lecture notes in

Geographic Information Systems

Introduced By

Prof.DR. Ahmed Aboul-fotouh Saleh

Professor of Information Systems

Faculty of Computer & Information Sciences

Information Systems Dept.

Index

Preface :

This book focuses on many of the terms and concepts which related with geographic information systems (GIS) . the terms and concepts introduced in this book have been compiled from a variety of different sources. These include some of the prevalent text books in the field, but most have been firmly implanted from practical experience.

The book is intended to provide resource beginner and specialists with an exposure to GIS technology and terminology, so that papers, applications, and exhibits on GIS will be more understandable and useful in the future. This book presents a substantial amount of GIS theory, and a whole bunch of buzzwords, but is also supplemented with relevant forestry based applications to illustrate the concepts and techniques for applying GIS technology.

In addition, The focus of this book is on practical issues concerned with the implementation and application of GIS technology. This book is directed at the resource specialist and researchers who has had limited exposure to GIS. It is intended purely as an introductory text with an emphasis on identifying and clearly illustrating primitive concepts in geographic information system applications.

Any comments or suggestions on the content or format of this book are welcome. Please send any comments to "elfetouh@gmail.com".

Chapter 1

Geographic Information Systems: An Overview

The short history of GIS (it goes back to the late 1960's) was founded in attempts in the UK, Canada and US to automate some of the land-management and census activities of government.

Figuring out how to do that, and how to explain what went wrong when they tried, was the start of the science of GIS. It was realized that many map-related concepts that seem so simple to us (scale, a boundary), required a lot of effort to teach to a computer.

1.1 WHAT IS GIS ?

Geographical Information Systems(GIS) *are computer-based systems that enable users to collect, store, process, analyze and present spatial data.*

It provides an electronic representation of information, called spatial data, about the Earth's natural and man-made features. A GIS references these real-world spatial data elements to a coordinate system. These features can be separated into different layers. A GIS system stores each category of information in a separate "layer" for ease of maintenance, analysis, and visualization. For example, layers can represent terrain characteristics, census data, demographics information, environmental and ecological data, roads, land use, river drainage and flood plains, and rare wildlife habitats. Different applications create and use different layers. A GIS can also store attribute data, which is descriptive information of the map features.

This attribute information is placed in a database separate from the graphics data but is linked to them. A GIS allows the examination of both spatial and attribute data at the same time. Also, a GIS lets users search the attribute data and relate it to the spatial data.

Therefore, a GIS can combine geographic and other types of data to generate maps and reports, enabling users to collect, manage, and interpret location-based information in a planned and systematic way. In short, a GIS can be defined as a computer system capable of

assembling, storing, manipulating, and displaying geographically referenced information.

GIS systems are dynamic and permit rapid updating, analysis, and display. They use data from many diverse sources such as satellite imagery, aerial photos, maps, ground surveys, and global positioning systems (GPS).

In general, a GIS provides facilities for data capture, data management, data manipulation and analysis, and the presentation of results in both graphic and report form, with a particular emphasis upon preserving and utilizing inherent characteristics of spatial data.

The ability to incorporate spatial data, manage it, analyze it, and answer spatial questions is the distinctive characteristic of geographic information systems.

A geographic information system, commonly referred to as a GIS, is an integrated set of hardware and software tools used for the manipulation and management of digital spatial (geographic) and related attribute data.

OR

"a Computer tool for managing geographic feature location data and data related to those features."

Fig 1.1 GIS Concept

The power of GIS is the ability to combine geospatial information in unique ways—by layers or themes—and extract something new. For instance, a GIS analysis might include the location of a highway intersection and the average number of vehicles that flow through the intersection throughout the day, and extract information useful for locating a business. GIS might include both the location of a river and the water depth along its course by season, and enable an analysis of the effects of development on runoff within the watershed. Overlaying the path of a severe thunderstorm with geospatial data on the types of structures encountered—homes, stores, schools, post offices—could inform an analysis of what types of building construction can survive high winds and hail.

1.2 GIS SUBSYSTEMS :

A GIS comprises of **four main functional subsystems. These are:**

a data input subsystem;

a data storage and retrieval subsystem;

a data manipulation and analysis subsystem; and a data output and display subsystem.

1.2.1 Data Input

A data input subsystem allows the user to capture, collect, and transform spatial and thematic data into digital form. The data inputs are usually derived from a combination of hard copy maps, aerial photographs, remotely sensed images, reports, survey documents, etc.

1.2.1.1 GIS Data Types

These data will contain maps of different detail levels (maps of the county, its main cities and villages, maps of the archaeological and historical sites etc.), photos of places and monuments, video images, text (in many languages), music and sound. For more complex applications, multimedia data can be remotely sensed imagery, scanned maps, digitized video clips, DTMs, one or more dimensional measurements, simulation model outputs and others. Most of them are complicated objects, which have large data volumes, intensive processing requirements and rich semantics.

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

(1) Spatial data

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

(2) Attribute data

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.

The coordinate location of a forestry stand would be spatial data, while the characteristics of that forestry stand, e.g. cover group, dominant species, crown closure, height, etc., would be attribute data. Other data types, in particular image and multimedia data, are becoming 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, narrations, etc.

1.2.1.2 Sources of Data

A wide variety of data sources exist for both spatial and attribute data. The most common general sources for spatial data are:

- Hard copy maps
- Aerial photographs
- Remotely-sensed imagery
- Point data samples from surveys
- Existing digital data files

This spatial data is usually in analog form and needs to be converted to digital form before it can be used. Maps can be digitized, or hand traced with at computer mouse, to collect the coordinates of features. Attribute data has an even wider variety of data sources. Any textual or tabular data than can be referenced to a geographic feature, e.g. a point, line, or area, can be input into a GIS.

Attribute data is usually input by manual keying or via a bulk loading utility of the DBMS software.

1.2.2 Data Storage and Retrieval

The data storage and retrieval subsystem organizes the data, spatial and attribute, in a form which permits it to be quickly retrieved by the user for analysis, and permits rapid and accurate updates to be made to the database. This component usually involves use of a database management system (DBMS) for maintaining attribute data. Spatial data is usually encoded and maintained in a proprietary file format.

1.2.2.1 Organizing Data for Analysis

Most GIS software organizes spatial data in a thematic approach that categorizes data in vertical layers. The definition of layers is fully dependent on the organization.s requirements.

Typical layers used in natural resource management agencies or companies include forest cover, soil classification, elevation, road network (access), ecological areas, hydrology, etc.

1.2.2.2 Editing and Updating of Data

Perhaps the primary function in the data storage and retrieval subsystem involves the
editing and updating of data. Frequently, the following data editing capabilities are required:

- Interactive editing of spatial data
- Interactive editing of attribute data
- The ability to add, manipulate, modify, and delete both spatial features and attributes (independently or simultaneously)
- Ability to edit selected features in a batch-processing mode.

1.2.2.3 Data Retrieval and Querying

The ability to retrieve data is based on the unique structure of the DBMS and command interfaces are commonly provided with the software. Most GIS software also provides a programming ubroutine library, or macro language, so the user can write their own specific data retrieval routines if required.
Querying is the capability to retrieve data, usually a data subset, based on some userdefined formula. These data subsets are often referred to as logical views. Often the querying is closely linked to the data manipulation and analysis subsystem. Querying can be either by example of by content.

1.2.3 Data Manipulation and Analysis

The data manipulation and analysis subsystem allows the user to define and execute spatial and attribute procedures to generate derived information. This subsystem is commonly thought of as the heart of a GIS, and usually distinguishes it from other database information systems and computer-aided drafting (CAD) systems.

1.2.3.1 Manipulation and Transformations of Spatial Data

The maintenance and transformation of spatial data concerns the ability to input, manipulate, and transform data once it has been created. Some specific functions are:

- Coordinate thinning: involves the reduction of the coordinate pairs (X and Y) from arcs.
- Geometric Transformations
- Map Projection Transformations
- Edge Matching
- Interactive Graphic Editing

1.2.3.2 Analytical Functions in a GIS

The primitive analytical functions that must be provided by any GIS are:

- Retrieval, Reclassification, and Generalization
- Topological Overlay Techniques
- Transformation , Dissolve, Buffers....
- Connectivity Functions

1.2.4 Data Output and Display

The data output subsystem allows the user to generate graphic displays, normally maps, and tabular reports representing derived information products.

The critical function for a GIS is, by design, the analysis of spatial data. It is important to understand that the GIS is not a new invention. In fact, geographic information processing has a rich history in a variety of disciplines. In particular, natural resource specialists and environmental scientists have been actively processing geographic data and promoting their techniques since the 1960's.

Today's generic, geographic information system, is distinguished from the geo-processing of the past by the use of computer automation to integrate geographic data processing tools in a friendly and comprehensive environment. The advent of sophisticated computer techniques has proliferated the multi-disciplinary application of geo-processing methodologies, and provided data integration capabilities that were logistically impossible before.

1.3 THE COMPONENTS OF GIS :

Definitions of a geographic information system can vary considerably. The definition provided here combines both the components and functions of a GIS.

An operational GIS has a series of components that combine to make the system work. These components are critical to a successful GIS.

A working GIS integrates five key components: HARDWARE, SOFTWARE, DATA, PEOPLE, METHODS.

Fig. 1.2 GIS Components

(1) GIS Hardware

Hardware is the computer system on which a GIS operates. Today, GIS software runs on a wide range of hardware types, from centralized computer servers to desktop computers used in stand-alone or networked configurations.

The selection of hardware will vary according to the application and data requirements.

Data	Hardware

Entry (Input)	• Digitizing tablet • CCD digitizing camera • Scanner • Tape drive
Output Hardcopy (maps, lists. Tables) Softcopy Electronic	 • Printers(laser, dot-matrix,ink-jet) • Plotters (pen, electro-static) • Colour graphics monitor • Tape drive • Optical disks • Optical disk drives • Telecommunication media.

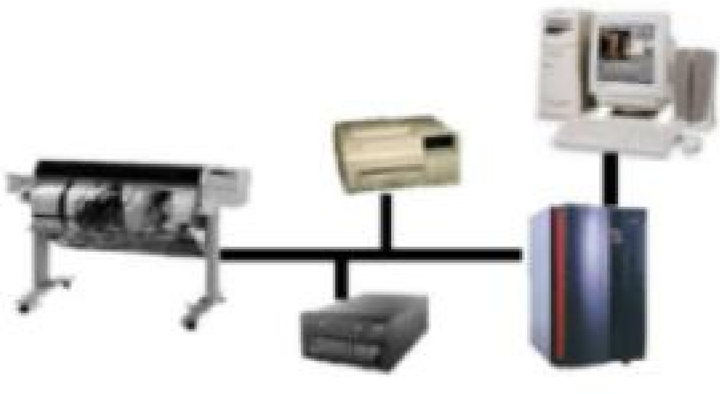


Fig 1.3 Computer Hardware & Software

(2) Software

GIS software provides the functions and tools needed to store, analyze, and display geographic information. This element includes not only actual GIS software, but also various database, drawing, statistical, imaging, or other software.

GIS software typically consist of three components which be explan in the chapter 3 :

- The Graphics database

- The attribute database

- The user Interface

The next figures illustrate some examples of GIS Software:

Examples of Software

- IDRISI (www.clarklabs.org)
- MapInfo (www.pbinsight.com/welcome/mapinfo/)
- GRASS (grass.fbk.eu)
- ArcGIS (www.esri.com/software/arcgis)

ArcGIS

- Latest version: 10.0
- Three levels of license: ArcInfo, ArcEditor and ArcView
- Applications:
  - ✓ArcMap- used to create maps, view, edit, and analyze spatial data.
  - ✓ArcScene- allows you to overlay many layers of data in a 3D environment
  - ✓ArcTool box- has tools for geoprocessing, data conversion, and defining and changing map projections
  - ✓ArcCatalog-used to manage and organize GIS data, preview datasets, view and manage metadata.

The storage, manipulation and reporting of locational information is handled by the graphics database, while descriptive data are handled by the attribute database. These two databases are independent of one another, however they are linked so that both sets of information are always available to the user.

The user interface provides access to the capabilities and functions each particular GIS offers.

The software provided to manipulate the graphics database in conjunction with the attribute database, is the distinguishing factors between different GIS software products. All GIS packages provide basic functions of capture, storage, manipulation, display, and output, but not all systems will necessarily provide the mix of capabilities and type of interface required by a particular application or group of users.

Fig 1.4 ArcGIS Software Package

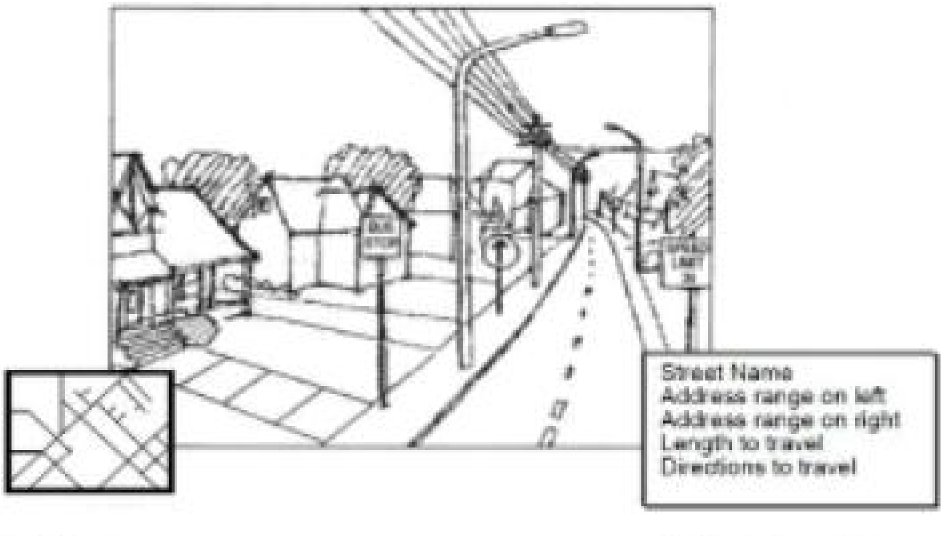
(3) Data

Perhaps the most important component of a GIS is the data. Geographic data and related tabular data can be collected in-house, compiled to custom specifications and requirements, or

occasionally purchased from a commercial data provider. A GIS can integrate spatial data with other existing data resources, often stored in a corporate DBMS. The integration of spatial (often proprietary to the GIS software), and tabular data stored in a DBMS is a key functionality afforded by GIS.

3/1 Geography Information System Data :

- **Spatial Data :** Represents features that have a known location on earth.
- **Attribute Data** : The information linked to the geographic features (spatial data) that describe those features.
- **Data Layers :** Are the result of combining spatial and attribute data. Essentially adding the attribute database to the spatial location.
- **Layer Types :** A layer type refers to the way spatial and attribute information are connected. There are two major layer types, vector and raster.
- **Topology** : How geographic features are related to one another and where they are in relation to one another.



Spatial Data **Attribute Data**

Fig 1.5

Geographic features are represented by two types of data.

(4) Users

This is the most important component in a GIS. People must develop the procedures and define the tasks of the GIS. GIS users 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 specialists versus end users is often critical to the proper implementation of GIS technology.

(5) Procedures (or Methods)

A successful GIS operates according to a well-designed implementation plan and business rules, which are the models and operating practices unique to each organization.

As in all organizations dealing with sophisticated technology, new tools can only be used effectively if they are properly integrated into the entire business strategy and operation. To

do this properly requires not only the necessary investments in hardware and software, but also in the retraining and/or hiring of personnel to utilize the new technology in the proper organizational context. Failure to implement your GIS without regard for a proper organizational commitment will result in an unsuccessful system ! Many of the issues concerned with organizational commitment are described in Implementation Issues and Strategies.

1.4 GIS FUNCTIONS :

According to the above , the basic functions of a GIS , in terms of practically are: Capturing, storing, Analyzing, Quering and Dsiplaying

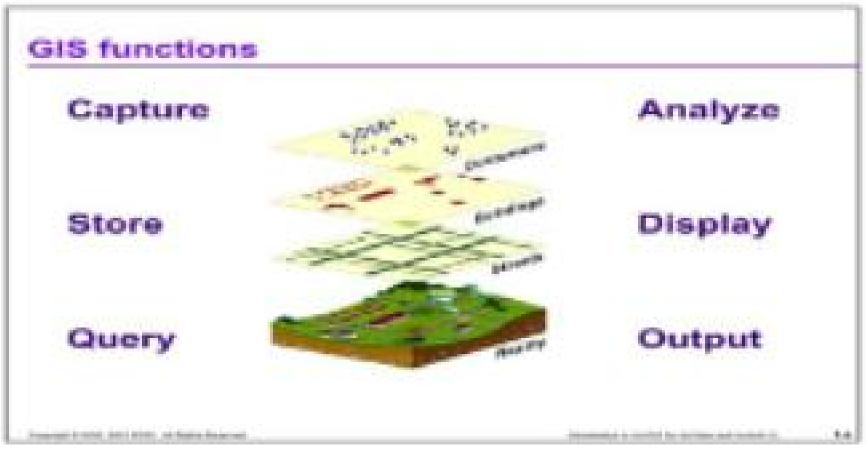


Fig 1.6 GIS Functions

1.4.1 Capturing data

A GIS must provide methods for inputting geographic (coordinate) and tabular (attribute) data. The more input methods available, the more versatile the GIS.

