the negative impacts of urban sprawl and increased development, it's important for city planners to carefully determine the best way to grow urban areas. Urban development needs to take into consideration today's requirements, potential changes in demand and the long-term effects of building upward and outward.

3D GIS software can help city planners visualize what their proposed changes will look like and predict the outcomes for current residents and future generations. One example was the 2012 revitalization of the Mulheim Sud district in Cologne, Germany, located on the Rhine River. The project set out to make the district, which included a mix of residential, commercial and industrial buildings, more environmentally friendly over the course of two decades.

A 3D model spotlighted building information, aerial photos, energy performance, air pollution, lidar elevation data, noise and traffic. The wide range of integrated information allowed architects, planning engineers and others to collaborate effectively. As the district develops further, the 3D model will help future planners with energy and environmental modeling and guide public participation initiatives.

4.5.2. Building Information Modeling

Building information modeling (BIM) is a technology that generates digital representations of facilities and relevant processes. BIM has given facilities managers the ability to closely review structures, beginning with the construction planning phase.

Used in conjunction with 3D GIS data, BIM can help create robust building management plans and allow for more detailed analysis. For example, before breaking ground on a construction project, stakeholders can review findings from GIS and BIM to draw conclusions about environmental impact, sustainability, disaster readiness and how to optimize the use of assets and space.

BIM and 3D GIS can also come together to support the preservation and restoration of historical buildings. An effort to digitally record cultural heritage sites in Dublin, Ireland, drew on Historic Building Information Modeling (HBIM) and 3D GIS to document and analyze selected locations.

One project focused on restoring buildings along Henrietta Street, which dates its earliest construction to 1730. By the 21st century, the street was lined with buildings in serious need of care. HBIM technology made it possible to map the extent of the damage and visualize what the area looked like when new. Researchers employed GIS tools to note attributes of individual buildings like the years of construction and address information. In the process, they developed a store of information that could be used to generate in-depth visualizations or guide tourists.

4.5.3. Coastal Modeling and Analysis

A nation’s coastline is a crucial gateway for imports and exports, and about 40 percent of the world’s population lives within 60 miles of a coast. But these areas also pose numerous challenges for development.

Its critical for planners to understand the factors that affect construction and maintenance for shipping ports, fisheries, mineral mining operations and wilderness preservation areas. Responsible coastal development must be informed by underwater topography, local vegetation and predictions for the long-term environmental impact.

Resource planning systems that draw on GIS can provide insights into the economic, environmental and cultural results of activities along the coast. The right data makes all the difference in sustainably performing operations like construction or excavation. When preparing for the extraction of resources on the coastline, organizations benefit from synthesizing information like:

- Findings from 3D GIS mapping that suggest the likely outcomes of dredging material in the water
- Lidar topographical surveys
- Data sets from past extraction activities
- Trends in coastal change

4.5.4. Wind Farm Assessment

Planning a wind farm requires a detailed analysis of an environment and the potential effects of the structures. By using 3D GIS modeling, planners and other stakeholders can get a better idea of the impact from wind farm development on wildlife and people.

4.16 GIS

For example, when assessing possible wind farm locations in two dimensions, a bird's migratory path might make a location seem inaccessible. However, reviewing that same space using 3D GIS data may reveal that the elevation of birds' flight paths and the height of the wind farm are compatible.

In Switzerland, developers wanted to find a way to accurately determine the noise that would be generated from the installation of a new wind farm. A team developing a visual-acoustic simulation tool decided to study Mont Crosin in the Canton of Bern, Switzerland, which is home to 16 wind turbines. The researchers analyzed recordings taken on days with varying wind and weather conditions and wind speed measurements taken with a 3D ultrasonic anemometer. They generated 3D models representing vegetation, infrastructure and the wind turbines themselves. The data allowed planners to predict the noise and environmental impact that would be produced by the proposed wind farm.

UNIT V DATA MANAGEMENT AND OUTPUT

Import/Export – Data Management functions- Raster to Vector and Vector to Raster Conversion -Data Output - Map Compilation – Chart/Graphs – Multimedia – Enterprise Vs. Desktop GISdistributed GIS.

5.1 Import/Export

Importing and exporting data in a Geographic Information System (GIS) are essential functions that facilitate the integration, sharing, and utilization of spatial data across different systems and applications. Here's a detailed overview of how import and export processes work in GIS:

Importing Data into GIS

1. **File Formats**: GIS systems support a variety of data formats for import. Common formats include:
   - **Vector Formats**: Shapefiles (.shp), GeoJSON (.geojson), KML/KMZ (.kml, .kmz), DXF (.dxf)
   - **Raster Formats**: TIFF (.tif), JPEG (.jpg), PNG (.png), GRID
   - **Database Formats**: SQL databases (e.g., PostgreSQL/PostGIS, Microsoft SQL Server), GeoPackage (.gpkg)
   - **Spreadsheet Formats**: CSV (.csv), Excel (.xls, .xlsx)
2. **Geocoding**: For importing address data, geocoding processes convert addresses into geographic coordinates (latitude and longitude) that can be mapped.
3. **Import Tools**: Most GIS software includes import tools or wizards to guide users through the process of loading data into the system. This often involves:
   - Specifying file paths and formats
   - Setting coordinate systems or projections
   - Configuring data field mappings
4. **Data Transformation**: During import, data may need to be transformed to match the GIS's coordinate system or projection. Transformation tools help align spatial data to a common reference frame.
5. **Data Validation**: After import, it's crucial to check the integrity of the data, ensuring there are no errors or inconsistencies. This might involve visual inspection or running validation tools.

Exporting Data from GIS

1. **File Formats**: GIS can export data in various formats, depending on the needs of the users or systems that will receive the data:
   - **Vector Formats**: Shapefiles, GeoJSON, KML/KMZ, DXF
   - **Raster Formats**: TIFF, JPEG, PNG, GRID
   - **Database Formats**: GeoPackage, SQL dump
   - **Spreadsheet Formats**: CSV, Excel
2. **Export Tools**: GIS software typically provides export functions that allow users to:
   - Select the data layers or datasets to export
   - Choose the output format and specify export parameters (e.g., file name, location)

- Set coordinate systems or projections for the exported data

3. **Data Simplification**: When exporting, it may be necessary to simplify or generalize the data to reduce file size or make it compatible with other systems. This might involve:
    - Reducing the level of detail in vector data
    - Adjusting raster resolution
4. **Metadata Preservation**: Ensure that metadata (information about the data, such as source, accuracy, and purpose) is preserved or properly documented during export, especially if the data is being shared with others.
5. **Quality Assurance**: Verify the exported data to ensure it meets the intended specifications and that the integrity of the data has been maintained. This may involve checking for data completeness, accuracy, and format compatibility.