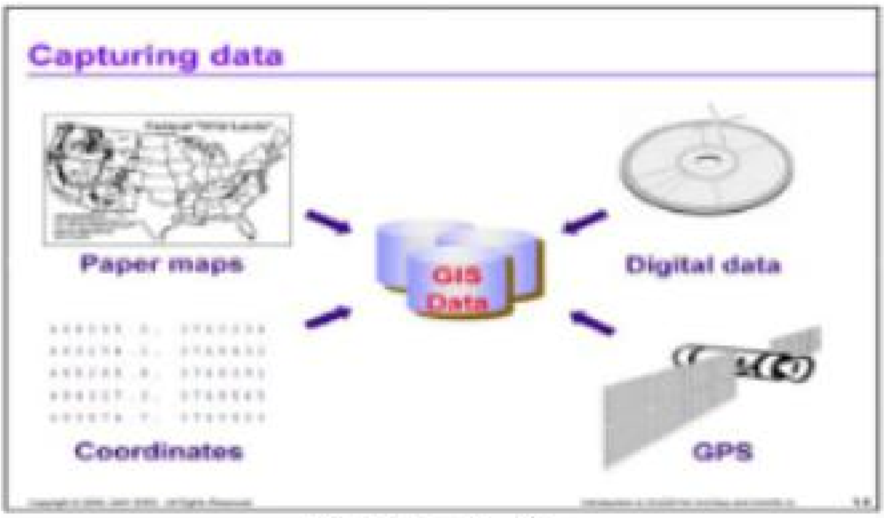


Fig 1.7 Capturing Data

The geographic database is the expensive and long-lived component of the GIS, thus making data entry an important consideration. Because ArcGIS integrates a variety of data types from a variety of sources, it provides multiple data entry options. ArcGIS offers efficient data entry methods for automating paper maps and other non-digital data sources. To take advantage of the vast collection of geographically referenced data that already exists in digital format, ArcGIS provides the most comprehensive data conversion capability of any GIS on the market. ArcGIS software's integrative capabilities also allow data sharing with other applications without the need for conversion.

1.4.2 Storing data

There are two basic data models for geographic data storage: vector and raster. A GIS should be able to store geographic data in both.

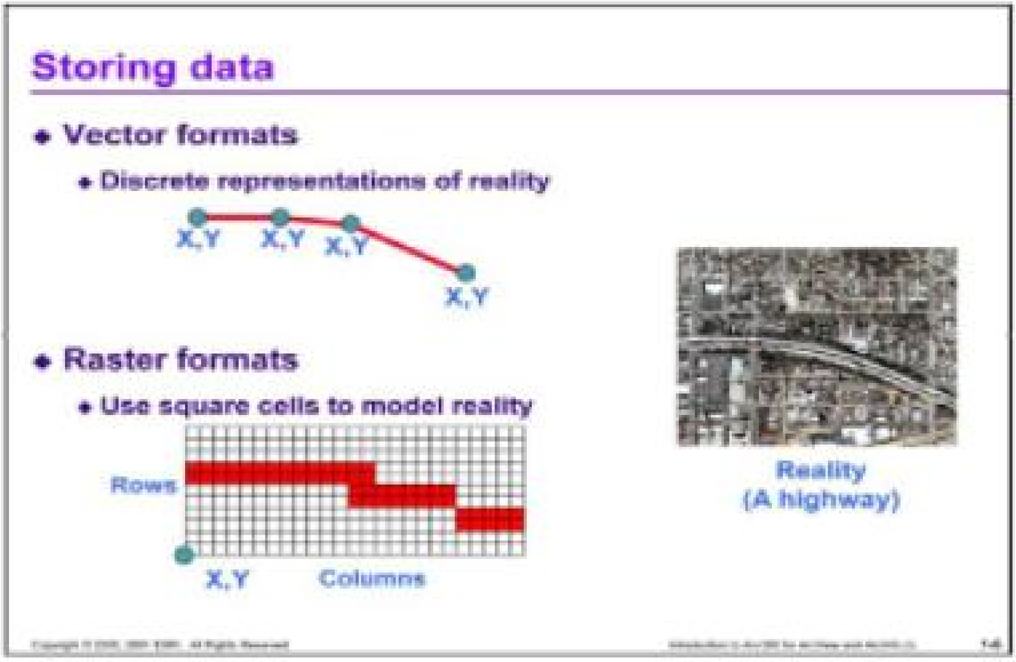


Fig 1.8 Storing Data of GIS

Vector data

The vector data model represents geographic features similarly to the way maps do—using points, lines, and areas. An x,y (Cartesian) coordinate system references real-world locations.

Raster data

Instead of representing features by their explicit x,y coordinates, p g y p ,, the raster data model assigns values to cells that cover their locations. Raster format is well suited to spatial analysis and is also appropriate for the storage of data that is collected in grid format. The amount of detail you can show for a particular feature depends on the size of the cells in the grid. This makes raster data inappropriate for applications where discrete boundaries must be known, such as parcel management.

1.4.3 Querying data

A GIS must provide utilities for finding specific features based on their locations or attribute values.

- **Identifying specific features**

  One common type of GIS query is determining what exists at a particular location. In this type of query, the user understands where the features of interest are, but wants to know what characteristics are associated with them. This can be accomplished with GIS because the spatial features are linked to the descriptive characteristics.

- **Identifying features based on conditions**

  Another type of GIS query is to determine which location or locations satisfy certain conditions. In this case the user knows what characteristics are important, and wants to find out where the features are that have those characteristics.

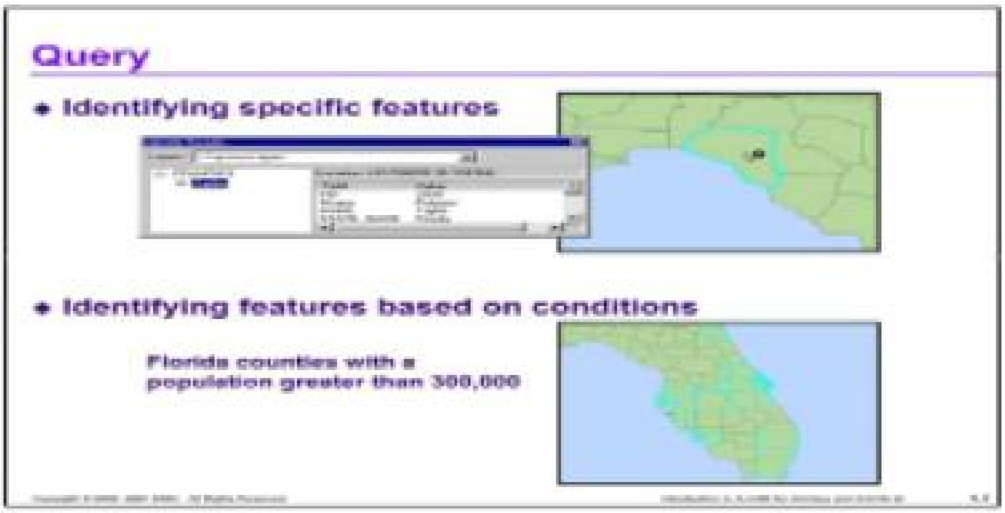


Fig 1.9 Query

1.4.4 Analyzing data

A GIS must have the ability to answer questions regarding the interaction of spatial relationships between multiple datasets.

Fig 1.10 Analysis

You can perform analysis to obtain the answers to a particular question or solutions to a particular problem.

Geographic analysis usually involves more than one geographic dataset and requires the analyst to proceed through a series of steps to reach a final result. Three common types of geographic analysis are:

Proximity analysis

- How many houses lie within 100 m of this water main?
- What is the total number of customers within 10 km of this store?
- What proportion of the alfalfa crop is within 500 m of the well?

To answer such questions, GIS technology uses a process called buffering to determine the proximity relationship between features.

Overlay analysis

The integration of different data layers involves a process called overlay. At its simplest, this could be a visual operation, but analytical operations require one or more data layers to be joined physically. This overlay, or spatial join, can integrate data on soils, slope, and vegetation or land ownership with tax assessment.

Network analysis

This type of analysis examines how linear features are connected and how easily resources can flow through them.

Displaying data

There must be tools for visualizing the geographic features using a variety of symbology.

For many types of geographic operations the end result is best visualized as a map or graph. Maps are very efficient at storing and communicating geographic information. While cartographers have created maps for millennia, GIS provides new and exciting tools to extend the art and science of cartography. Map displays can be integrated with reports, three-dimensional views, photographic images, and other output such as multimedia.

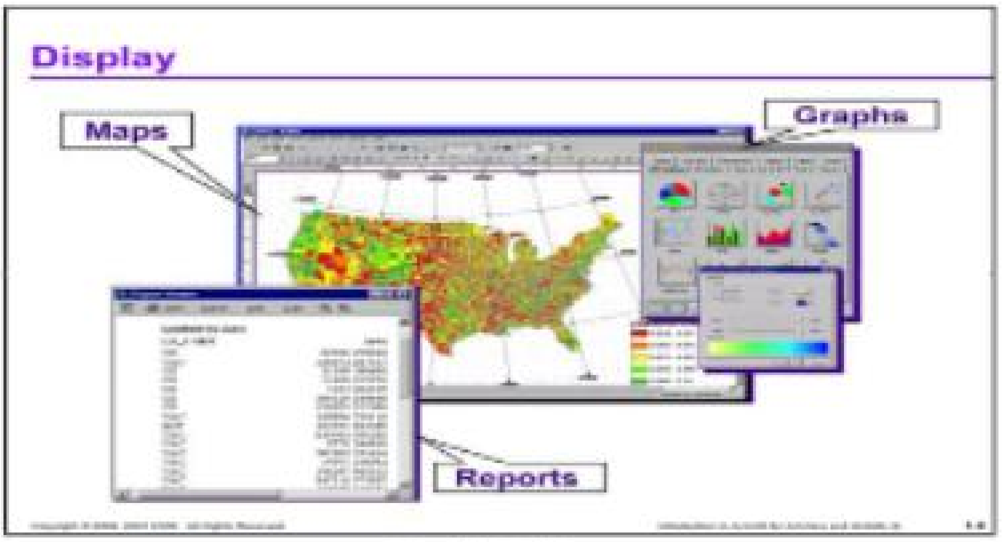


Fig 1.11 Display

Output

Results of display should be able to be output in a variety of formats such as maps, reports, and graphs. The more avenues for output a GIS can offer, the greater the potential for reaching the right audience with the right information.

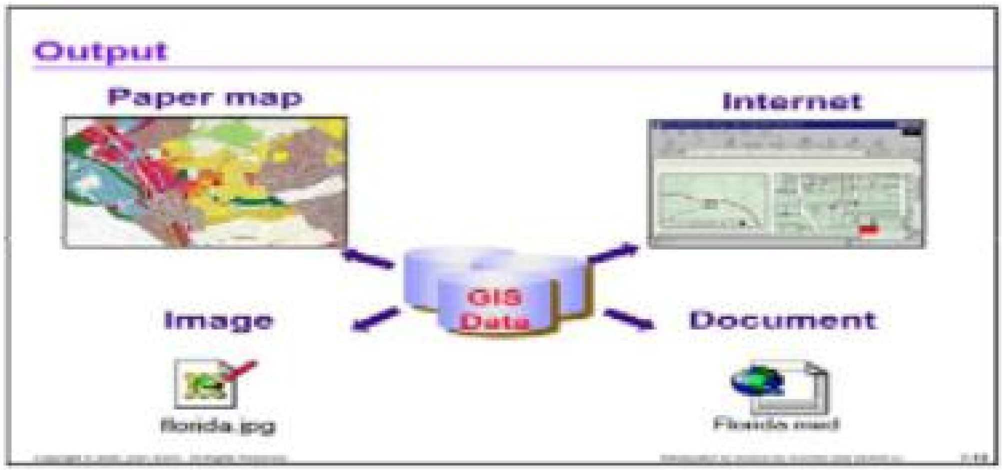


Fig 1.12

1.5 APPLICATION OF GIS :

1/5/1 Education

Education is a field where integration of multimedia and GIS can bring enormous benefits. Students will learn faster and more efficiently. In addition, it will be possible to individualize learning and tune it to particular preferences of each student. In this model a teacher becomes a guide rather than a repository of facts. It is the computer that takes on a role of "an infinitely patient teacher."

1/5/2 Mapmaking

GIS can use and combine all layers that are available for an area, in order to produce anoverlay that can be analyzed by using the same GIS. Such overlays and their analysis radically change decision-making process that include, among others:

Site selection
Simulation of environmental effects (for example, creating perspective views of a terrain before and after mining)
Emergency response planning (for example, combining road network and earth science information to analyze the effects of a potential earthquake)

1/5/3 Land Information

GIS has aided management of land information by enabling easy creation and maintenance of data for land records, land planning and land use. GIS makes input, updates, and retrieval of data such as tax records, land-use plan, and zoning codes much easier then during the paper-map era. Typical uses of GIS in land information management include managing land registry for recording titles to land holdings, preparing land-use plan and zoning maps, cadastral mapping etc. Input of data into a land information GIS includes: political and administrative boundaries, transportation, and soil cover.

1/5/4 Infrastructure and Utilities

GIS technologies are also widely applied to the planning and management of public utilities. Typical uses include management of the following services: electric, gas, water, roads, telecommunication, storm sewers, TV/FM transmitting facilities, hazards analysis, and dispatch and emergency services.

Typical data input includes street network, topographic data, demographic data and local government administration boundary.

1/5/5 Environmental

The environmental field has long used GIS for a variety of applications that range from simple inventory and query, to map analysis and overlay, to complex spatial decision-making systems. Examples include: forest modeling, air/water quality modeling and monitoring, environmentally sensitive zone mapping, analysis of interaction between economic, meteorological, and hydrological & geological change. Typical data input into an environmental GIS include: elevation, forest cover, soil quality and hydrogeology coverage.

1/5/6 Archaeology

Archaeology, as a spatial discipline, has used GIS in a variety of ways. At the simplest level, GIS has found applications as database management for archaeological records, with the added benefit of being able to create instant maps. It has been implemented in cultural resource management contexts, where archaeological site locations are predicted using statistical models based on previously identified site locations. It has also been used to simulate diachronic changes in past landscapes, and as a tool in intra-site analysis.

1/5/7 Natural Hazards

Areas vulnerable to earthquakes, floods, cyclones, storms, drought, fire, volcano, land slides, soil erosion can be used to accurately predict future disasters.

1/5/8 Forestry

GIS has been emerging as a strong tool for many areas of forestry, from harvesting schedules to urban forestry.

1/5/9 Military GIS

GIS offers a virtually unique ability to aggregate, automate, integrate and analyze. geographical data, which further enhance the intelligence base for defense operations.

1/5/10 Oceanography

GIS enables study of sea level change, marine population, sea surface temperature, and coral reef ecosystem.

1/5/11 Water Resources

GIS enables spatial representation of ground water resources, waste quality, watershed management, surface water management, and water pollution.

1/5/12 GIS in agriculture and soil

Data includes information on the country's land resources including physiography, soils, climate, hydrology, cropping systems and crop suitability.

Chapter 2

Introduction to Map Elements

Maps are the integral components of a GIS whether they represent an input to, or product of, the system.

2.1 WHAT IS a MAP ?

according to the International Cartographic Association, a map is:

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

Maps show more than the Earth's surface

- the term "map" is often used in mathematics to convey the notion of transferring information from one form to another, just as cartographers transfer information from the surface of the Earth to a sheet of paper
- the term "map" is used loosely to refer to any visual display of information, particularly if it is abstract, generalized or schematic

Cartographic abstraction

production of a map requires:

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

2.2 MAPS AND SPATIAL INFORMATION

The main method of identifying and representing the location of geographic features on the landscape is a map. A map is a graphic representation of where features are, explicitly and relative to one another. A map is composed of different geographic features represented as either points, lines, and/or areas. Each feature is defined both by its location in space (with reference to a coordinate system), and by its characteristics (typically referred to as attributes). Quite simply, a map is a model of the real world.

The map legend is the key linking the attributes to the geographic features. Attributes, e.g. such as the species for a forest stand, are typically represented graphically by use of different symbology and/or color. For GIS, attributes need to be coded in a form in which they can be used for data analysis (Burrough, 1986). This implies loading the attribute data into a database system and linking it to the graphic features.

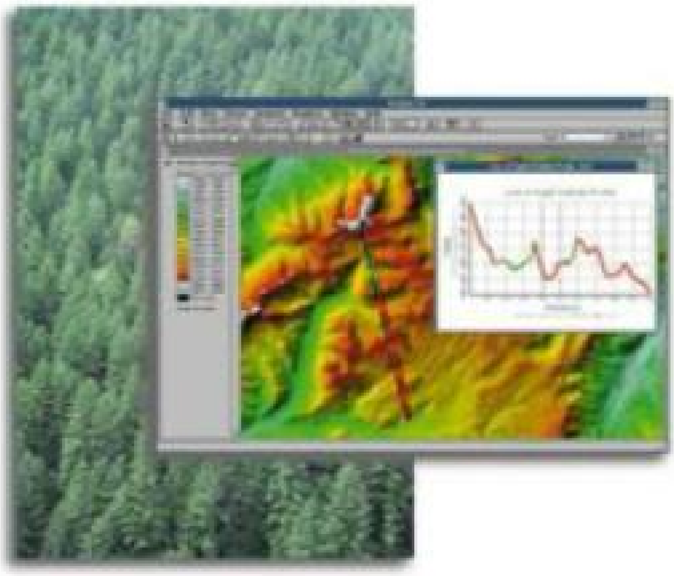


Fig 2.1

Maps are simply models of the real world. They represent snapshots of the land at a specific map scale. The map legend is the key identifying which features are represented on a map.

For geographic data, often referred to as spatial data, features are usually referenced in a coordinate system that models a location on the earth's surface. The coordinate system may be of a variety of types.

Maps are the traditional method of storing and displaying geographic information.