Common Use Cases

- **Integration**: Importing data from various sources to combine and analyze in GIS, such as integrating survey data with existing maps.
- **Sharing**: Exporting maps or data for use in reports, presentations, or other software applications.
- **Interoperability**: Facilitating data exchange between different GIS platforms or with non-GIS systems.
- **Archiving**: Exporting data for long-term storage or for backup purposes.

Best Practices

- **Standardize Formats**: Use widely accepted formats (e.g., GeoJSON for web applications, shapefiles for traditional GIS) to ensure compatibility and ease of use.
- **Maintain Metadata**: Always include or retain metadata to provide context and information about the data.
- **Ensure Compatibility**: Check that the data formats and coordinate systems are compatible with the receiving system or application.
- **Regular Updates**: Keep software and tools up-to-date to support the latest formats and features.

By effectively managing the import and export of data, GIS users can ensure seamless integration, accurate analysis, and effective sharing of geographic information across different platforms and applications.

5.2 Data Management functions

Data management in Geographic Information Systems (GIS) involves a range of functions that ensure spatial data is collected, organized, maintained, and used effectively. Here's a comprehensive look at the key data management functions within GIS:

1. Data Collection

- **Source Identification**: Identifying and accessing various sources of geographic data, including satellite imagery, aerial surveys, field surveys, existing maps, and online databases.

- **Data Acquisition**: Gathering data through different methods such as GPS, remote sensing, crowdsourcing, or purchasing from data vendors.
- **Data Entry**: Inputting data into the GIS system, which may involve manual digitizing of maps or importing from digital files.

2. Data Storage

- **Database Management Systems (DBMS)**: Storing data in spatial databases like PostGIS, Oracle Spatial, or Microsoft SQL Server with spatial extensions. These systems support complex queries and large datasets.
- **File Storage**: Using file-based systems such as shapefiles, GeoJSON, or KML for simpler or smaller-scale projects.
- **Data Warehouses**: For large-scale and enterprise-level GIS implementations, data warehouses are used to store vast amounts of spatial data efficiently.

3. Data Organization

- **Data Structuring**: Organizing data into layers or datasets based on themes or categories (e.g., land use, transportation, hydrology).
- **Metadata Management**: Maintaining metadata that describes the data's source, accuracy, format, and other characteristics to support data quality and usability.
- **Spatial Indexing**: Creating spatial indexes to improve the efficiency of querying and analyzing spatial data. This includes R-trees, Quad-trees, and other indexing methods.

4. Data Integration

- **Data Merging**: Combining data from different sources or formats into a unified GIS system. This often requires aligning coordinate systems and projections.
- **Data Transformation**: Converting data from one format or coordinate system to another to ensure compatibility and accuracy.
- **Data Linking**: Integrating attribute data with spatial data by linking tables or databases, allowing for more comprehensive analysis.

5. Data Analysis

- **Spatial Analysis**: Performing analyses such as buffering, overlay analysis, spatial joins, and proximity analysis to extract insights from spatial data.
- **Geostatistics**: Applying statistical techniques to analyze spatial patterns and relationships, such as kriging or spatial regression.
- **Modeling and Simulation**: Creating models to simulate geographic phenomena or predict outcomes based on spatial data.

6. Data Quality Management

- **Data Validation**: Checking for errors or inconsistencies in the data, such as topology errors or incorrect attribute values.
- **Data Cleaning**: Correcting or removing erroneous data and filling in missing values to improve data accuracy.
- **Quality Control**: Implementing procedures to ensure ongoing data accuracy and consistency, including regular audits and updates.

7. Data Backup and Recovery

- **Backup Procedures**: Regularly creating backups of GIS data to prevent data loss due to hardware failure, corruption, or other issues.
- **Disaster Recovery**: Developing plans and processes for restoring data and functionality in case of data loss or system failure.

8. Data Security

- **Access Control**: Implementing permissions and authentication measures to control who can access and modify the GIS data.
- **Data Encryption**: Protecting data through encryption, both during transmission and at rest, to prevent unauthorized access.
- **Audit Trails**: Keeping records of changes made to the data to track and review modifications.

9. Data Sharing and Distribution

- **Data Export**: Facilitating the export of GIS data in various formats for use in other applications or for sharing with stakeholders.
- **Web Services**: Providing data access through web services such as Web Map Services (WMS), Web Feature Services (WFS), or Web Coverage Services (WCS).
- **Public and Private Sharing**: Managing how data is shared with the public or within an organization, ensuring appropriate levels of access and confidentiality.

10. Data Maintenance

- **Updating Data**: Regularly updating spatial data to reflect new information or changes in the geographic features being monitored.
- **Version Control**: Managing different versions of the data to track changes over time and maintain historical records.
- **Archiving**: Storing older versions of data or less frequently accessed data in an archive to keep the primary database current and efficient.

11. Data Visualization and Reporting

- **Map Production**: Creating and customizing maps to visually represent spatial data and communicate findings.

- **Reporting Tools**: Generating reports and summaries based on GIS data, which can include tables, charts, and descriptive text.

Effective data management in GIS ensures that spatial data is accurate, accessible, and useful for analysis and decision-making. By implementing robust practices in each of these areas, GIS professionals can optimize the value and effectiveness of their geographic information systems.

5.2 Data Management functions

Data management in Geographic Information Systems (GIS) involves a suite of functions essential for organizing, maintaining, and utilizing spatial data effectively. Here's a detailed breakdown of these functions:

1. Data Collection

- **Source Identification**: Locating and identifying sources of spatial data, such as satellite imagery, aerial photography, GPS data, surveys, and existing maps.
- **Data Acquisition**: Collecting data through various methods including direct field surveys, remote sensing, crowdsourcing, and purchasing from data providers.
- **Data Entry**: Inputting data into the GIS system, which can involve digitizing features from maps, entering attribute information, or importing data files.