A map portrays 3 kinds of information about geographic features. The:

- **Location and extent of the feature**
- **Attributes (characteristics) of the feature**
- **Relationship of the feature to other features.**

Geography has often been described as the study of why what is where. This description is quite appropriate when considering the three kinds of information that are portrayed by the traditional map;

- the location and extent of a feature is identified explicitly by reference to a coordinate system representing the earth's surface. This is where a feature is.
- the attributes of a feature describe or characterize the feature. This is what the feature is.
- The relationship of a feature to other features is implied from the location and attributes of all features. Relationships can be defined explicitly, e.g. roads connecting towns, regions adjacent to one another, or implicitly, e.g. close to, far from, similar to, etc. Implicit relationships are interpreted according to the knowledge that we have about the natural world. Relationships are described as how or why a feature is.

2.3 TYPES OF MAPS :

- in practice we normally think of two types of map:
- ***topographic map*** - a reference tool, showing the outlines of selected natural and man-made features of the Earth
  - often acts as a frame for other information
  - "Topography" refers to the shape of the surface, represented by contours and/or shading, but topographic maps also show roads and other prominent features
- ***thematic map*** - a tool to communicate geographical concepts such as the distribution of population densities, climate, movement of goods, land use etc.

topographic map

thematic map

Fig 2.2 topographic map & thematic map

Thematic maps in GIS :

- several types of thematic map are important in GIS:

- **<u>a choropleth map</u>** uses reporting zones such as counties or census tracts to show data such as average incomes, percent female, or rates of mortality
  - the boundaries of the zones are established independently of the data, and may be used to report many different sets of data
- **<u>an area class map</u>** shows zones of constant attributes, such as vegetation, soil type, or forest species
  - the boundaries are different for each map as they are determined by the variation of the attribute being mapped, e.g. breaks of soil type may occur independently of breaks of vegetation
- **<u>an isopleths map</u>** shows an imaginary surface by means of lines joining points of equal value, "isolines" (e.g. contours on a topographic map)
  - used for phenomena which vary smoothly across the map, such as temperature, pressure, rainfall or population density

Line maps versus photo maps :

- an important distinction for GIS is between a line map and a photo map
- a line map shows features by conventional symbols or by boundaries
- a photo map is derived from a photographic image taken from the air
    - features are interpreted by the eye as it views the map
    - certain features may be identified by overprinting labels
    - photomaps are relatively cheap to make but are rarely completely free of distortions

Fig 2.3 Photo Map

2.4 Characteristics of maps :

- maps are often stylized, generalized or abstracted, requiring careful interpretation
- usually out of date
- show only a static situation - one slice in time

- often highly elegant/artistic
- easy to use to answer certain types of questions:
  - how do I get there from here?
  - what is at this point?
- difficult or time-consuming to answer other types:
  - what is the area of this lake?
  - what places can I see from this TV tower?
  - what does that thematic map show at the point I'm interested in on this topographic map?

2.5 The CONCEPT OF SCALE :

- **the scale of a map is the ratio between distances on the map and corresponding distances in the real world**
  - if a map has a scale of 1:50,000, then 1 cm on the map equals 50,000 cm or 0.5 km on the Earth's surface
- **the use of the terms "small scale" and "large scale" is often confused, so it is important to be consistent**
  - **a large scale map shows great detail, small features**
    - representative fraction is large, e.g. 1/10,000
  - **a small scale map shows only large features**
    - representative fraction is small, e.g. 1/250,000
- **the scale controls not only how features are shown, but what features are shown**
  - a 1:2,500 map will show individual houses and lamp posts while a 1:100,000 will not
- **different scales are used in different countries**

- in the US, 1:100,000 is the largest scale at which complete coverage of the continental states exists, but there is limited coverage at 1:62,500 and 1:24,000
- in the UK, there is complete coverage at much larger scales (1:1,250 to 1:10,000)

2.6 MAP PROJECTION :

- the Earth's surface is curved but as it must be shown on a flat sheet, some distortion is inevitable
  - distortion is least for when the map only shows small areas, and greatest when a map attempts to show the entire surface of the Earth
- a projection is a method by which the curved surface of the earth is represented on a flat surface
  - it involves the use of mathematical transformations between the location of places on the earth and their projected locations on the plane
- numerous projections have been invented, and arguments continue about which is best for which purposes
- projections can be identified by the distortions which they avoid - in general a projection can belong to only one of these classes:
  - equal area projections preserve the area of features by assigning them an area on the map which is proportional to their area on the earth - these are useful for applications which require measuring area, and are popular in GIS

– conformal projections preserve the shape of small features, and show directions (bearings) correctly - they are useful for navigation

– equidistant projections preserve distances to places from one or two points

2.7 CARTOGRAPHY & DIGITIZING :

2.7.1 Basic Concepts of Cartography

- Cartography is the art and science of map making.
- Communication is the traditional objective.
- Analysis has become an important objective with the development of GIS.

2.7.2 The Importance of Maps

- - To record and store information
- - To analyze locational distributions and spatial patterns
- - To present information and communicate findings

2.7.3 Basic Elements of Map Composition

***A) Map Scale*:** Map scale defines the amount of reduction of reality. Scale defines the precision of the location and the level of detail Be care when using small scale maps as input and then enlarging it is always better to reduce a map after analysis than to enlarge it for analysis.

Scale is expressed in three primary ways:

1. Verbal Scale
2. Representative fraction (RF)
3. Graphic scale (bar)

1. Verbal Scale

Map scale is expressed as ordinary text words

For example : 1 centimeter equals (represents) 1 meter .

2. Representative fraction (RF)

Map scale is expressed as a ratio in the same units.

For example : 1:2,000 means that one inch (or one meter) on the map represents 2,000 inches (or meters) on the ground.

3. Graphic Bar

The graphic bar places visual measure of ground distances on the map. Used on printed maps (output of GIS) to aid in communicating the scale. Most software can automatically generate a graphic scale.

Ex:

Remains accurate after mechanical enlargement of map, printed ratio or printed scale will be wrong after "zooming" the page on the copy machine.

Map Scale: Small vs. Large

Small scale refers to the RF ratio. A 1:250,000 scale is small compared to a 1:2,000.The ratio is small and the amount of reduction is large, producing a map of a large area.

Large scale means less reduction and a map covering a small area.

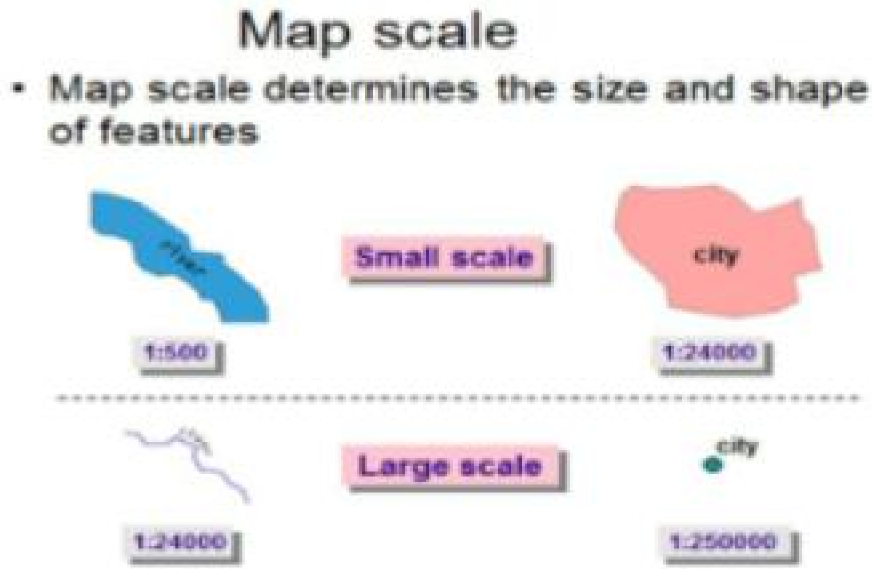


Fig 2.4 Map Scale

B) Legend

1. The reference area on a map that lists the colors, symbols, line patterns, shadings and other annotation used on the map, and their meanings. The legend often includes the map's title, scale, origin, orientation and other information.
2. The symbol key on a map used to describe a map's symbols and how they are interpreted.

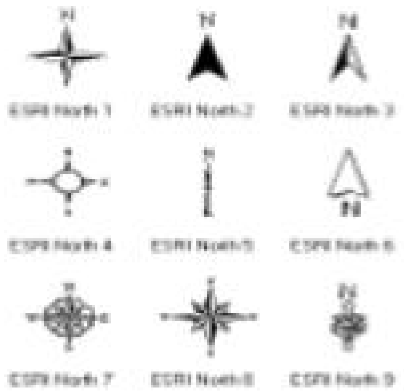
C. Direction

The question of **what is north** can be an issue on some maps.
On the earth, ***true north*** (the direction to the North Pole) differs from magnetic north, and the ***magnetic north*** pole moves due to changing geophysical conditions of the earth's crust and core.

Many reference maps indicate both. Most maps we compose are oriented to true north, even though compass readings in the field are angled to the magnetic pole. Adjustments for these compass deviations are made routinely.

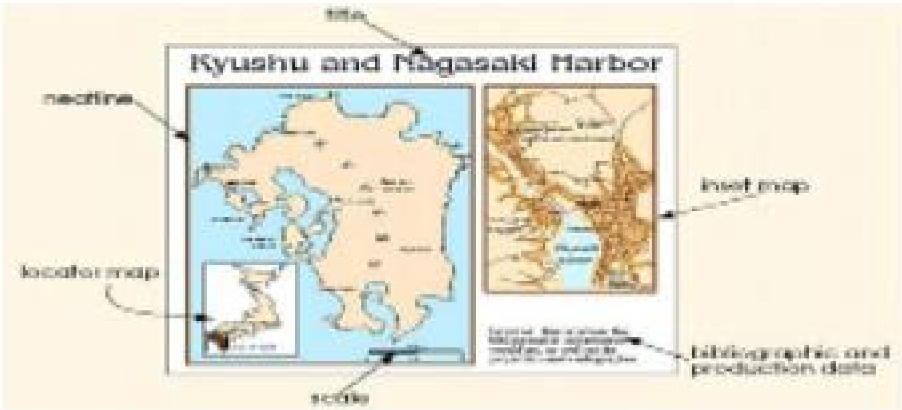
D. Sources of information and how processed

Unless it is absolutely clear from the context in which a map appears, readers will need to know about the sources from which the map was derived. Often the age, accuracy, and reliability of sources is critical to the interpretation of a map and should be noted.



E. Title

The title of a map is usually one of its most essential features. As such, it should receive very careful attention so as to match the needs of the theme and audience. The content of the title should also be measured against other lettering applied to the map, for example in the legend or annotations.



F. Projection

The projection used to create a map influences the representation of area, distance, direction, and shape. It should be noted when these characteristics are of prime importance to the interpretation of the map.

G. Cartographer

The authority lying behind the composition of a map can be of prime importance in some situations. Most maps note the name, initials, or corporate identity of the cartographer(s).

H. Date of production

The meaning and value of some maps--such as those relating to current affairs or weather--are time sensitive. The reader must know when they were produced to estimate whether to trust them or not. An out-of- date road atlas or city map can cause tremendous frustration. Other maps are less sensitive to the passage of time, but the date of production can still be important if, for example, better information becomes available in the period after publication.

I. Neatlines

Neatlines or clipping lines are used to frame a map and to indicate exactly where the area of a map begins and ends. The outer neatline of a map--its border--helps to frame the entire map composition to draw the reader's attention to the various elements of information. Neatlines are also used to "clip" the area of the body of the map and of locator, and inset maps.

J. Locator maps

Some maps portray areas whose locations may be unfamiliar to readers. In such cases, the cartographer adds a "helper" or locator map that places the body of the map within a larger geographical context with which the reader can be expected to be familiar.

k. Inset maps

Sometimes observations and data are so densely clustered in small sections of a larger map that the cartographer must provide the reader with additional close-up, "zoomed-in" maps of these smaller areas. Otherwise the data will obscure itself. These close-up detailed maps are called insets.

Chapter 3
GIS 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.

The coordinate location of a forestry stand would be spatial data, while the characteristics of that forestry stand, e.g. cover group, dominant species, crown closure, height, etc., would be attribute data. Other data types, in particular image and multimedia data, are becoming 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.

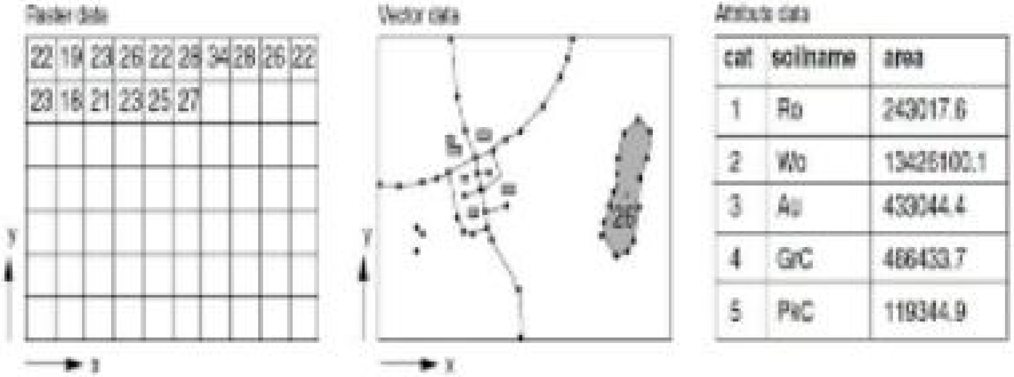


Fig 3.1 Data models in GIS – raster and vector data with attribute table

(1) SPATIAL DATA MODEL:

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.

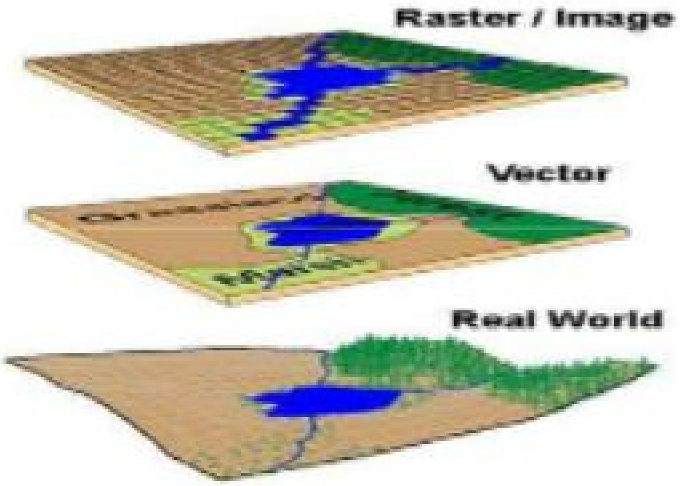


Fig 3.2 diagram reflects *vector and Raster*

Representation of the real world and showing differences in how a vector and a raster GIS will represent this real world.

1.1 Vector Data Model :

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.

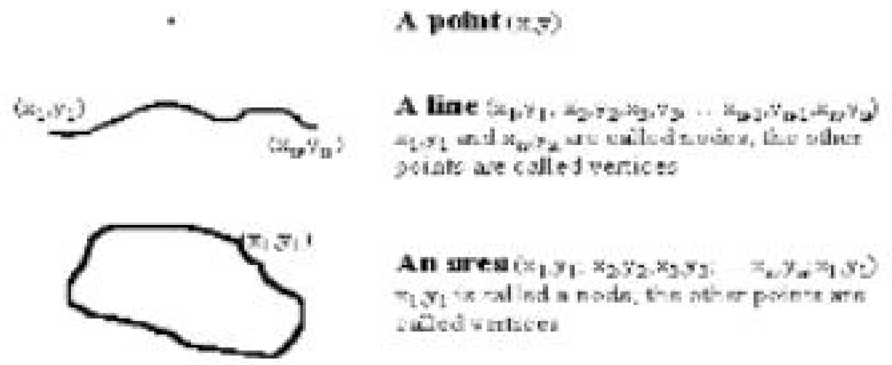


Fig 3.3 Basic Graphical elements of Vector Data

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.

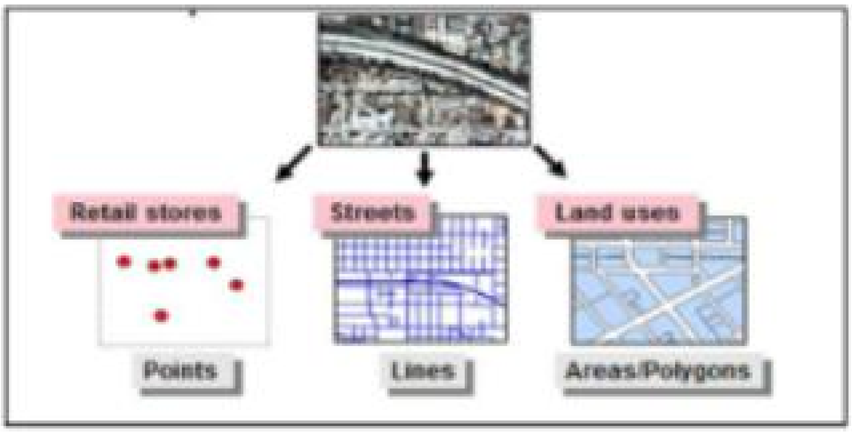


Fig 3.4 Representing feature in vector data

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.

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 (see table 3.1) 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.

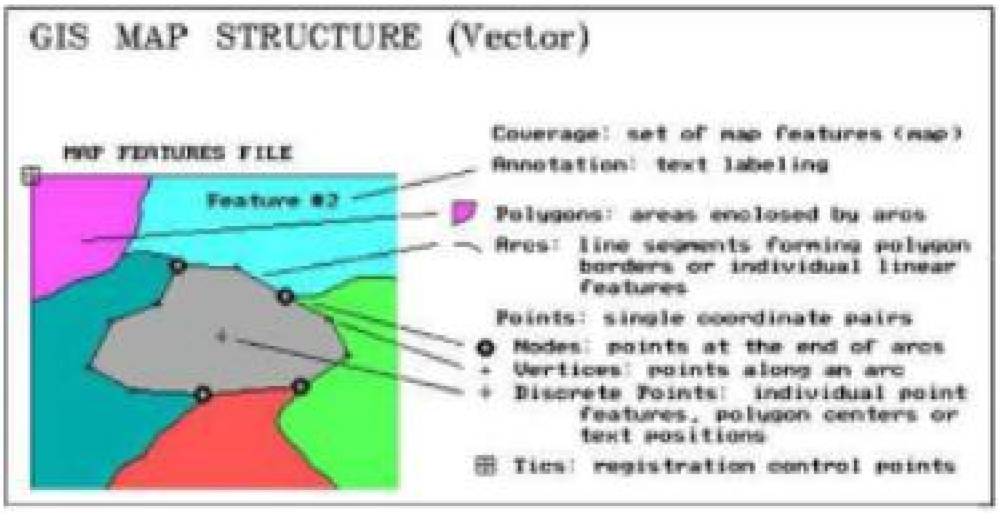


Fig 3.5 GIS MAP Structure - VECTOR systems (Adapted from Berry)

points	lines	polygons
Retail Stores	Highways	Countries
Cities, towns	City Streets	Postal Zones
Manhole Covers	Power Lines	Tax Parcels
Telephone Poles	Rivers	Blocks or Tracts
Airports	Water, Sewer Lines	Airports
Businesses	Railroads	Building Outlines
Warehouses	Shorelines	Military Installations
Customers	Bus Routes	Lakes
Prospects	Pipelines	Area Code Boundaries
Disease cases	Runways	Counties

Table 3.1 example of Layer Types

1.2 Raster Data Format :

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.

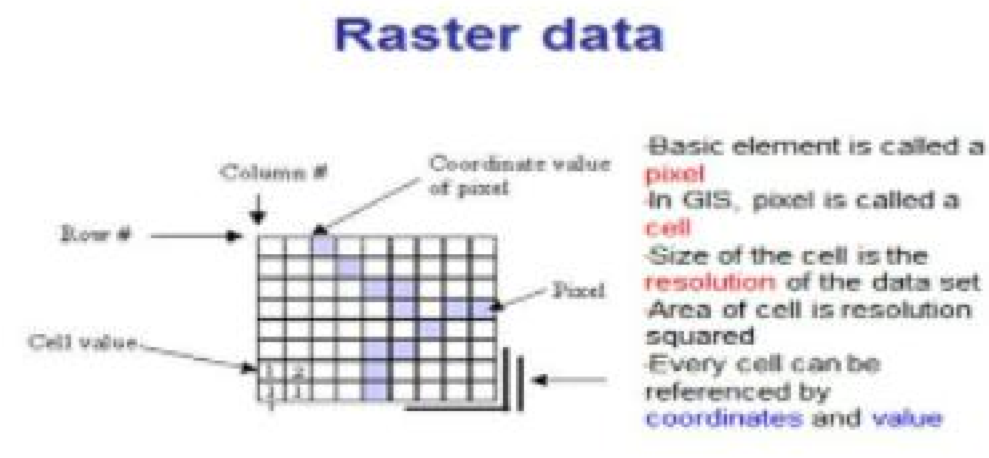


Fig 3.6 Raster Data Structure

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.

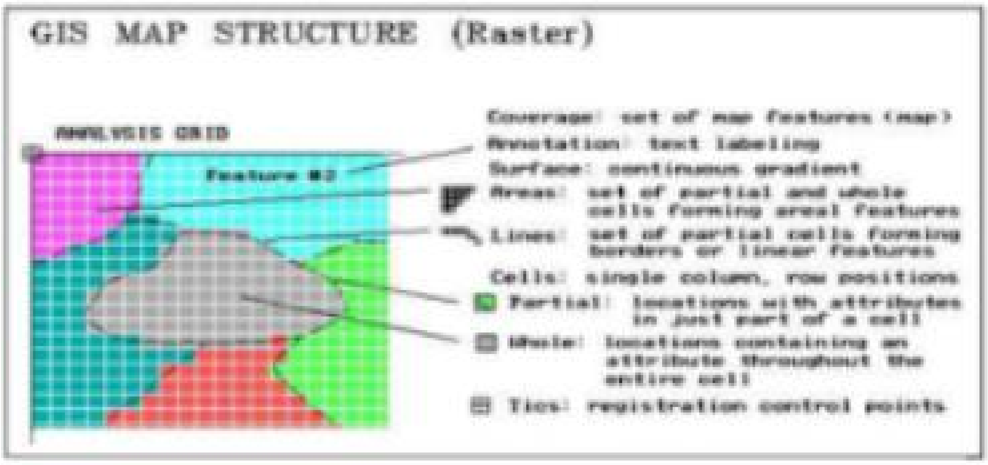


Fig 3.7 GIS MAP Structure - RASTER systems (Adapted from Berry)

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.

1.3 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 3.8 Image data is most often used for remotely sensed imagery such as satellite imagery or digital orthophotos.

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 cell size determines the resolution at which the data is represented.;
- It is especially difficult to adequately represent linear features depending on the cell resolution. Accordingly, network linkages are difficult to establish.
- Processing of associated attribute data may be cumbersome if large amounts of data exists. Raster maps inherently reflect only one attribute or characteristic for an area.
- Since most input data is in vector form, data must undergo vector-to-raster conversion. Besides increased processing requirements this may introduce data integrity concerns due to generalization and choice of inappropriate cell size.
- Most output maps from grid-cell systems do not conform to high-quality cartographic needs.

It is often difficult to compare or rate GIS software that use different data models. Some personal computer (PC) packages utilize vector structures for data input, editing, and display but convert to raster structures for any analysis. Other more comprehensive GIS offerings provide both integrated raster and vector analysis techniques. They allow users to select the data structure appropriate for the analysis requirements. Integrated raster and vector processing capabilities are most desirable and provide the greatest flexibility for data manipulation and analysis.

Finally, the next table illustrate Raster versus Vector

Table 3.2 : Raster vs Vector: a comparison

	Raster	Vector
1	Cell structure	Point, line (arc) and polygon structure
2	Location and attribute stored in one table	Location and attribute stored in separate tables
3	No topological information	Topological information recorded
4	An object has one attribute recorded	An object may have multiple attributes
5	More layers required	Less layers required
6	All cells are given a value in database	Only features are recorded
7	Data volume large	Data volume small
8	Data structure simple	Data structure complex
9	Fast processing	Slower processing
10	Resolution limited by pixel size	In theory, unlimited resolution
11	Average graphics	Good graphics
12	Low geometrical accuracy	High geometrical accuracy
13	RS data requires little conversion	RS data requires conversion
14	Spatial modelling easier	Spatial modelling harder
15	Network analysis poor	Network analysis good
16	Area analysis good	Area analyses average.

When Vector models are better ?

Vector models are better when :

- Boundaries are abrupt
- You look at relationships along a network (i.e. roads, streams, etc.)
- You want to store many attributes for spatial queries (especially if those attributes are for a single set of features)
- You want to create precise, high-quality maps

When Raster models are better ?

Raster Model are better when :

- Boundaries are gradual (e.g. topography, veg.)
- Features vary along a continuous surface
- Need to combine lots of data layers quickly and cheaply
- Using remotely sensed data
- Used for predictive modeling

Why not In most packages now *both* can be displayed together :

- To analyze and compare, must be in the *same* format
- Can store in one format and process in another - to save space
- Can convert from raster to vector for mapping (presentation)
- Can overlay the raster output with a vector coverage to allow better interpretation of the raster data and for effective mapping

(2) 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**

- **Hierarchial**
- **Network**
- **Relational**
- **Object Oriented**

The tabular model is the manner in which most early GIS software packages stored their attribute data. The next three models are those most commonly implemented in database management systems (DBMS). The object oriented is newer but rapidly gaining in popularity for some applications. A brief review of each model is provided.

2/1 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.

2/2 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. Hierarchial 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.

2/3 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.

2/4 Relational Model

The relational database model is the most widely accepted for managing the attributes of geographic data. 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. The following figure presents a sample table for forest inventory features. This table has 4 rows and 5 columns. The forest stand number would be the label for the spatial feature as well as the primary key for the database table. This serves as the linkage between the spatial definition of the feature and the attribute data for the feature.

UNIQUE STAND NUMBER	DOMINANT COVER GROUP	AVG. TREE HEIGHT	STAND SITE INDEX	STAND AGE
001	DEC	3	G	100
002	DEC-CON	4	M	80
003	DEC-CON	4	M	60
004	CON	4	G	120

Data is often stored in several tables. Tables can be joined or referenced to each other by common columns (relational fields). Usually the common column is an identification number for a selected geographic feature, e.g. a forestry stand polygon number. This identification number acts as the primary key for the table. The ability to join tables through use of a common column is the essence of the relational model. Such relational joins are usually ad hoc in nature and form the basis of for querying in a relational GIS product. Unlike the other previously discussed database types, relationships are implicit in the character of the data as opposed to explicit characteristics of the database set up.

There are many different designs of DBMSs, but in GIS the relational design has been the most useful. In the relational design, data are stored conceptually as a collection of tables. Common fields in different tables are used to link them together. This surprisingly simple design has been so widely used primarily because of its flexibility and very wide deployment in applications both within and without GIS.

In the relational design, data are stored conceptually as a collection of tables. Common fields in different tables are used to link them together.

In fact, most GIS software provides an internal relational data model, as well as support for commercial off-the-shelf (COTS) relational DBMS'. COTS DBMS' are referred to as external DBMS'. This approach supports both users with small data sets, where an internal data model is sufficient, and customers with larger data sets who utilize a DBMS for other corporate data storage requirements. With an external DBMS the GIS software can simply connect to the database, and the user can make use of the inherent capabilities of the DBMS. External DBMS' tend to have much more extensive querying and data integrity capabilities than the GIS' internal relational model. The emergence and use of the external DBMS is a trend that has resulted in the proliferation of GIS technology into more traditional data processing environments.

The relational DBMS is attractive because of its:

- simplicity in organization and data modelling.
- flexibility - data can be manipulated in an ad hoc manner by joining tables.
- efficiency of storage - by the proper design of data tables redundant data can be minimized; and
- the non-procedural nature - queries on a relational database do not need to take into account the internal organization of the data.
- The relational DBMS has emerged as the dominant commercial data management tool in GIS implementation and application.

The following diagram illustrates the basic linkage between a vector spatial data (topologic model) and attributes maintained in a relational database file.

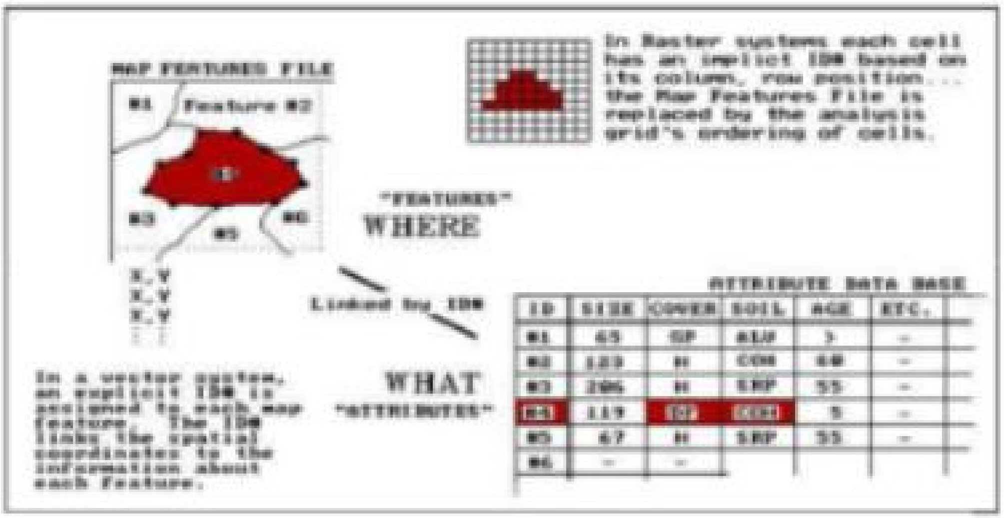


Fig 3.9 Basic linkages between a vector spatial data (topologic model) and attributes maintained in a relational database file (From Berry)

2/5 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. Fulfilment of this promise with a commercial GIS product remains to be seen.

(3) SPATIAL DATA RELATIONSHIPS :

The nature of spatial data relationships are important to understand within the context of GIS. In particular, the relationship between geographic features is a complex problem in which we are far from understanding in its entirety. This is of concern since the primary role of GIS is the manipulation and analysis of large quantities of spatial data. To date, the accepted theoretical solution is to topologically structure spatial data.

It is believed that a topologic data model best reflects the geography of the real world and provides an effective mathematical foundation for encoding spatial relationships, providing a data model for manipulating and analyzing vector based data.

Most GIS software segregate spatial and attribute data into separate data management systems. Most frequently, the topological or raster structure is used to store the spatial data, while the relational database structure is used to store the attribute data. Data from both structures are linked together for use through unique identification numbers, e.g. feature labels and DBMS primary keys. This coupling of spatial features with an attribute record is usually maintained by an internal number assigned by the GIS software. A label is required so the user can load the appropriate attribute record for a given geographic feature. Most often a single attribute record is automatically created by the GIS software once a clean topological structure is properly generated. This attribute record normally contains the internal number for the feature, the user's label identifier, the area of the feature, and the perimeter of the feature. Linear features have the length of the feature defined instead of the area.

TOPOLOGY

The topologic model is often confusing to initial users of GIS. Topology is a mathematical
approach that allows us to structure data based on the principles of feature adjacency and feature connectivity. It is in fact the mathematical method used to define spatial relationships. Without a topologic data structure in a vector based GIS most data manipulation and analysis functions would not be practical or feasible.

The most common topological data structure is the arc/node data model. This model contains two basic entities, the arc and the node. The arc is a series of points, joined by straight line segments, that start and end at a node. The node is an intersection point where two or more arcs meet. Nodes also occur at the end of a dangling arc, e.g. an arc that does not connect to another arc such as a dead end street. Isolated nodes, not connected to arcs represent point features. A polygon feature is comprised of a closed chain of arcs.

In GIS software the topological definition is commonly stored in a proprietary format.

However, most software offerings record the topological definition in three tables. These tables are analogous to relational tables. The three tables represent the different types of features, e.g. point, line, area. A fourth table containing the coordinates is also utilized. The

node table stores information about the node and the arcs that are connected to it. The arc table contains topological information about the arcs. This includes the start and end node, and the polygon to the left and right that the arc is an element of. The polygon table defines the arcs that make up each polygon. While arc, node, and polygon terminology is used by most GIS vendors, some also introduce terms such as edges and faces to define arcs and polygons.

This is merely the use of different words to define topological definitions. Do not be confused by this.

Since most input data does not exist in a topological data structure, topology must be built with the GIS software. Depending on the data set this can be an CPU intensive and time consuming procedure. This building process involves the creation of the topological tables and the definition of the arc, node, and polygon entities. To properly define the topology there are specific requirements with respect to graphic elements, e.g. no duplicate lines, no gaps in arcs that define polygon features, etc. These requirements are reviewed in the Data Editing section of the book.

The topological model is utilized because it effectively models the relationship of spatial entities. Accordingly, it is well suited for operations such as contiguity and connectivity analyses. Contiguity involves the evaluation of feature adjacency, e.g. features that touch one another, and proximity, e.g. features that are near one another. The primary advantage of the topological model is that spatial analysis can be done without using the coordinate data. Many operations can be done largely, if not entirely, by using the topological definition alone. This is a significant advantage over the CAD or spaghetti vector data structure that requires the derivation of spatial relationships from the coordinate data before analysis can be undertaken.

The major disadvantage of the topological data model is its static nature. It can be a time consuming process to properly define the topology depending on the size and complexity of the data set. For example, 2,000 forest stand polygons will require considerably longer to build the topology that 2,000 municipal lot boundaries. This is due to the inherent complexity of the features, e.g. lots tend to be rectangular while forest stands are often long and sinuous. This can be a consideration when evaluating the topological building capabilities of GIS software.

The static nature of the topological model also implies that every time some editing has occurred, e.g. forest stand boundaries are changed to reflect harvesting or burns, the topology must be rebuilt. The integrity of the topological structure and the DBMS tables containing the attribute data can be a concern here. This is often referred to as referential integrity. While topology is the mechanism to ensure integrity with spatial data, referential integrity is the concept of ensuring integrity for both linked topological data and attribute data.

Chapter 4
GIS Data Sources

This chapter reviews different sources, formats, and input techniques for GIS data. The focus is on reviewing different data input techniques for spatial data. This chapter also describes data input errors, spatial and attribute, and reviews typical procedures to correct input errors. This chapter will be of most interest to technical staff and GIS operators.

- **Sources of Data**
- **Data Input Techniques**
- **Data Editing and Quality Assurance**

4.1 SOURCES OF DATA

As previously identified, two types of data are input into a GIS, spatial and attribute. The data input process is the operation of encoding both types of data into the GIS database formats.

The creation of a clean digital database is the most important and time consuming task upon which the usefulness of the GIS depends. The establishment and maintenance of a robust spatial database is the cornerstone of a successful GIS implementation.

As well, the digital data is the most expensive part of the GIS. Yet often, not enough attention is given to the quality of the data or the processes by which they are prepared for automation. The general consensus among the GIS community is that 60 to 80 % of the cost incurred during implementation of GIS technology lies in data acquisition, data compilation and database development.

A wide variety of data sources exist for both spatial and attribute data. The most common general sources for spatial data are:

- hard copy maps;
- aerial photographs;
- remotely-sensed imagery;
- point data samples from surveys; and
- existing digital data files.