2. Data Storage

- **Database Management Systems (DBMS)**: Storing spatial data in specialized databases such as PostgreSQL/PostGIS, Oracle Spatial, or Microsoft SQL Server with spatial extensions. These systems support spatial queries and handle large datasets efficiently.
- **File-Based Storage**: Using formats like shapefiles, GeoJSON, KML, and raster files for data storage, especially for simpler or smaller-scale projects.
- **Data Warehousing**: Implementing data warehouses to store and manage large volumes of spatial data for enterprise-level applications.

3. Data Organization

- **Layer Management**: Organizing data into thematic layers (e.g., land use, roads, water bodies) to facilitate analysis and visualization. Each layer represents a specific type of spatial data.
- **Metadata Management**: Creating and maintaining metadata that describes the data's origin, accuracy, format, and other relevant attributes to support data quality and usability.
- **Spatial Indexing**: Creating spatial indexes (e.g., R-trees, Quad-trees) to optimize spatial queries and improve performance when working with large datasets.

4. Data Integration

- **Data Merging**: Combining data from multiple sources into a single GIS framework. This may involve aligning different datasets and resolving discrepancies.

- **Coordinate System Transformation**: Converting data from one coordinate system or projection to another to ensure consistency across datasets.
- **Data Linking**: Connecting spatial data with attribute data through relational databases, allowing for comprehensive analysis and visualization.

5. Data Analysis

- **Spatial Analysis**: Performing operations such as buffering, overlay analysis, and spatial joins to extract insights from spatial relationships and patterns.
- **Geostatistics**: Applying statistical methods to analyze spatial data, including techniques like interpolation, kriging, and spatial regression.
- **Modeling and Simulation**: Creating models to simulate spatial phenomena or predict outcomes based on geographic data, such as flood modeling or land-use planning.

6. Data Quality Management

- **Data Validation**: Ensuring data accuracy and consistency by checking for errors, such as topology errors in vector data or anomalies in raster data.
- **Data Cleaning**: Correcting or removing erroneous data, addressing issues such as duplicate records, incorrect values, or incomplete information.
- **Quality Assurance**: Implementing procedures to maintain data quality over time, including routine checks and updates.

7. Data Backup and Recovery

- **Backup Strategies**: Regularly creating backups of GIS data to prevent loss due to hardware failures, corruption, or other issues. This includes full backups and incremental backups.
- **Disaster Recovery Planning**: Developing plans and procedures to restore data and resume operations in case of data loss or system failure.

8. Data Security

- **Access Control**: Implementing authentication and authorization measures to manage who can view or modify data. This includes setting user roles and permissions.
- **Data Encryption**: Protecting sensitive data through encryption methods both during transmission and while at rest.
- **Audit Trails**: Keeping logs of data access and modifications to track changes and ensure accountability.

9. Data Sharing and Distribution

- **Export Functions**: Facilitating the export of GIS data in various formats to be used in other applications or for sharing with external stakeholders.

- **Web Services**: Providing access to spatial data through web services such as Web Map Services (WMS), Web Feature Services (WFS), and Web Coverage Services (WCS) for integration with web applications.
- **Public and Private Access**: Managing data sharing options to control how data is distributed, whether publicly, within an organization, or to selected partners.

10. Data Maintenance

- **Regular Updates**: Keeping data current by incorporating new information and changes. This involves updating datasets to reflect the latest developments or corrections.
- **Version Control**: Managing different versions of datasets to track changes over time and maintain historical records.
- **Archiving**: Storing older or less frequently accessed data in archives to keep the primary database streamlined and efficient.

11. Data Visualization and Reporting

- **Map Production**: Creating and customizing maps to visually represent spatial data and convey information effectively. This includes designing thematic maps, interactive maps, and dashboards.
- **Reporting Tools**: Generating reports and summaries that incorporate spatial data, including tables, charts, and descriptive text to support decision-making.

Effective data management in GIS ensures that spatial data is accurate, accessible, and effectively utilized for analysis and decision-making. By employing these functions, GIS professionals can optimize the performance and value of their geographic information systems.

5.3 Raster to Vector and Vector to Raster Conversion

In Geographic Information Systems (GIS), converting between raster and vector data formats is a common task that can be essential for various types of analysis and data integration. Here's a detailed overview of the processes involved in raster-to-vector and vector-to-raster conversion:

Raster to Vector Conversion

Raster-to-vector conversion involves transforming grid-based raster data (which represents spatial information as a matrix of cells or pixels) into vector data (which represents spatial information as discrete points, lines, or polygons). This process is often used to extract features from raster images, such as converting scanned maps or satellite imagery into vector format for more detailed analysis.

1. Conversion Process

1. **Thresholding/Binarization**:
    - **Purpose**: Separate relevant features from the background.

   - **Method**: Apply a threshold to the raster to create a binary image where features of interest are distinguished from non-features.
2. **Edge Detection**:
   - **Purpose**: Identify boundaries or edges of features.
   - **Method**: Use edge detection algorithms (e.g., Sobel, Canny) to find the contours of features in the raster data.
3. **Vectorization**:
   - **Purpose**: Convert the detected edges or boundaries into vector format.
   - **Method**: Trace the edges and convert them into vector lines or polygons. Tools like the `Raster to Polygon` function in GIS software can automate this process.
4. **Simplification**:
   - **Purpose**: Reduce the complexity of vector data.
   - **Method**: Simplify vector lines or polygons by reducing the number of vertices while maintaining the general shape and accuracy.
5. **Attribute Assignment**:
   - **Purpose**: Attach attributes to vector features.
   - **Method**: Assign attribute data (e.g., land use types, class names) to the vector features based on the original raster values or other reference data.

2. Applications

- **Feature Extraction**: Extracting land cover types, roads, or other features from satellite imagery.
- **Cartographic Processing**: Converting scanned maps into vector formats for editing and analysis.
- **Land Use Mapping**: Converting land cover rasters to vector format for detailed land use planning.

Vector to Raster Conversion

Vector-to-raster conversion involves transforming vector data (which represents spatial information as points, lines, or polygons) into grid-based raster data. This process is useful for creating raster layers from vector datasets for various types of spatial analysis.