Existing hard copy maps, e.g. sometimes referred to as analogue maps, provide the most popular source for any GIS project.

Attribute data has an even wider variety of data sources. Any textual or tabular data than can be referenced to a geographic feature, e.g. a point, line, or area, can be input into a GIS Attribute data is usually input by manual keying or via a bulk loading utility of the DBMS software. ASCII format is a de facto standard for the transfer and conversion of attribute information. The following figure describes the basic data types that are used and created by a GIS.

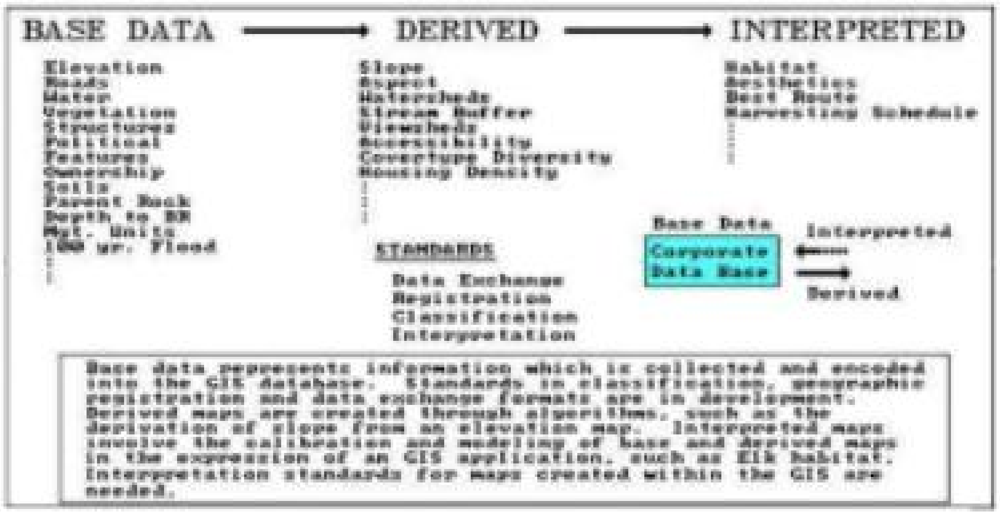


Fig 4.1 The basic data types that are used and created by a GIS (after Berry).

4.2 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.

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.

4.2.1 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 line work 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.

4.2.2 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.

4.2.3 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.

4.2.4 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.

4.3 DATA EDITING AND QUALITY ASSURANCE

Data editing and verification is in response to the errors that arise during the encoding of spatial and non-spatial data. The editing of spatial data is a time consuming, interactive process that can take as long, if not longer, than the data input process itself.

Several kinds of errors can occur during data input. They can be classified as:

- Incompleteness of the spatial data. This includes missing points, line segments, and/or polygons.
- Locational placement errors of spatial data. These types of errors usually are the result of careless digitizing or poor quality of the original data source.
- Distortion of the spatial data. This kind of error is usually caused by base maps that are not scale-correct over the whole image, e.g. aerial photographs, or from material stretch, e.g. paper documents.
- Incorrect linkages between spatial and attribute data. This type of error is commonly the result of incorrect unique identifiers (labels) being assigned during manual key in or digitizing. This may involve the assigning of an entirely wrong label to a feature, or more than one label being assigned to a feature.
- Attribute data is wrong or incomplete. Often the attribute data does not match exactly with the spatial data. This is because they are frequently from independent sources and often different time periods. Missing data records or too many data records are the most common problems.

The identification of errors in spatial and attribute data is often difficult. Most spatial errors become evident during the topological building process. The use of check plots to clearly determine where spatial errors exist is a common practice. Most topological building functions in GIS software clearly identify the geographic location of the error and indicate the nature of the problem. Comprehensive GIS software allows users to graphically walk through and edit the spatial errors. Others merely identify the type and coordinates of the error. Since this is often a labour intensive and time consuming process, users should consider the error correction capabilities very important during the evaluation of GIS software offerings.

4/3/1 SPATIAL DATA ERROR :

A variety of common data problems occur in converting data into a topological structure. These stem from the original quality of the source data and the characteristics of the data capture process. Usually data is input by digitizing. Digitizing allows a user to trace spatial data from a hard copy product, e.g. a map, and have it recorded by the computer software. Most GIS software has utilities to clean the data and build a topologic structure. If the data is unclean to start with, for whatever reason, the cleaning process can be very lengthy. Interactive editing of data is a distinct reality in the data input process.

The most common problems that occur in converting data into a topological structure include:

- slivers and gaps in the line work;
- dead ends, e.g. also called dangling arcs, resulting from overshoots and undershoots in the line work; and
- bow ties or weird polygons from inappropriate closing of connecting features.

Of course, topological errors only exist with linear and areal features. They become most evident with polygonal features. Slivers are the most common problem when cleaning data.

Slivers frequently occur when coincident boundaries are digitized separately, e.g. once each for adjacent forest stands, once for a lake and once for the stand boundary, or after polygon overlay. Slivers often appear when combining data from different sources, e.g. forest inventory, soils, and hydrography. It is advisable to digitize data layers with respect to an existing data layer, e.g. hydrography, rather than attempting to match data layers later. A proper plan and definition of priorities for inputting data layers will save many hours of interactive editing and cleaning.

Dead ends usually occur when data has been digitized in a spaghetti mode, or without snapping to existing nodes. Most GIS software will clean up undershoots and overshoots based on a user defined tolerance, e.g. distance. The definition of an inappropriate distance often leads to the formation of bow ties or weird polygons during topological building. Tolerances that are too large will force arcs to snap one another that should not be connected. The result is small polygons called bow ties. The definition of a proper tolerance for cleaning requires an understanding of the scale and accuracy of the data set.

The other problem that commonly occurs when building a topologic data structure is duplicate lines. These usually occur when data has been digitized or converted from a CAD system. The lack of topology in these type of drafting systems permits the inadvertent creation of elements that are exactly duplicate. However, most GIS packages afford automatic elimination of duplicate elements during the topological building process. Accordingly, it may not be a concern with vector based GIS software. Users should be aware of the duplicate element that retraces itself, e.g. a three vertice line where the first point is also the last point. Some GIS packages do not identify these feature inconsistencies and will build such a feature as a valid polygon. This is because the topological definition is mathematically correct, however it is not geographically correct. Most GIS software will provide the capability to eliminate bow ties and slivers by means of a feature elimination command based on area, e.g. polygons less than 100 square metres. The ability to define custom topological error scenarios and provide for semiautomated correction is a desirable capability for GIS software.

The adjoining figure illustrates some typical errors described above. Can you spot them ? They include undershoots, overshoots, bow ties, and slivers. Most bow ties occur when inappropriate tolerances are used during the automated cleaning of data that contains many overshoots. This particular set of spatial data is a prime candidate for numerous bow tie polygons.

4/3/2 Attribute Data Errors

The identification of attribute data errors is usually not as simple as spatial errors. This is especially true if these errors are attributed to the quality or reliability of the data. Errors as such usually do not surface until later on in the GIS processing. Solutions to these type of problems are much more complex and often do not exist entirely. It is much more difficult to spot errors in attribute data when the values are syntactically good, but incorrect.

Simple errors of linkage, e.g. missing or duplicate records, become evident during the linking operation between spatial and attribute data. Again, most GIS software contains functions that check for and clearly identify problems of linkage during attempted operations. This is also an area of consideration when evaluating GIS software.

4/4 Data Verification

Six clear steps stand out in the data editing and verification process for spatial data. These are:

- ***Visual review***. This is usually by check plotting.
- ***Cleanup of lines and junctions***. This process is usually done by software first and interactive editing second.
- ***Weeding of excess coordinates***. This process involves the removal of redundant vertices by the software for linear and/or polygonal features.
- ***Correction for distortion and warping***. Most GIS software has functions for scale correction and rubber sheeting. However, the distinct rubber sheet algorithm used will vary depending on the spatial data model, vector or raster, employed by the GIS. Some raster techniques may be more intensive than vector based algorithms.
- ***Construction of polygons***. Since the majority of data used in GIS is polygonal, the construction of polygon features from lines/arcs is necessary. Usually this is done in conjunction with the topological building process.
- ***The addition of unique identifiers or labels***. Often this process is manual. However,some systems do provide the capability to automatically build labels for a data layer.

These data verification steps occur after the data input stage and prior to or during the linkage of the spatial data to the attributes. Data verification ensures the integrity between the spatial and attribute data. Verification should include some brief querying of attributes and cross checking against known values.

Chapter 5

Designing a GIS Database

5/1 Overview :

Nothing is more critical to the success of a GIS project than good planning. In this lesson, you will be introduced to the issues and processes involved in designing a GIS database. You will then put some of these concepts into practice in the following lessons as you build a geodatabase and use it to conduct some simple analysis.

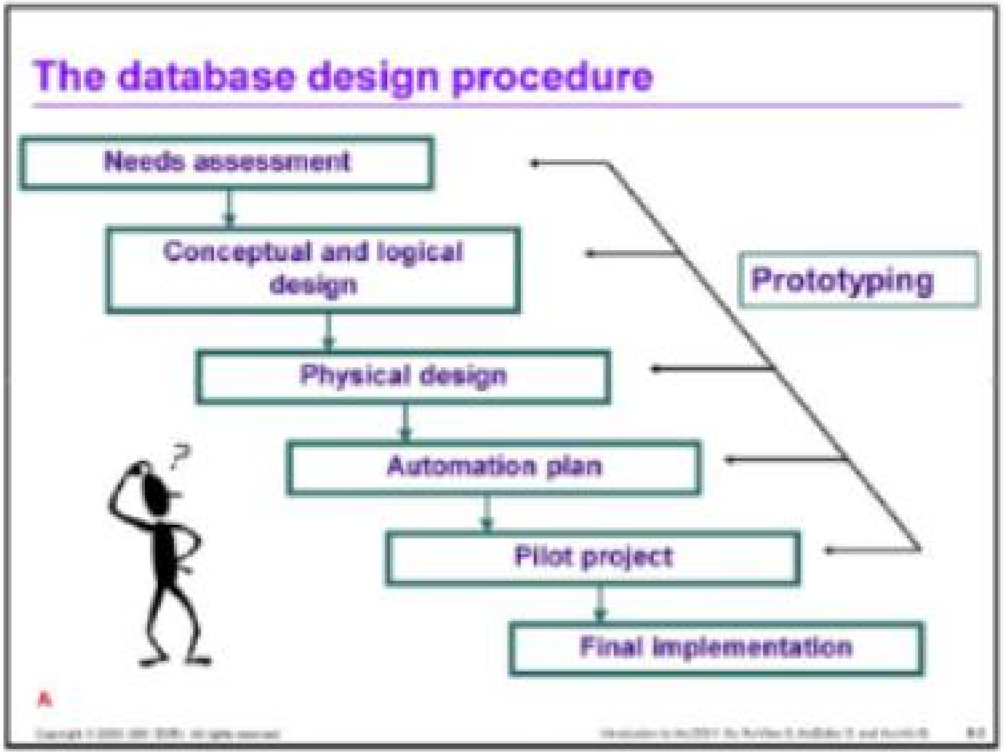


Fig 5.1 GIS database design Phase

5/2 The database design procedure

There is no single correct procedure for database design; the procedure outlined here is a guideline. The database design process is not linear. There are feedback loops, especially between the pilot study, the conceptual and logical designs, and the physical design.

5/2/1 Needs assessment :

Determine the functions that will be supported by the GIS in this time-consuming step. To get the information necessary to complete this task, you could interview potential users, tour operations, conduct inventories of data, hardware, software, and personnel, and conduct educational seminars.

5/2/2 Conceptual and logical design :

Determine database contents and how to logically organize the data in the database. The conceptual and logical design for the project in this course is based on the concept of a shared database using ESRI data models, particularly the geodatabase.

5/2/3 Physical design :

Physically structure the data so that it conforms to the ESRI data structures. A detailed database schema is also implemented in this phase, along with plans for documentation and naming conventions.

5/2/4 Automation plan :

Establish automation procedures. Prepare data for automation and implement the plan. If there are problems during this step, you may need to reevaluate the design of the database.

5/2/5 Pilot project :

Test the functionality, performance, and flexibility of the database design. At least one pilot study should be performed before full implementation of the database.

5/2/6 Prototype :

Remember to test as you go. Consistent feedback from users can avoid problems later.

(1)Assessing needs :

Perhaps no step in the database design process is as critical and potentially time-consuming as the needs assessment. There is a good reason for committing the proper attention and resources to this step. The ability of the GIS to operate acceptably is directly related to how well the necessary functions the GIS will perform are understood. To gather the right information, you must ask the right people the right questions. This means conducting detailed interviews with end users, data managers, supervisory personnel, and anyone else who might be impacted by the new system. From these interviews you should try to determine:

•What tasks are performed using spatial data.

•What tasks are performed that do not currently use spatial data, but have a spatial component.

•What tasks are not currently performed, but are desired.

•Who will be using the system and what their roles will be.

•What types of products the system will need to produce (maps, reports, etc.).

•What applications will need to be developed.

•What data will be needed to fulfill the stated tasks.

•Will the data be shared among multiple users.

•What special security measures are necessary.

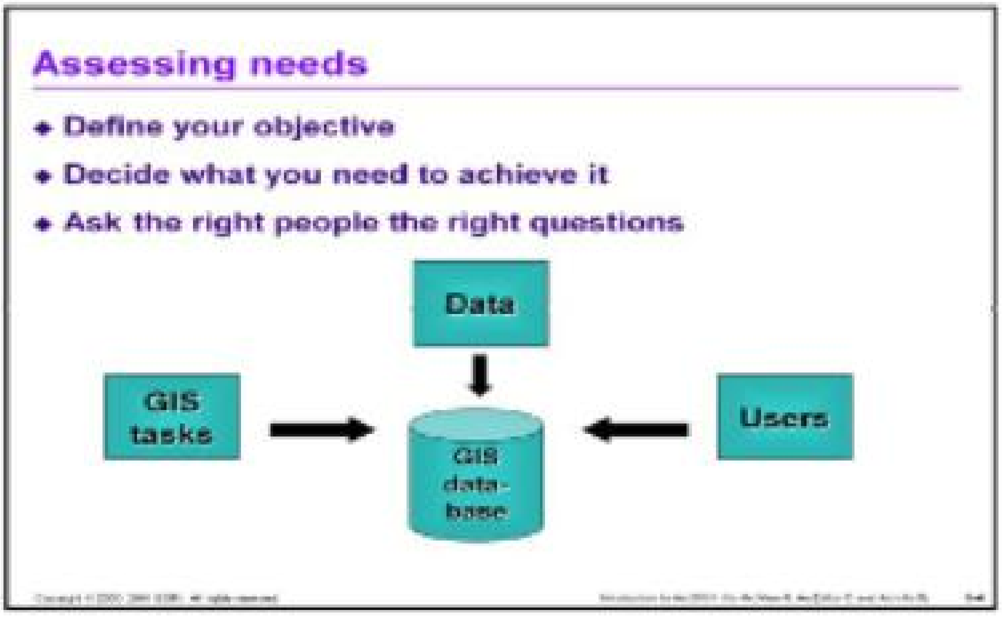


Fig 5.2 Needs Assessment

(2)Conceptual and logical design

Once you have determined which data to store in your database, your next task is to select feature classes and organize these feature classes into layers. The conceptual design is a top-level concept of how the database will work. The logical design is a detailed layout that fills in the conceptual design in accordance with a specific data model. This procedure includes determining the database contents (spatial, attribute, and behavior), selecting appropriate geographic datasets, and organizing the content into a series of themes.

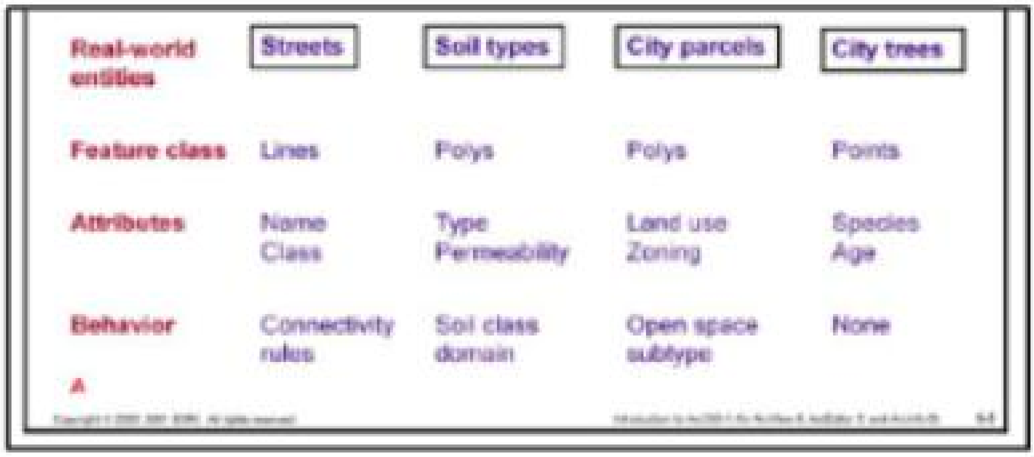


Fig 5.3 Conceptual and logical design

A) Determining the data storage format Data storage formats :

ArcInfo and ArcView offer you many choices for storing spatial and attribute data. Each format has it strengths and limitations, but if you understand your needs, you should be able to select the ones that will work best for the GIS being developed. ***Some issues to consider when choosing data formats are:***

- **Topology:** Coverages store polygon and line topology. The geodatabase, at present, stores only line topology. Shapefiles do not store topology.
- **Vector versus raster data:** Raster format is especially suited to data without clear boundaries (continuous data) such as temperature, pollution, and elevation. Vector formats store discrete lines, so they are more suited to discrete data such as streets and parcel boundaries.
- **Consistency**: The formats you choose should be standardized within your organization, and easily converted if necessary.