1. Conversion Process

1. **Rasterization**:
   - **Purpose**: Convert vector features into a raster grid.
   - **Method**: For each pixel in the raster grid, determine whether it falls within a vector feature (e.g., point, line, or polygon). Assign values to pixels based on the presence of vector features.
2. **Cell Resolution**:
   - **Purpose**: Define the level of detail in the raster.
   - **Method**: Set the cell size or resolution of the raster. A smaller cell size provides higher detail but results in larger file sizes.
3. **Attribute Encoding**:
   - **Purpose**: Encode vector attributes into the raster.

   - **Method**: Assign values to raster cells based on vector attributes (e.g., classifying land cover types or assigning numerical values).
4. **Aggregation/Resampling**:
   - **Purpose**: Adjust raster resolution or combine multiple raster layers.
   - **Method**: Use aggregation or resampling techniques to modify the raster grid to match the desired resolution or format.

2. Applications

- **Density Analysis**: Converting point data (e.g., locations of incidents) into raster format to perform density or hotspot analysis.
- **Suitability Modeling**: Creating raster layers from vector data for suitability or risk modeling (e.g., land suitability for agriculture).
- **Map Creation**: Generating raster maps from vector data for visualization and cartographic purposes.

Tools and Software

- **ArcGIS**: Offers tools such as `Raster to Polygon` for raster-to-vector conversion and `Feature to Raster` for vector-to-raster conversion.
- **QGIS**: Provides tools like `Rasterize (vector to raster)` and `Polygonize (raster to vector)` to perform these conversions.
- **GDAL**: A popular open-source library that includes command-line tools like `gdal_rasterize` for vector-to-raster and `gdal_polygonize` for raster-to-vector conversion.

Best Practices

- **Resolution Matching**: Ensure that the resolution and extent of the raster output match the requirements of your analysis.
- **Data Quality**: Verify the accuracy and quality of the converted data, and correct any errors or artifacts that may have occurred during the conversion process.
- **Post-Processing**: After conversion, perform necessary post-processing steps such as data cleaning, attribute assignment, or spatial adjustments to refine the results.

By understanding and utilizing these conversion techniques, GIS professionals can effectively manage and analyze spatial data in various formats to support diverse analytical and decision-making needs.

5.4 Data Output

In Geographic Information Systems (GIS), data output refers to the methods and formats used to present and share spatial information after processing and analysis. Effective data output ensures that the results of GIS analyses are communicated clearly and can be utilized for decision-making or further analysis. Here's a detailed look at the various aspects of data output in GIS:

1. Data Visualization

- **Maps**: Creating static or interactive maps to visually represent spatial data. Maps can be thematic (showing specific themes like land use or population density) or general-purpose (providing a broad overview of geographic areas).
    - **Static Maps**: Printable maps with fixed content, often used in reports or publications.
    - **Interactive Maps**: Online maps that users can interact with, zooming in and out, querying features, and more.
- **Charts and Graphs**: Representing spatial data through charts (e.g., bar charts, pie charts) and graphs to illustrate trends or distributions.
    - **Histograms**: Showing the distribution of attribute values.
    - **Pie Charts**: Illustrating proportions of different categories.
- **3D Visualization**: Using 3D models to represent terrain, buildings, or other spatial features. This can include terrain elevation models, 3D city models, or virtual reality environments.

2. Reporting and Analysis Results

- **Reports**: Generating detailed reports that include maps, charts, tables, and textual descriptions of GIS analyses. These reports are often used for decision-making and presentation purposes.
    - **Automated Reports**: Reports generated automatically based on predefined templates and analysis results.
    - **Custom Reports**: Tailored reports created manually to address specific questions or audiences.
- **Dashboards**: Interactive platforms that combine maps, charts, and other data visualizations in a single interface, allowing users to monitor and analyze key metrics and spatial data in real-time.

3. Data Export

- **File Formats**: Exporting GIS data into various formats for use in other applications or for sharing with stakeholders. Common formats include:
    - **Vector Formats**: Shapefiles (.shp), GeoJSON (.geojson), KML/KMZ (.kml, .kmz)
    - **Raster Formats**: TIFF (.tif), JPEG (.jpg), PNG (.png)
    - **Database Formats**: GeoPackage (.gpkg), SQL databases
    - **Spreadsheet Formats**: CSV (.csv), Excel (.xls, .xlsx)
- **Export Tools**: GIS software provides tools to export data, allowing users to specify the format, extent, and attributes to include.

4. Web and Mobile Applications

- **Web Mapping Services**: Publishing GIS data and maps online through services like Web Map Services (WMS), Web Feature Services (WFS), or Web Coverage Services (WCS). These services allow users to access and interact with GIS data via web browsers.
    - **Map Servers**: Tools like ArcGIS Online or QGIS Server that provide web access to GIS data.

- **Mobile Apps**: Creating mobile applications that provide access to GIS data and functionality on smartphones and tablets. Mobile apps can include field data collection tools, interactive maps, and location-based services.

5. Data Sharing and Collaboration

- **File Sharing**: Distributing GIS data files through file-sharing services or direct transfers, ensuring that recipients have the necessary software and permissions to access the data.
- **Collaborative Platforms**: Using platforms like ArcGIS Online, Google Earth Engine, or open-source solutions to share data and collaborate with other users or organizations.

6. Printing and Publishing

- **Map Printing**: Creating high-quality prints of maps for physical distribution or display. This can include adjusting map layouts, scales, and legend details.
- **Publishing**: Preparing GIS data and maps for inclusion in publications, reports, or presentations, ensuring that the output meets professional standards and is suitable for the intended audience.

7. Alerts and Notifications

- **Automated Alerts**: Setting up automated notifications based on changes or thresholds in GIS data. For example, alerts for significant changes in land use or environmental conditions.
- **Email Notifications**: Sending email updates or notifications that include links to updated maps or reports.

8. Custom Output

- **Custom Applications**: Developing specialized applications or scripts to generate tailored output based on specific requirements, such as custom data visualizations or automated analysis reports.

Best Practices

- **Understand Your Audience**: Tailor data output to meet the needs and technical capabilities of your intended audience, whether they are technical professionals, decision-makers, or the general public.
- **Ensure Accuracy**: Verify that all output data and visualizations are accurate and reflect the results of the GIS analysis properly.
- **Maintain Consistency**: Use consistent symbols, colors, and formats across outputs to avoid confusion and ensure clarity.
- **Provide Context**: Include necessary contextual information, such as legends, scales, and data sources, to make the output easily understandable.

Effective data output in GIS enhances the ability to communicate and utilize spatial information, supporting informed decision-making and enabling collaboration across various stakeholders.

5.5 Map Compilation

Map compilation in Geographic Information Systems (GIS) involves the process of creating, assembling, and producing maps that represent spatial data. This process integrates various types of data and visualizes them in a coherent and meaningful way. Here's an in-depth look at the map compilation process in GIS:

1. Define Map Objectives

- **Purpose**: Determine the purpose of the map (e.g., thematic map, topographic map, cadastral map) and the target audience (e.g., general public, researchers, planners).
- **Content**: Decide what information will be included on the map, such as land use, infrastructure, demographics, or environmental features.

2. Gather and Prepare Data

- **Data Collection**: Collect spatial data from various sources, including:
    - **Satellite Imagery**: Provides high-resolution images of the Earth's surface.
    - **Aerial Photography**: Offers detailed views from airborne platforms.
    - **Field Surveys**: Collects ground-based data through GPS or manual surveys.
    - **Existing Maps and Databases**: Integrates data from historical maps, government databases, and other sources.
- **Data Cleaning**: Ensure data accuracy and consistency by:
    - **Correcting Errors**: Fixing inaccuracies and inconsistencies in spatial and attribute data.
    - **Standardizing Formats**: Converting data into compatible formats and coordinate systems.

3. Design Map Layout

- **Map Elements**: Determine the key elements to include:
    - **Title**: Clearly states the purpose and content of the map.
    - **Legend**: Explains the symbols, colors, and patterns used on the map.
    - **Scale**: Shows the relationship between distances on the map and real-world distances.
    - **North Arrow**: Indicates map orientation.
    - **Labels**: Provides names and descriptions for features and areas.
    - **Grids and Coordinates**: Adds reference grids or coordinate systems if needed.
- **Design Principles**: Apply cartographic design principles to ensure clarity and readability:
    - **Symbology**: Use appropriate symbols and colors to represent different types of data.
    - **Hierarchical Design**: Highlight important features and de-emphasize less critical elements.
    - **Color Schemes**: Choose colors that are distinguishable and meaningful for different data types.

4. Map Production

- **Digital Map Creation**:
  - **Layer Management**: Organize data into layers (e.g., roads, land use, water bodies) and adjust their visibility and order.
  - **Data Integration**: Combine various datasets into a single map project, ensuring they align correctly.
  - **Map Rendering**: Generate the map view based on the design and data layers, using GIS software tools.
- **Map Customization**: Adjust the map to enhance its appearance and functionality:
  - **Label Placement**: Position labels to avoid overlap and ensure readability.
  - **Symbol Adjustment**: Modify symbols and styles to match the map's purpose and improve visual clarity.
  - **Map Insets**: Add inset maps to show areas of interest or provide additional context.

5. Review and Quality Assurance

- **Accuracy Check**: Verify that the map accurately represents the spatial data and adheres to the defined objectives.
- **Design Review**: Ensure the map layout is clear, visually appealing, and meets design standards.
- **Feedback**: Obtain feedback from stakeholders or target users and make necessary adjustments.

6. Export and Distribution

- **Export Formats**: Save the final map in appropriate formats for distribution:
  - **Digital Formats**: PDF, JPEG, PNG, TIFF for sharing online or through electronic media.
  - **Interactive Formats**: Web maps or applications for interactive exploration.
  - **Print Formats**: High-resolution files for printing, ensuring proper resolution and layout for physical copies.
- **Distribution Channels**: Share the map with the intended audience through:
  - **Web Services**: Publish online using platforms like ArcGIS Online, Google Maps, or custom web applications.
  - **Print**: Distribute physical copies through mail or at events.
  - **Mobile**: Provide access via mobile apps or devices.

7. Update and Maintenance

- **Data Updates**: Regularly update the map to reflect new data or changes in existing data.
- **Version Control**: Maintain version history to track changes and updates to the map.

Tools and Software for Map Compilation

- **ArcGIS**: Offers a comprehensive suite of tools for map design, layer management, and publishing.

- **QGIS**: An open-source GIS platform with robust map design and editing capabilities.
- **MapInfo**: Provides tools for creating, analyzing, and sharing maps.
- **Adobe Illustrator with GIS Plugins**: For advanced cartographic design and customization.

Best Practices

- **Clarity and Simplicity**: Aim for a clear and uncluttered design that effectively communicates the intended information.
- **Audience Awareness**: Tailor the map design and content to the needs and understanding of the target audience.
- **Consistency**: Ensure consistency in symbol usage, color schemes, and design elements across maps.

Map compilation in GIS is a critical process that transforms raw spatial data into informative and visually compelling maps. By following these steps and best practices, you can create effective maps that meet your analysis needs and communicate geographic information effectively.

5.6 Chart/Graphs

Charts and graphs in Geographic Information Systems (GIS) are essential for visualizing and analyzing spatial data beyond traditional maps. They help convey complex data patterns, trends, and relationships in a clear and interpretable manner. Here's a detailed look at how charts and graphs are used in GIS:

1. Types of Charts and Graphs in GIS

1.1. Bar Charts

- **Purpose**: Compare quantities across different categories or groups.
- **Use Cases**:
  - Displaying the number of features in different land use categories.
  - Showing the distribution of population by administrative regions.
- **Examples**: Comparing crime rates across neighborhoods, or land cover types by area.

1.2. Pie Charts

- **Purpose**: Show proportions or percentages of a whole.
- **Use Cases**:
  - Illustrating the composition of land use in a given area.
  - Representing the percentage distribution of different vegetation types in a region.
- **Examples**: Land use distribution in a city, or breakdown of different species in a habitat.

1.3. Line Graphs

- **Purpose**: Show trends or changes over time.
- **Use Cases**:

- Tracking changes in environmental conditions (e.g., temperature, pollution levels) over time.
- Analyzing temporal trends in population growth.

- **Examples**: Historical climate data, or seasonal changes in vegetation.