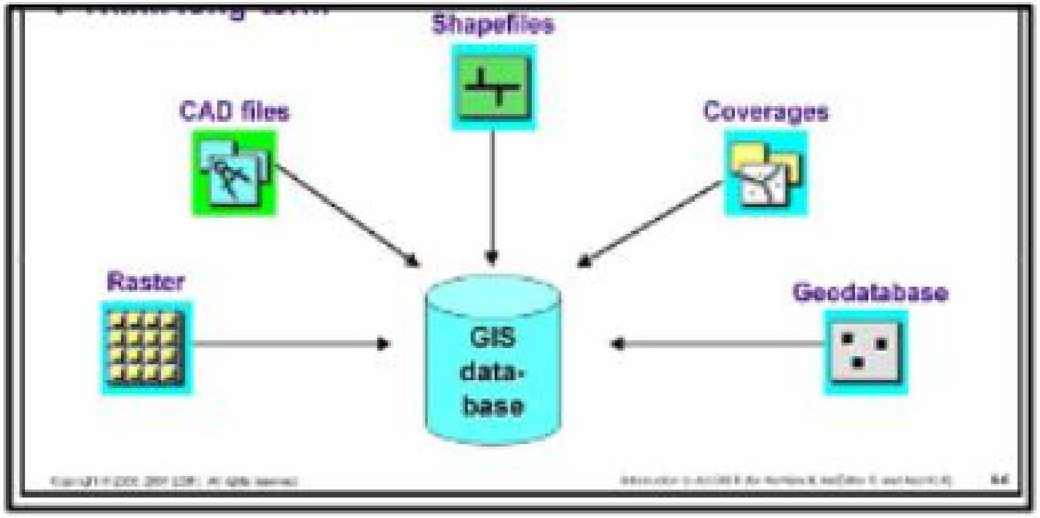


Fig 5.4 Data Storage Format

B) Why the geodatabase?

The geodatabase format offers many unique advantages for geographic data storage.

Scalability As user needs for security and data management grow, the geodatabase can meet them.

Custom features Because of its COM architecture, custom objects can be programmed that represent real-world features more accurately.

Domains and subtypes These properties, easily created and maintained in the geodatabase, make data creation, editing, and maintenance much more efficient and would require special programming to be achieved in other formats.

Geometric Networks Using the geodatabase, you can create geometric networks for modeling connectivity and performing trace and path finding analysis.

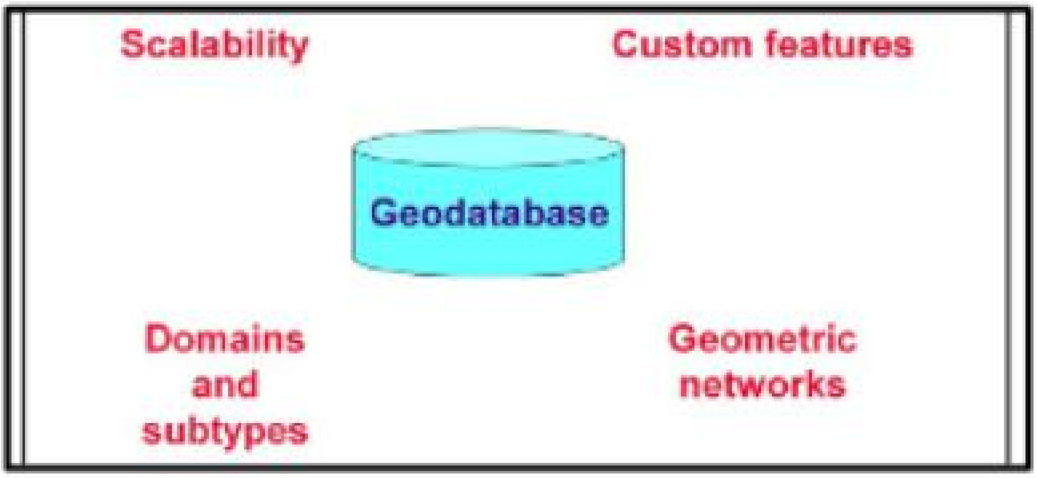


Fig5.5 Why geodatabase

(3)Physical design

There are some physical design issues to take into consideration when designing your database.

A) Data sources

Decide which source of data to use for the GIS. This step may involve searching outside the organization. This is an important step because both the cost and quality of the database can be profoundly affected by the choice of data sources.

B) Designing individual datasets

Make the final layout of each geographic dataset and independent table. Decide precisely how each data element will be stored and what coding schemes will be used. Determine what relationships will be stored and what the keys will be. Also determine what domains and subtypes will be established. This step is a detailed translation of the conceptual design into a physical layout.

C) Documentation

Establish a standardized naming convention for spatial, attribute, and related file data. Determine what information will be stored as the documentation for the database and what the procedures will be to implement it.

D) Database schema

The schema of a GIS database is its overall structure. This part of the physical design translates the conceptual or logical design into a detailed layout. In addition to the data components, the schema should also take into account the physical storage devices, security issues, and user needs.

Once designed, a database schema should be difficult to alter, but possible if truly necessary. The schema is similar to the blueprints for a house. It would be impractical to decide you want a French Tudor style house halfway through the construction of your Victorian cottage, yet you could do it. The key is to know what you want before you start and to plan for it.

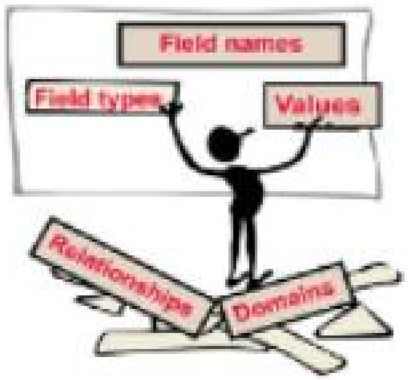


Fig5.6 Data Base Schema

E) Metadata: Documenting your data

Data is of limited use without metadata (data about data). Metadata is a reference that serves as a blueprint both for building the database, and as a guide for subsequent users. Storing this documentation and keeping it current has long been one of the more frustrating and least-performed tasks in GIS database development, but ArcInfo 8 makes the creation and maintenance of metadata much easier than it has been in the past.

The ArcCatalog application contains a special metadata tool that allows you to enter documentation for each component of the database. The information you can store conforms to the requirements of the Federal Geographic Data Committee (FGDC). ArcInfo will automatically capture and store some of the data (the Properties shown on this figure). The user is required to store any other information they believe is relevant.

The metadata is stored in Extensible Markup Language (XML), a simple internet compliant format. This means that your metadata created in ArcCatalog can be viewed from there or with a Web browser. There are different predefined formats, or style sheets, for viewing the metadata, but you can also create your own.

- **Types of metadata**

  Metadata created in ArcCatalog contains two components.

  - **Documentation**

    This is the information added to the metadata by the user. It usually describes the quality or content of the data. Examples include field descriptions, code definitions, and contact information.

  - **Properties**

This information is automatically stored in the metadata for each feature class by ArcCatalog. Examples include projection definition and number of features.

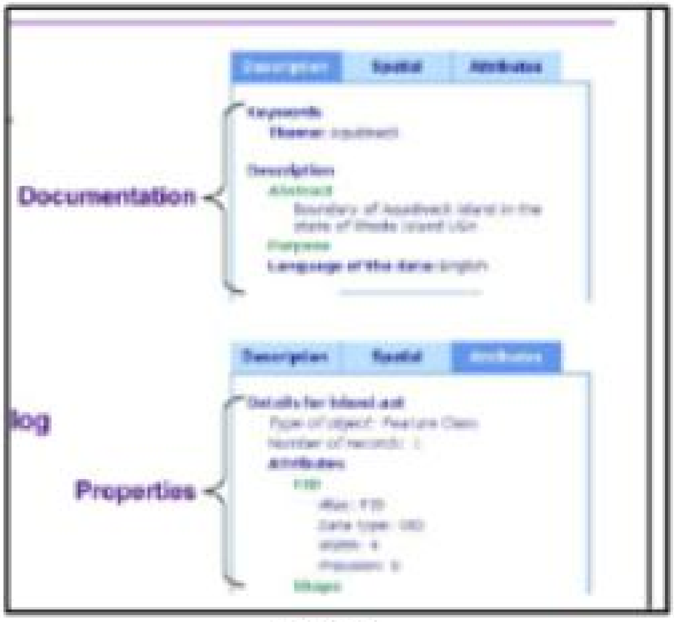


Fig 5.7

- **Viewing metadata**

You can view metadata in ArcCatalog by clicking on an item in the Catalog tree, then clicking the Metadata tab in the display area. There are a number of stylesheets available in ArcCatalog to view metadata: ESRI (which is the default), FGDC or FGDC FAQ, the Geography Network and XML. Stylesheets define how XML data is presented and are written using Extended Stylesheet Language (XSL).

XML: The Extensible Markup Language (XML) stylesheet presents the raw XML data and all the tags and values contained in the metadata.

FGDC or FGDC FAQ: The Federal Geographic Data Committee (FGDC) stylesheet selects only the information defined by the FGDC standard. The FAQ formats the metadata by frequently asked questions.

The Geography Network: Formats metadata as it is seen in the Geography Network.

ESRI: The ESRI stylesheet is the default stylesheet; it selects a subset of the entire body of metadata and presents it as if it was in a tabbed dialog.

You can also create your own customized stylesheet for displaying metadata in ArcCatalog.

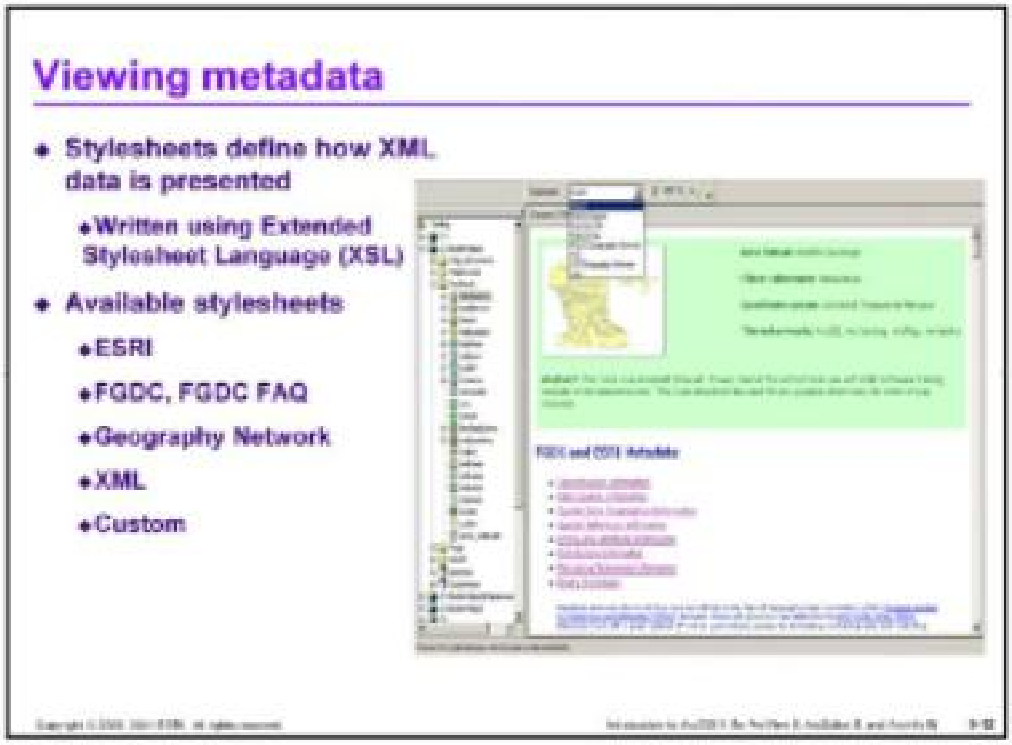


Fig 5.8 viewing metadata

- **Editing metadata**

Metadata can be edited using a special editor that you can access by clicking the Edit Metadata button on the Metadata toolbar. The metadata editor allows you to edit individual fields of the metadata.

In addition, files describing the contents of a data source can be enclosed in the metadata. A copy of the file is contained within the metadata. Enclosing files in metadata works the same way that you enclose files in an E-mail message.

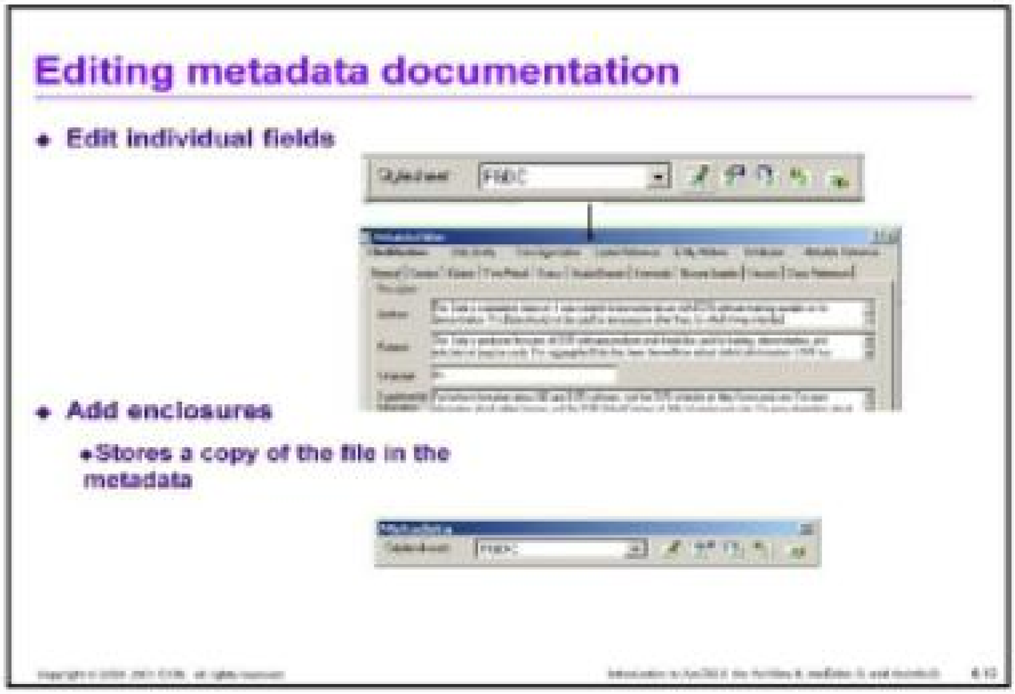


Fig 5.9 editing metadata

- **Setting properties**

Options are available in the properties of the metadata to control automatic creation and updating of metadata. By default, when you go to view the metadata for an item in the Catalog tree for the first time, ArcCatalog will automatically create metadata for that item. You may also choose to have ArcCatalog automatically update metadata for a selected item in the Catalog tree.

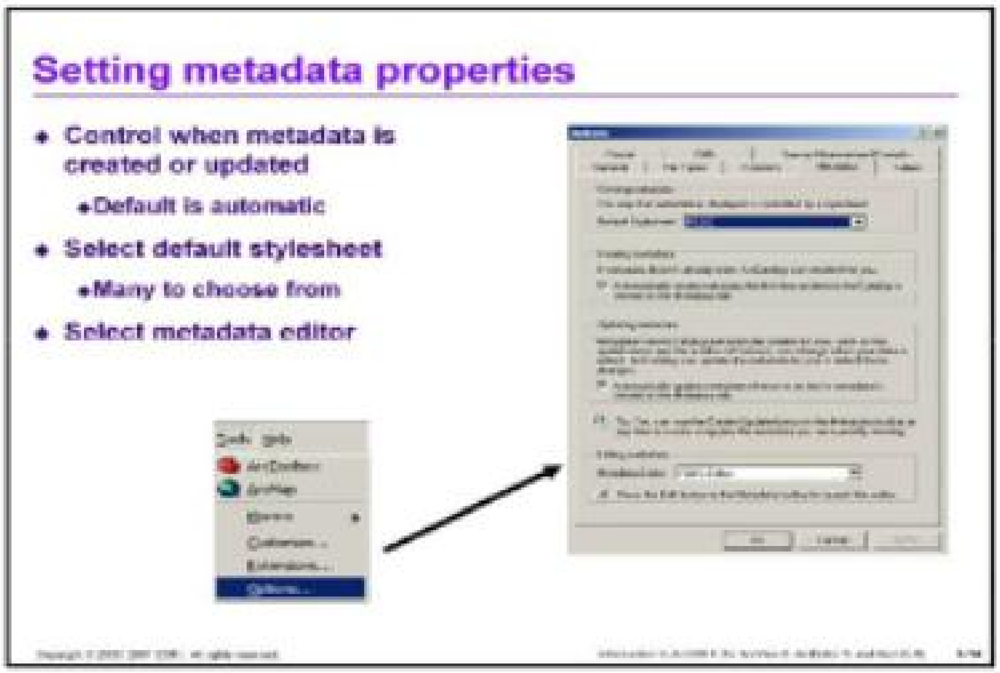


Fig 5.10 setting metadata properties

- **Importing and exporting metadata**

The standard format that ArcCatalog uses for exporting and importing metadata is FGDC CSDGM (Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata). In addition, metadata can be imported or exported in different formats into ArcCatalog. The following formats are supported for importing metadata into ArcCatalog:

FGDC CSDGM (SGML, TXT, XML)

The following formats are supported for exporting metadata from ArcCatalog:

FGDC CSDGM (FAQ, HTML, SGML, TXT, XML) and HTML

F) Choosing a database projection

Because projecting a three-dimensional surface onto a flat one causes distortion of spatial properties, choosing the appropriate projection can have an important impact on your database. This slide lists some of the issues you need to consider when making this decision.