1.4. Scatter Plots

- **Purpose**: Display relationships between two variables.
- **Use Cases**:
  - Analyzing the correlation between two spatial variables, such as land price vs. proximity to amenities.
  - Studying patterns of environmental variables, like air quality versus distance from industrial areas.
- **Examples**: Relationship between elevation and temperature, or correlation between green space area and urban health indicators.

1.5. Histograms

- **Purpose**: Show the distribution of a single variable.
- **Use Cases**:
  - Displaying the distribution of elevation values within a specific area.
  - Analyzing the frequency of different land cover types.
- **Examples**: Elevation ranges in a study area, or distribution of building heights.

1.6. Heat Maps

- **Purpose**: Visualize the density or intensity of a variable across a geographic area.
- **Use Cases**:
  - Showing areas with high concentrations of incidents, such as crime or traffic accidents.
  - Representing density of certain species or land cover types.
- **Examples**: Crime hotspots in a city, or areas of high biodiversity.

2. Creating and Using Charts/Graphs in GIS

2.1. Data Preparation

- **Aggregation**: Summarize or aggregate spatial data to prepare it for charting (e.g., total population by district).
- **Classification**: Classify or categorize data to facilitate meaningful comparisons (e.g., land use types or income brackets).

2.2. Integration with Maps

- **Overlaying Charts**: Place charts on or alongside maps to provide context and enhance understanding (e.g., pie chart showing land use percentages on a map of a city).

- **Interactive Elements**: Incorporate interactive charts that allow users to explore data by clicking on different map features or regions.

2.3. Software Tools

- **ArcGIS**: Offers tools for creating various types of charts and graphs, including bar charts, pie charts, and histograms, through its Charting and Reporting tools.
- **QGIS**: Provides charting capabilities through plugins or built-in tools to create bar charts, pie charts, and other visualizations.
- **Tableau**: Can be used to create advanced visualizations and dashboards from GIS data by connecting to GIS data sources.
- **Microsoft Excel**: Useful for creating basic charts and graphs from GIS data exported into spreadsheets.

2.4. Best Practices

- **Clarity**: Ensure that charts and graphs are easy to understand, with clear labels, legends, and titles.
- **Relevance**: Choose the type of chart or graph that best represents the data and highlights the key insights.
- **Accuracy**: Verify that data is accurately represented and that charts are correctly scaled and labeled.
- **Interactivity**: Where possible, incorporate interactive features that allow users to explore and filter data dynamically.

3. Applications of Charts and Graphs in GIS

- **Urban Planning**: Visualizing and analyzing spatial data to support city planning, such as land use analysis or population density.
- **Environmental Monitoring**: Tracking changes in environmental conditions, like pollution levels or deforestation, through time-series graphs and heat maps.
- **Public Health**: Analyzing spatial patterns in health data, such as the incidence of diseases or access to healthcare facilities, using charts and graphs.
- **Transportation**: Evaluating traffic patterns, accident hotspots, or public transportation usage through various types of graphs and heat maps.

Charts and graphs are powerful tools in GIS for enhancing spatial data analysis and communication. By selecting the appropriate type of visualization and effectively integrating it with spatial data, GIS professionals can gain deeper insights and make more informed decisions.

5.7 Multimedia

Multimedia in Geographic Information Systems (GIS) refers to the integration and use of various forms of media—such as images, videos, audio, and interactive elements—in spatial data analysis and presentation. Multimedia enhances the richness of GIS outputs, making them more engaging, informative, and accessible. Here's a detailed overview of how multimedia is used in GIS:

1. Types of Multimedia in GIS

1.1. Images

- **Raster Images**: Include satellite imagery, aerial photographs, and scanned maps that provide visual context and detailed information about geographic areas.
  - **Applications**: Land cover classification, urban planning, environmental monitoring.
- **Static Images**: Maps with embedded photographs or illustrations that offer additional context or highlight specific features.
  - **Applications**: Annotated maps, educational materials, and historical comparisons.

1.2. Videos

- **Dynamic Visualization**: Videos that capture changes over time, such as temporal changes in land use or environmental conditions.
  - **Applications**: Monitoring deforestation, urban expansion, or seasonal changes in landscapes.
- **Educational Content**: Instructional videos explaining GIS concepts, tools, and techniques.
  - **Applications**: Training materials for GIS users or educational resources for students.

1.3. Audio

- **Voice Narration**: Audio clips that provide explanations or descriptions of spatial data and maps.
  - **Applications**: Interactive maps with voice instructions or guided tours.
- **Soundscapes**: Audio recordings representing environmental sounds or urban noise levels.
  - **Applications**: Acoustic studies, noise pollution mapping, and soundscape analysis.

1.4. Interactive Elements

- **Web-Based Maps**: Interactive maps that allow users to explore spatial data dynamically, zooming, panning, and querying features.
  - **Applications**: Online map services, interactive dashboards, and public engagement tools.
- **Geospatial Storytelling**: Combining maps with narrative text, images, and multimedia elements to tell a story or provide a comprehensive view of a geographic issue.
  - **Applications**: Story maps for community outreach, project presentations, and advocacy.

1.5. Augmented Reality (AR) and Virtual Reality (VR)

- **AR**: Overlaying spatial information onto the real world through mobile devices or AR glasses.
  - **Applications**: Navigation aids, immersive educational experiences, and field data collection.
- **VR**: Creating immersive virtual environments that simulate real-world locations or scenarios.

  - **Applications**: Urban planning simulations, virtual tours of historical sites, and disaster preparedness training.

2. Integrating Multimedia in GIS

2.1. Data Integration

- **Linking Multimedia**: Associating multimedia files with spatial features or attributes in GIS databases. For example, linking images of buildings or videos of environmental changes to specific locations.
- **Embedded Media**: Incorporating multimedia directly into GIS platforms, such as embedding videos or images in interactive maps or dashboards.

2.2. Tools and Software

- **ArcGIS**: Provides tools for integrating multimedia with spatial data, including StoryMaps for creating multimedia narratives and Web AppBuilder for interactive web applications.
- **QGIS**: Supports multimedia integration through plugins and customizations, such as linking external media files to spatial features.
- **Google Earth Engine**: Allows for the visualization of satellite imagery and video data over time.
- **GIS Cloud**: Offers web-based solutions for embedding multimedia content in GIS projects.

2.3. Creation and Editing

- **Image Editing**: Using software like Adobe Photoshop or GIMP to prepare and enhance raster images before integrating them into GIS.
- **Video Editing**: Editing video clips with tools like Adobe Premiere Pro or Final Cut Pro to produce time-lapse videos or instructional content.
- **Audio Editing**: Using audio editing software like Audacity to create or refine voice narrations or soundscapes.

3. Applications and Use Cases

3.1. Urban and Regional Planning

- **Visualizing Changes**: Using time-lapse videos or animated maps to show urban growth or redevelopment projects.
- **Public Engagement**: Creating interactive story maps or augmented reality experiences to engage the public in planning processes.

3.2. Environmental Monitoring

- **Change Detection**: Analyzing satellite imagery and video data to monitor environmental changes, such as deforestation or coastal erosion.
- **Educational Resources**: Developing multimedia content to raise awareness about environmental issues and conservation efforts.

3.3. Disaster Management

- **Response Coordination**: Using interactive maps with embedded videos and audio for real-time coordination during emergencies.
- **Training**: Providing VR simulations or instructional videos for disaster preparedness and response training.

3.4. Tourism and Education

- **Virtual Tours**: Creating immersive virtual tours of historical sites or natural landmarks for educational and tourist purposes.
- **Interactive Guides**: Developing interactive maps with multimedia elements to enhance visitor experiences at museums or parks.

4. Best Practices

- **Ensure Accessibility**: Make sure multimedia elements are accessible to all users, including those with disabilities, by providing alternative text for images, transcripts for audio, and subtitles for videos.
- **Maintain Quality**: Use high-quality multimedia content to ensure clarity and effectiveness in communication.
- **Optimize Performance**: Optimize multimedia files to balance quality and performance, avoiding large files that can slow down web applications or GIS software.
- **Provide Context**: Include descriptive text and explanations to provide context for multimedia elements and enhance understanding.

Integrating multimedia into GIS enhances the ability to communicate complex spatial information, engage audiences, and provide interactive and immersive experiences. By leveraging various forms of media, GIS professionals can create richer, more informative, and more engaging spatial data presentations.

5.8 Enterprise Vs. Desktop GIS

Enterprise GIS and Desktop GIS are two primary types of Geographic Information Systems (GIS) used for managing, analyzing, and visualizing spatial data. Each has distinct features, capabilities, and use cases that cater to different needs and organizational structures. Here's a detailed comparison of Enterprise GIS and Desktop GIS:

1. Overview

Enterprise GIS

- **Definition**: A comprehensive, scalable GIS solution designed to support the needs of an entire organization or enterprise. It typically involves a centralized GIS server and databases that can be accessed and utilized by multiple users across different departments or locations.

- **Components**: Includes GIS servers, enterprise databases, web-based applications, and tools for data management, analysis, and sharing.
- **Examples**: Esri ArcGIS Enterprise, GeoServer, MapServer.

Desktop GIS

- **Definition**: A GIS software application installed on a single workstation or desktop computer, primarily used by individual users or small teams for local data analysis and map creation.
- **Components**: Includes GIS software for data visualization, editing, and analysis, often with tools for creating maps and conducting spatial analysis on local datasets.
- **Examples**: Esri ArcGIS Desktop (ArcMap, ArcGIS Pro), QGIS, MapInfo Professional.

2. Key Features

Enterprise GIS

- **Scalability**: Designed to handle large volumes of data and support many users simultaneously.
- **Centralized Data Management**: Data is stored in centralized databases (e.g., relational databases) and managed from a central server, ensuring consistency and integrity.
- **Access Control**: Robust user management and security features to control access to data and functionalities based on user roles and permissions.
- **Integration**: Can integrate with other enterprise systems (e.g., CRM, ERP) and data sources (e.g., IoT sensors, external databases).
- **Collaboration**: Facilitates collaboration among multiple users and departments through shared data, web-based applications, and enterprise-wide access.
- **Customization and Extensions**: Allows for extensive customization and development of applications tailored to organizational needs.

Desktop GIS

- **Local Data Handling**: Data is typically managed and analyzed on the local machine, which can limit data sharing and collaboration.
- **Ease of Use**: Designed for individual users with a focus on user-friendly interfaces for map creation, data editing, and spatial analysis.
- **Limited Scalability**: Suitable for handling smaller datasets and single-user scenarios. Performance may be affected with large datasets or complex analyses.
- **Standalone Functionality**: Operates independently without needing a server or internet connection, though some functionalities may be enhanced with online capabilities.
- **Cost-Effective**: Generally more affordable compared to enterprise solutions, making it suitable for small businesses or individual users.

3. Deployment and Infrastructure

Enterprise GIS

- **Deployment**: Installed on servers within an organization's IT infrastructure or hosted on cloud platforms. It requires significant IT resources and infrastructure.
- **Maintenance**: Requires dedicated IT staff for server maintenance, data backups, and software updates. Typically includes support for high availability and disaster recovery.
- **Infrastructure**: Involves complex architecture including GIS servers, databases, web services, and network infrastructure.

Desktop GIS

- **Deployment**: Installed on individual desktop or laptop computers. Requires minimal IT infrastructure beyond the local machine.
- **Maintenance**: Maintained by the end-user or local IT staff. Updates and backups are managed individually.
- **Infrastructure**: Minimal infrastructure is required; usually operates with local storage and occasional access to external data sources or services.

4. Use Cases

Enterprise GIS

- **Large Organizations**: Suitable for large organizations with complex data needs, such as government agencies, utilities, and multinational corporations.
- **Collaborative Projects**: Ideal for projects that require collaboration across different departments or locations.
- **Complex Analyses**: Used for large-scale spatial analysis, real-time data processing, and integrating with other enterprise systems.

Desktop GIS

- **Individual Projects**: Ideal for personal or small team projects where in-depth spatial analysis and map creation are required.
- **Field Work**: Suitable for field data collection and analysis where a standalone tool is sufficient.
- **Educational and Training**: Often used in educational settings and for training purposes due to its ease of use and lower cost.

5. Licensing and Cost

Enterprise GIS

- **Licensing**: Typically involves enterprise-wide licensing agreements, which can be complex and expensive.
- **Cost**: Higher initial investment and ongoing costs due to infrastructure, licensing, and maintenance.

Desktop GIS

- **Licensing**: Often available as individual licenses or subscriptions. Some desktop GIS solutions offer free or open-source versions.
- **Cost**: Generally lower cost, making it accessible for individuals and small organizations.