You should also remember that even if you store datasets in different projections, the ArcMap application can still project them on the fly so they all display in the same coordinate space. For this reason, some database designers elect to store their data in decimal degrees of latitude and longitude, then apply a projection when the need arises.

(4) Automating data

Once a database design has been finalized, you can begin to incorporate data into the database. Although there are a myriad of commercial sources and techniques for capturing data, the process usually follows these three steps:

1. Convert the data into the desired format. This could require multiple steps.
2. Correct any spatial errors and add the appropriate attribute data.
3. Aggregate individual pieces of data into the complete representation of your study area. This may also require eliminating some superfluous data.

Remember to plan the methods of data automation before beginning this phase of the project and to document each step.

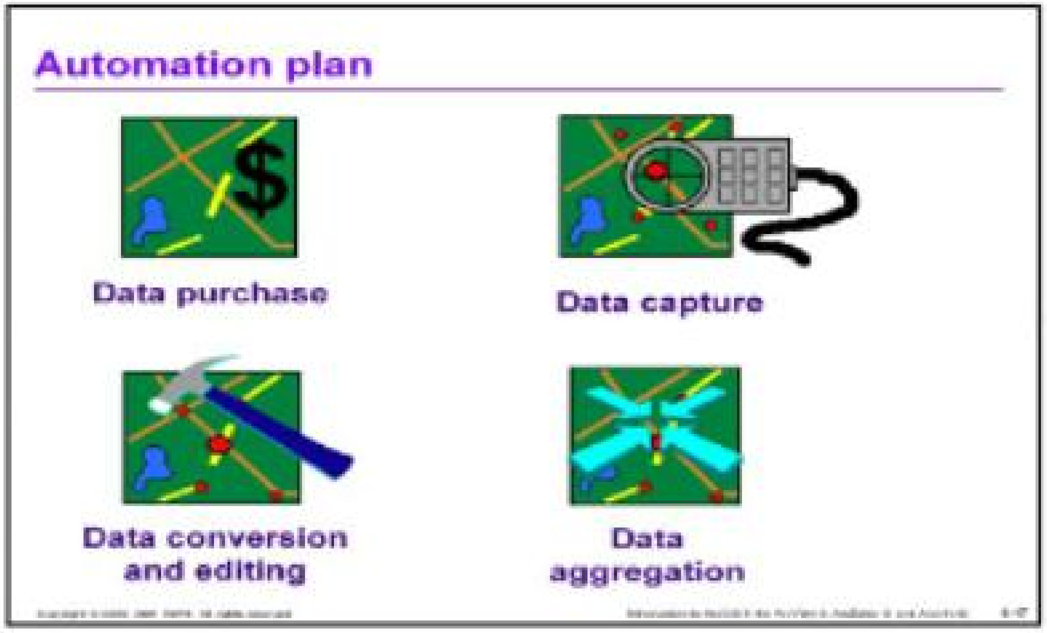


Fig 5.11 Automating plan

(5)Pilot project

Conducting at least one pilot study lets you evaluate the design of your database at a relatively low cost prior to full-scale implementation. The design should be tested for functionality, performance, and flexibility.

To meet this objective, a pilot study should reflect the types of tasks the GIS will be required to perform. It should also use several different datasets to ensure that the design is working across the board. More than one pilot study may be required.

Chapter 6
Data Automation
(Digitizing)

The next step in building the database is to automate the data ; that is to convert features on a map to a digital format on the computer. In most GIS methods a digital format is called coverage and on a map requires preparation of the manuscripts for automation. Capturing the features from a map manuscript and visually evaluation of the quality of the data captured . Data on a map can be captured by digitizing each feature. One by one , or by using and electronic scanner by typing in the exact x- and y coordinates by digitizing.

6.1 Data automation options

Geographic data can come from many sources: map manuscripts, digital scanned data, coordinate files, or other standard digital formats. ArcGIS software has a number of commands and procedures to convert data from various sources. Data can be entered, stored, and manipulated in either a vector or raster data structure. Vector data consists of points, lines, and polygons. Each feature has a set of x,y coordinates and associated attributes. Raster data use rows and columns of grid cells. Each cell can be assigned an attribute and a unique location in the grid.

6/1/1 Analog data Analog data can be digitized or scanned.

•Digitizing captures spatial data in digital form as a series of x,y coordinate pairs and stores it directly in a coverage, shapefile, or geodatabase.

•Scanning produces a digital raster file that can be vectorized and then converted into a vector format.

6/1/2 Digital data Conversion programs available in ArcInfo and ArcView convert other digital data from vector- or raster-formatted files to an ESRI data format. Conversion tools will differ depending on the level of software, ArcInfo or ArcView, you are using. Refer to Data Conversion in the online documentation system for information about conversion tools provided by your ArcGIS software.

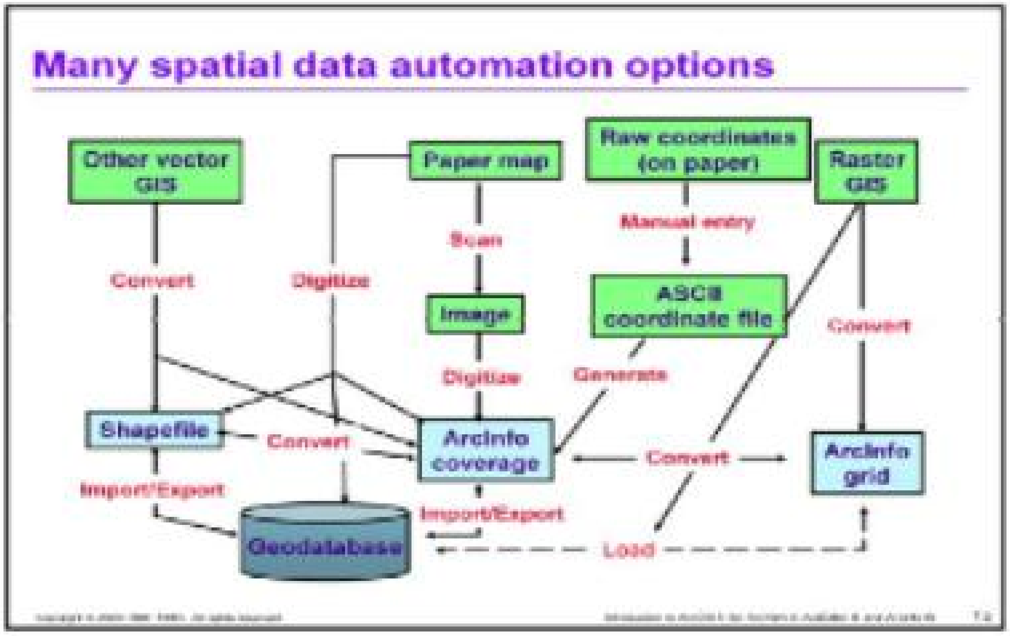


Fig 6.1 Spatial Data Automation

6.2 Key Terms :

Before beginning digitizing , here is a brief review of the terms needed for digitizing and the symbols used to represent them :

Table 6.1 Key terms

Terms	Symbols	Descriptions
Arc	----	A line feature of the border of a polygon. An are is defined by a from-node and a to-node and additional vertices in between.
Node		An are endpoint. In many cases, the end of one are makes the beginning of another , so it can be said that nodes also mark intersection.
Vertics	●	A point at within an arc providing shape
Pesudo node	◆	The point at which an arc connects to itself, or to only one other arc.
Dangling node	■	An Arc endpoint not connected
Label point	+	Used either to represent a point feature or to identify polygon .
User-ID	70	A number assigned to each features. These values should be unique . once established , these values can be altered by the user as needed.
TIC		A registration of geographic control point. These allow all coverage features to be registered to the same coordinated system.

6.3 Digitizing :

Digitizing is the process of converting the spatial features on a map into a digital format : point , line , and area features that from a map are covered into x , y coordinates A point is represented by a single coordinate a line by a string of coordinates end when combined one or more lines with a label point inside outline and identify an area or polygon . thus digitizing is a procedure on a map . this can be demonstrated by taking any map manuscript and breaking it into its component parts a number of point and lines. The spatial data shown in the following portion of mapped region can be captured by digitizing five lines and five label points.

- ***TIC points digitizing :***

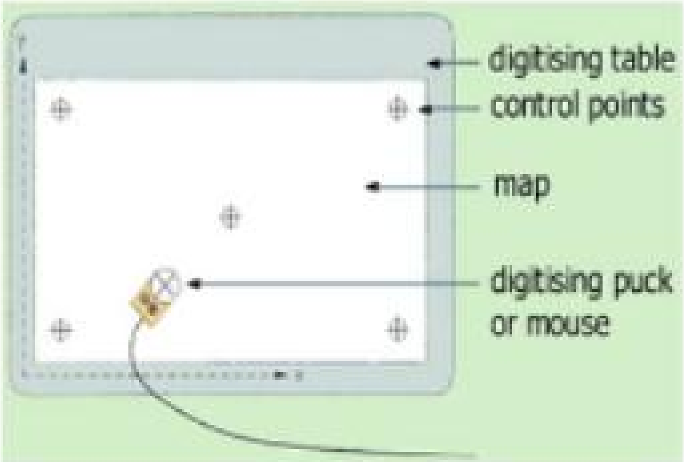


Figure 6.2 The digitizer table and puck

A tic is a registration or geographic control point for a coverage.
Tics allow coverage coordinates to be registered to a common coordinate system (universal transverse Mercator [UTM] meters, state plane feet, and so on) and, therefore, relate locations of features in a

coverage to locations on the earth's surface. Tics are important for registering map sheets during digitizing and editing.

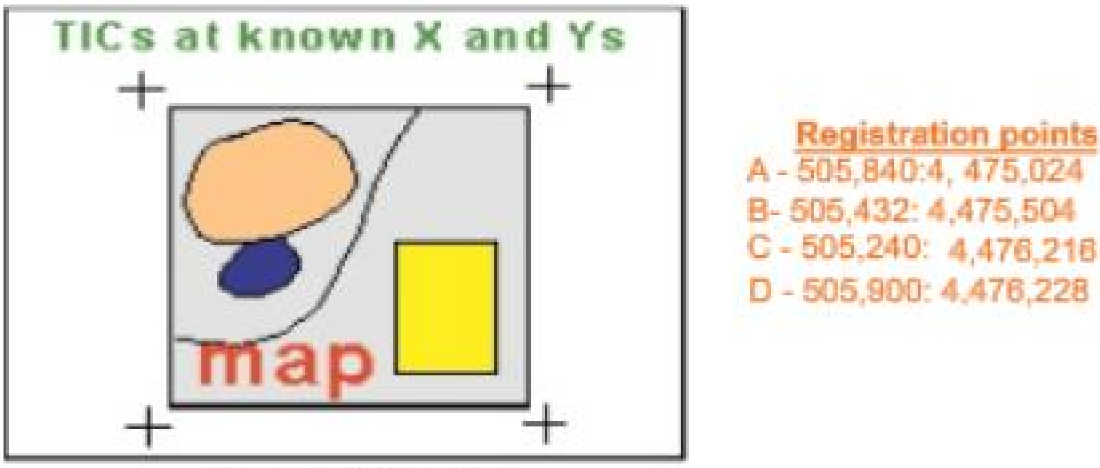
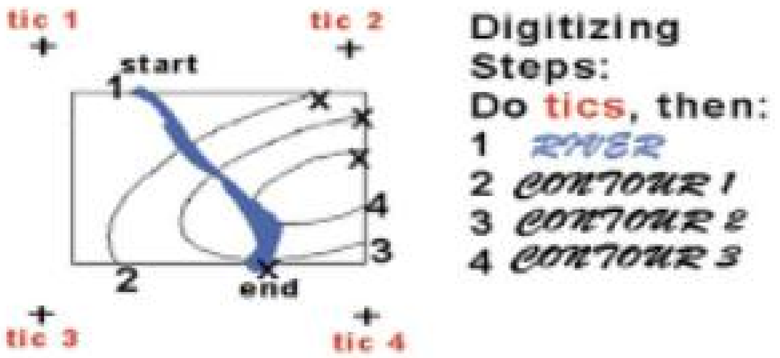
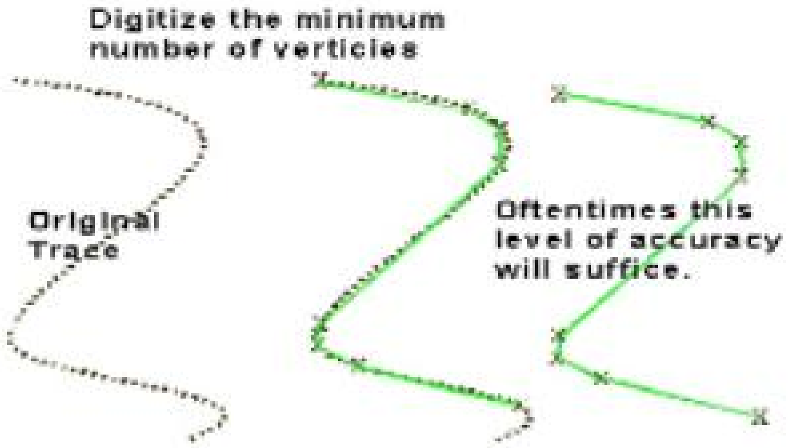


Fig 6.3 control point (TIC)

Prepare a systematic order for digitizing

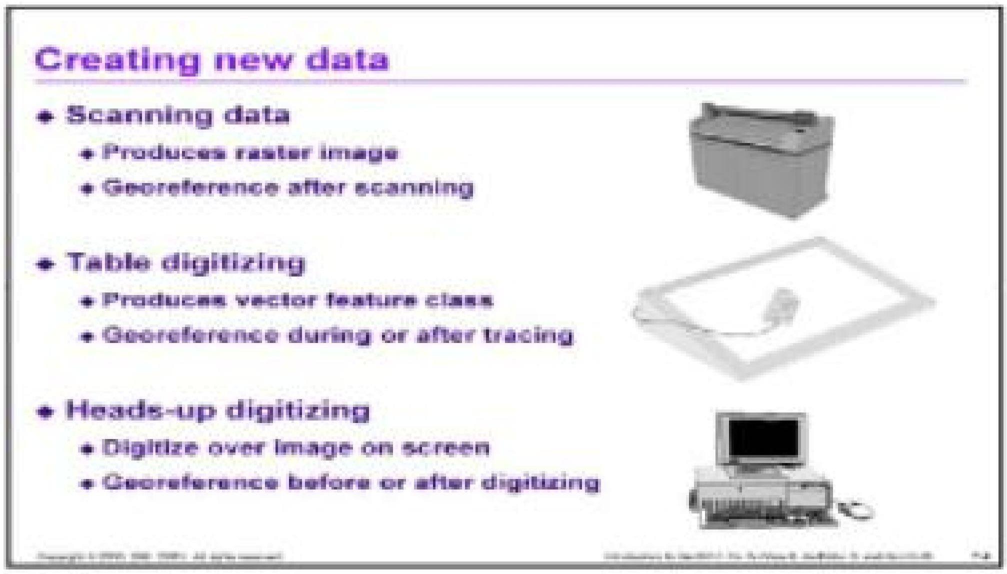


Here the **X** is a digitized point taken with "point mode" where instead one could have used "stream mode" and captured all the points represented by the dots, at a high cost in effort and storage without adding to the accuracy.



6.3 Creating new data (Digitizing teahouse)

Hardcopy spatial data can be digitally captured by scanning (taking a digital picture of the map), by digitizing (tracing the map with a digital tracing device), or a combination of the two, called heads-up digitizing.



6/3/1 Scanning

Scanning requires special hardware that creates a raster image of a hardcopy map. The resulting product may be converted directly to a grid or vectorized—converted to a vector coverage. ArcInfo provides special vectorization software for converting grids and images to coverages. An automatic vectorization tool is provided with the core ArcInfo package; more specialized tools are supplied with the GRID™ and ArcScan™ software extensions. Georeferencing of the image, grid, or resulting vector dataset must be done before geographic analysis can be performed.

6/3/2 Table digitizing

The most common digitizer used for mapping consists of a line wire grid embedded in the output of a table. Smaller digitizing tables are summations reffered to as digitizing. To digitize map must be taped down to a digitizing table and the points and lines traces with digitizer cursor or " Keypad" , the " Keypad" (or cursor) records the punition at which the crosshair intersect. The buttons of the keypad have been programmed tp perform factions such as recording a point or beginning a line. As keypad button is pressed the computer records the current x , y coordinate location of that position in digitizer units (usually cm or inches). This then becomes the x , y coordinate of the point feature or one of the points comprising a line or polygon.

Often, data is captured with special digital tracing hardware. Tablet digitizing produces a vector coverage, shapefile, or geodatabase feature class, which may or may not be georeferenced while tracing. Tics can be used to reference locations with known, real-world coordinates to locations on the digitizing tablet, allowing coordinates to be georeferenced during automation. Alternatively, features can be moved in coordinate space after tracing. Both methods will produce the same result.

- ***TIC points digitizing :***

A tic is a registration or geographic control point for a coverage.

Tics allow coverage coordinates to be registered to a common coordin transverse Mercator [UTM] meters, state plane feet, and so on) ar locations of features in a coverage to locations on the earth's surface. T registering map sheets during digitizing and editing.

Fig 6.4 Digitizer Table

6/3/3 Heads-up digitizing (On-Screen Digitizing)

Heads-up digitizing is a combination of scanning and manual digitizing. With heads-up digitizing, a map is scanned and the resulting image is traced on-screen with a mouse. The image may be georeferenced, producing a georeferenced dataset when traced, or the resulting dataset may be moved to real-world coordinate space after tracing.

6/4 Georeferencing your data

One important issue you must confront when adding data to your database is how to insure that all of the features in the individual datasets are properly aligned to locations on the earth's surface. In the following pages, you will learn how a master dataset of established ground control points can serve as the spatial reference template for the rest of the database.