6. Integration and Interoperability

Enterprise GIS

- **Integration**: Strong support for integration with various enterprise systems and data sources, including APIs, web services, and custom applications.
- **Interoperability**: Facilitates interoperability with different data formats and systems through standards like OGC (Open Geospatial Consortium) protocols.

Desktop GIS

- **Integration**: Limited to integration with local or networked data sources and services.
- **Interoperability**: Supports standard data formats but may have fewer capabilities for integrating with other systems compared to enterprise solutions.

Summary

Enterprise GIS is designed for large-scale, organization-wide applications that require extensive data management, collaboration, and integration capabilities. It is well-suited for organizations with complex data needs and the resources to support a centralized GIS infrastructure.

Desktop GIS is geared towards individual users or small teams, offering a more straightforward, cost-effective solution for local data analysis and map creation. It is ideal for scenarios where scalability and enterprise-level integration are not critical requirements.

Both types of GIS have their specific strengths and are chosen based on the scale of operations, data requirements, and available resources.

5.9 Distributed GIS

Distributed GIS refers to a geographic information system architecture where spatial data and processing capabilities are distributed across multiple locations or systems rather than being centralized in a single server or database. This distributed approach allows for more scalable, flexible, and efficient management of spatial data and GIS services. Here's an in-depth look at Distributed GIS:

1. Overview of Distributed GIS

Definition

- **Distributed GIS**: A system where GIS components (data storage, processing, and user interfaces) are spread across multiple computers or servers. These components work together to provide a cohesive GIS experience.

Key Characteristics

- **Decentralization**: Unlike centralized GIS systems where data and applications are managed on a single server or database, distributed GIS involves multiple interconnected systems that share responsibilities.
- **Interoperability**: Various GIS components and services interact through standard protocols and interfaces to function as a unified system.
- **Scalability**: The architecture can be scaled horizontally by adding more nodes or servers to handle increasing data loads or user demands.

2. Components of Distributed GIS

2.1. Data Sources

- **Distributed Databases**: Spatial data can be stored across multiple databases or file systems, often linked through networked connections.
- **Data Repositories**: Data may reside in various formats and locations, including cloud storage, local servers, or remote databases.

2.2. Servers

- **GIS Servers**: Servers that provide GIS services, such as map rendering, spatial analysis, and geocoding, across a network. Examples include ArcGIS Server, GeoServer, and MapServer.
- **Database Servers**: Servers that manage and query spatial databases, such as PostgreSQL with PostGIS, Oracle Spatial, or Microsoft SQL Server with spatial extensions.

2.3. Client Interfaces

- **Web Applications**: Users interact with GIS through web-based applications or services that access distributed data and services.
- **Desktop Applications**: GIS desktop software that can connect to distributed data sources and servers.
- **Mobile Applications**: Apps that access GIS data and services from mobile devices, often relying on distributed servers for real-time updates.

2.4. Middleware

- **Integration Tools**: Software that enables different GIS components to communicate and work together, often using standards like Web Map Service (WMS) or Web Feature Service (WFS).
- **APIs and Web Services**: Interfaces that allow different systems and applications to interact with distributed GIS components.

3. Advantages of Distributed GIS

3.1. Scalability

- **Load Distribution**: By distributing data and processing tasks across multiple servers, the system can handle larger volumes of data and more concurrent users.
- **Flexibility**: Easily add or remove servers to scale the system according to demand.

3.2. Performance

- **Reduced Bottlenecks**: Distributing tasks can alleviate performance bottlenecks associated with a single centralized system.
- **Geographical Distribution**: Data and services can be closer to users, reducing latency and improving response times.

3.3. Fault Tolerance

- **Redundancy**: Distributed systems can offer redundancy and backup capabilities, ensuring system reliability and availability.
- **Failover**: If one server or component fails, others can take over to maintain service continuity.

3.4. Data Integration

- **Diverse Data Sources**: Facilitate the integration of data from various sources and formats, providing a comprehensive view of spatial information.
- **Real-Time Data Access**: Allows real-time access and updates from different geographic locations.

4. Challenges of Distributed GIS

4.1. Complexity

- **System Management**: Managing a distributed GIS system can be complex, requiring coordination between multiple components and servers.
- **Data Synchronization**: Ensuring consistency and synchronization of data across different systems can be challenging.

4.2. Security

- **Data Protection**: Securing data across multiple servers and networks requires robust security measures to prevent unauthorized access and data breaches.
- **Access Control**: Implementing and managing access controls across a distributed system can be more complex than in a centralized system.

4.3. Integration

- **Interoperability Issues**: Ensuring that different GIS components and services can effectively communicate and work together may require additional integration tools and standards.

4.4. Latency and Network Dependence

- **Network Issues**: Performance can be impacted by network connectivity issues, especially if data and services are spread across geographically distant locations.

5. Examples of Distributed GIS Implementations

5.1. Web-Based GIS Platforms

- **ArcGIS Online**: Offers distributed services and data access through the cloud, integrating various data sources and providing scalable GIS functionalities.
- **Google Earth Engine**: Provides cloud-based access to vast amounts of satellite imagery and geospatial data.

5.2. Collaborative GIS Projects

- **OpenStreetMap**: A collaborative project where spatial data is contributed and maintained by a global community of users, distributed across various servers.
- **Global Disaster Alert and Coordination System (GDACS)**: Provides distributed access to real-time data and alerts for natural disasters.

5.3. Enterprise Systems

- **City Information Systems**: Large cities may use distributed GIS to manage and integrate data from various municipal departments and services, including transportation, utilities, and emergency services.

6. Best Practices

- **Standardization**: Use standard protocols and interfaces to ensure interoperability between different GIS components and services.
- **Security Measures**: Implement robust security practices to protect data and manage access controls effectively.
- **Regular Maintenance**: Perform regular system maintenance, backups, and updates to ensure system reliability and performance.
- **Monitoring and Management**: Use monitoring tools to oversee system performance and address issues promptly.

Distributed GIS systems offer significant benefits in terms of scalability, performance, and data integration, but they also come with challenges related to complexity, security, and integration. By understanding these factors and following best practices, organizations can effectively leverage distributed GIS to enhance their spatial data management and analysis capabilities.