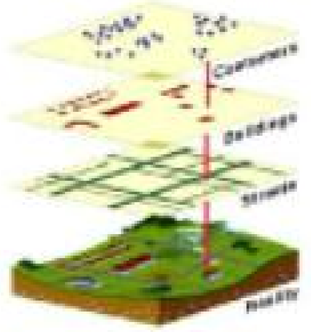


Fig 6.5 Georeferencing

6/5 Coordinate space

ArcGIS measures all coordinates in a unit less Cartesian coordinate system. Although a description of the measurement units can be stored with the data, most processing is done regardless of measurement units. In the graphic above, coordinate space is shown for an area measured in Universal Transverse Mercator (UTM) meters and another measured in UTM feet. Both boxes represent the same geographic location on earth. To ArcGIS, however, they do not exist in the same part of coordinate space. As a result of the different relative sizes of the two units, the data measured in feet has values roughly three times larger than those measured in meters.

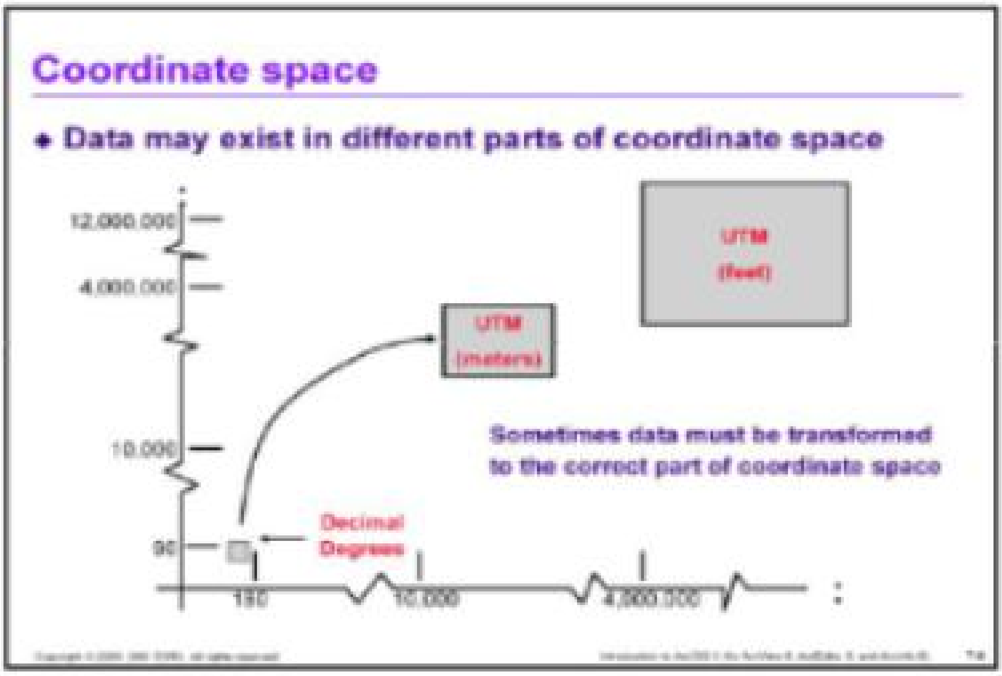


Fig 6.6 Coordinate Space

6/6 Spatial reference

When creating a new feature dataset or feature class, you must specify its spatial reference. The spatial reference for a feature class includes its coordinate system (for example, geographic, or a projected coordinate system such as UTM or State Plane) and its spatial or coordinate domain. The coordinate system is composed of a projection, a datum, an ellipsoid, and units. Other elements such as the prime meridian, a coordinate shift, and a zone may also be defined. The spatial reference also includes the coordinate extent or spatial domain of the feature class or the feature dataset. Extents can be set for the x-and y-coordinates as well as z, usually an elevation value, and m, a measurement along a linear feature.

The same spatial reference is shared by all the feature classes in a feature dataset. Any standalone feature class can have its own spatial reference associated with it. When importing spatial data (e.g., shapefile or coverage) to a feature dataset that has a different coordinate system, that data will automatically be projected to match the spatial reference of the feature dataset.

6/7 Digitizing Modes

Digitizing tablets generally operate in two modes: digitizing (absolute) mode and mouse (relative) mode.

When you are in digitizing mode, you can only digitize features; you can't choose buttons, menu commands, or tools from the ArcMap user interface because the screen pointer is locked to the drawing area. In mouse mode, however, there is no correlation between the position of the screen pointer and the digitizing tablet.

When digitizing, you can switch between digitizing mode and mouse mode using the Editing Options dialog box. This allows you to use the digitizer puck to digitize features as well as access user interface choices (as a substitute for the mouse). Also, you can use your mouse to choose interface elements at any time, whether your digitizer is in mouse mode or digitizing mode. You can digitize features on a paper map in two ways: point mode digitizing or stream mode digitizing (streaming).

(1) Point mode digitizing

- position the pointing device over the point element to be digitized.
- press the cursor button once.
- move the pointing device and a new point element and repeat the process. this mode is useful for individual locations (e.g. elevation benchmarks) as well as for straight lines that only require a few points to be digitized.Stream mode digitizing

(2) Stream Mode digitizing :

After pressing a button to begin the data collection, the digitizer continually collects points as the cursor is moved along a linear feature, until the operator presses another button to end digitizing .

(3) LINES

- place the cursor at the beginning of the line and press the button that initiates the recording of coordinates a start node will appear .
- move the cursor along the line at a fairly constant speed, following the curves in the line as carefully as possible
- at the end of the line, press the button to stop recording data coordinates, an end node will appear

for very long or complicated lines, digitize the line in portions. Trace a part of the line, then stop the data collection. Without moving the cursor, press the start button and recommence digitizing the subsequent portion of the line. Repeat this procedure until the line is complete.

(4) NETWORKS

– starting at one end of the network, digitize a line as described above

– where this first line meets another line, click to end digitizing. Then, without moving the cursor from this end node of the first line, begin digitizing the second line.

– repeat this procedure until the entire network has been digitized In order to be topologically correct, lines should never cross each other. A node should represent the intersection of two or more lines.

(5) Digitized as an POLYGONS/AREAS :

- also digitized in stream mode
- **digitized as an area edge or common line**
    1. using the previously marked starting point (on the Mylar sheet), trace out the feature and bring the cursor back to the starting point, before clicking to end the digitizing. The start node and end nodes should coincide.
    2. if the start and end nodes do not coincide, use the snapping or automatic closure menu

3. digitize each area edge only once, even if it is a boundary between two polygons. The topology about the nature of the adjacent polygons can be added in a later editing exercise

Practice on Building a master ground control feature class

(1) Building a master ground control feature class A master ground control feature class can be used to define a standard and uniform spatial reference for creating new feature classes. Establishing the coordinates for some known ground locations or ground control points can provide registration points for digitizing from a paper map, as well as providing a visual check that your data is georeferenced in the correct location. **What points can be used to establish ground control points** The known ground locations must be clearly identifiable features on the map. Some examples of points you could use to identify real-world locations are:

- Corners of the map sheet (latitude/longitude or projected units)
- Intersections of roads
- Control points located by survey markers
- Interpolated map reference points

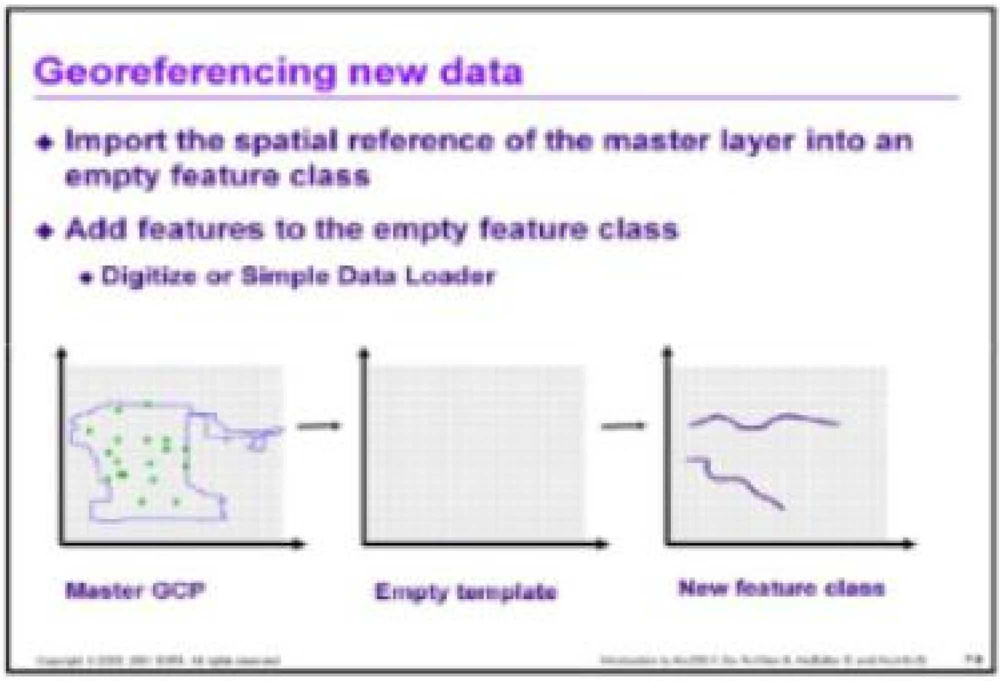
(2) Registering a paper map for digitizing

Registration points are points on the map for which you know the exact corresponding location on the ground. To register a feature class to the ground, you must digitally capture several registration points from your source map and store them along with their x- and y-coordinate values (ground locations). In an ArcInfo coverage, you refer to these registration points as tics. For more information on preparing to digitize and registering a paper map, see the online documentation.

(3) Georeferencing new data

When creating a new dataset, you can ensure new features will properly align with your existing data by importing the spatial reference of your ground control feature class into the new dataset.

By consistently referencing new datasets to a standard spatial reference, you can facilitate feature alignment between different datasets, as well as perform spatial analysis with results measured in real-world units.



(4) Tools for defining and changing a projection

ArcToolBox contains wizards for both updating the projection information associated with your spatial data and changing the projected coordinate system of your data. Defining a projection for your data does not alter the data in any way—it only tags on the projection information to the spatial data. Changing the projection will alter your data as it moves into a new projection.

4/1 Defining a projection

ArcMap expects coordinate system information to be stored with the data source. For a layer in a geodatabase, this information is part of the layer's metadata. For coverages and shapefiles, it is stored on disk in a separate file named after the data source but with a .prj file extension (for example, streets.prj). These files are optional files; thus, you may still need to define the coordinate system for one of these data sources. You can create a .prj file using the Define Projection Wizard in ArcToolBox.

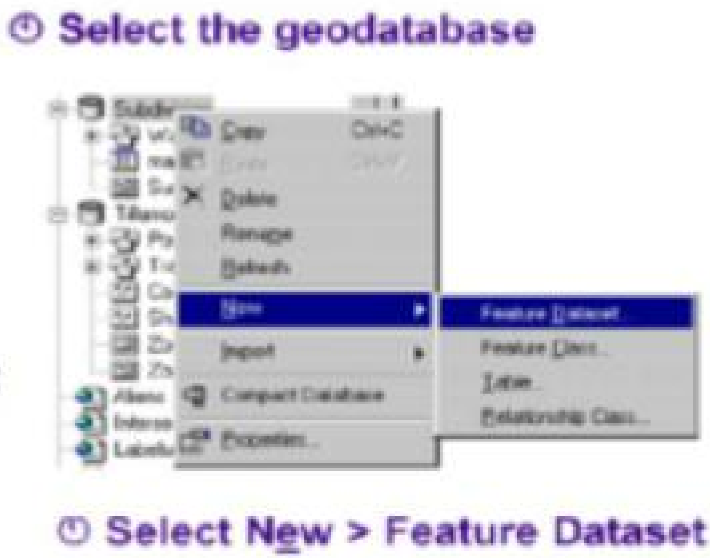
4/2 Changing the projection

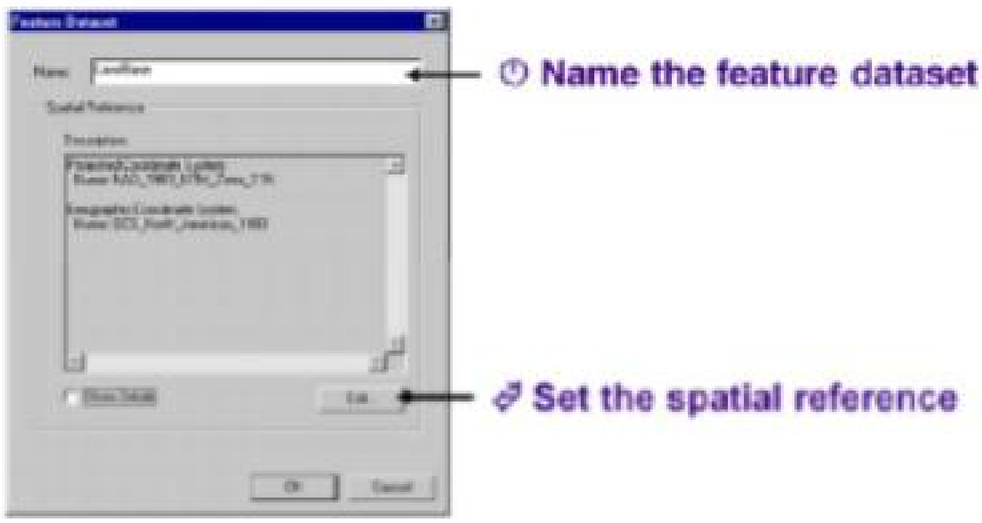
ArcGIS supports many different projections and is able to project your data from one projected coordinate system to another. This requires that the coordinate information is known for your data.

(5) Creating feature datasets

A new feature dataset in a geodatabase can be created in ArcCatalog by right-clicking on the geodatabase and choosing New>Feature Dataset. On the Feature Dataset dialog, fill in the name for the new feature dataset, and click the Edit button to set the spatial reference for the new feature dataset.

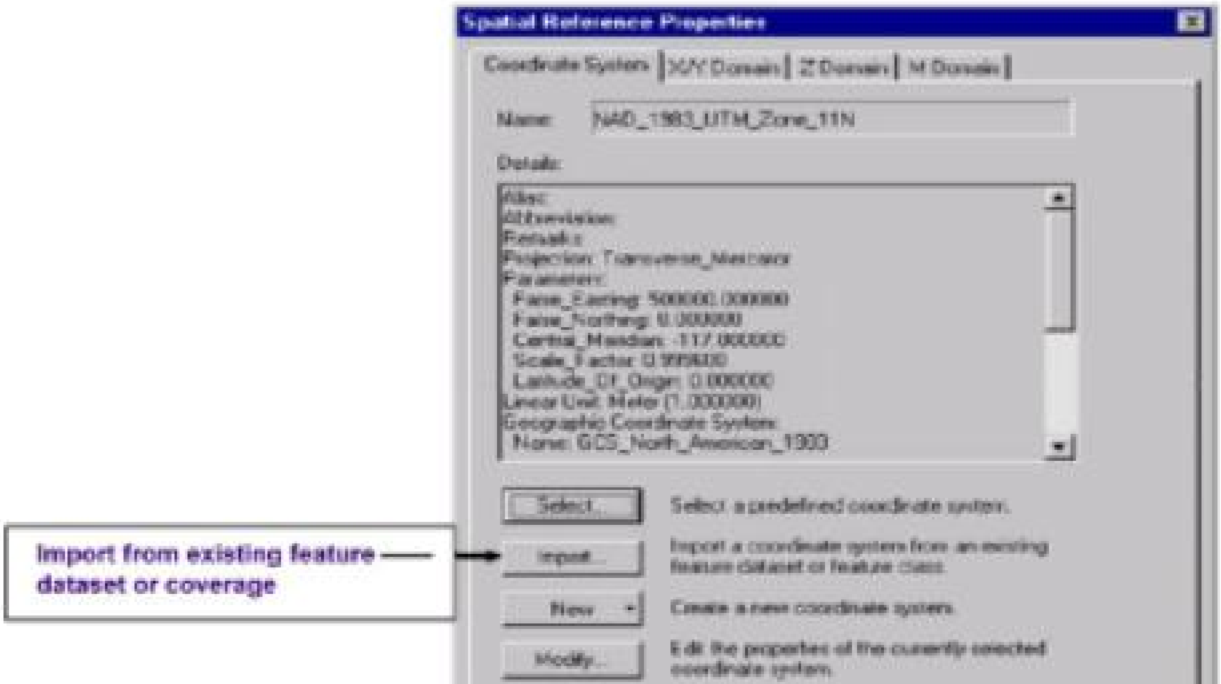
A feature dataset is a container for spatially or topologically related feature classes. The feature classes that will combine to form a geometric network or feature classes that share exact boundaries belong in a feature dataset. To maintain these topological relationships, all feature classes in a feature dataset must share the same spatial reference. The spatial reference can be set as a property of the feature dataset, guaranteeing that all feature classes you add to this feature dataset will automatically be projected to match the spatial reference.





(6) Setting the spatial reference

When setting the spatial reference for a brand new feature dataset or class, the Spatial Reference Properties dialog will appear blank. You have options to select a coordinate system from the many standard coordinate systems stored by ArcGIS, to import the spatial reference information from an existing source (perhaps another feature class), or to create a new coordinate system. Setting the spatial reference includes setting the coordinate system information and the domains for the x,y,z, and m coordinates. Z values usually store elevation values, while the m coordinate can be used to store a measure along a feature.



(7) Creating a new feature class

A new feature class can be created by right-clicking on the geodatabase in the Catalog tree, or by right-clicking on the feature dataset it will belong to. Next, you will define all its properties, the geometry type, the spatial reference, and the fields to store its attributes.

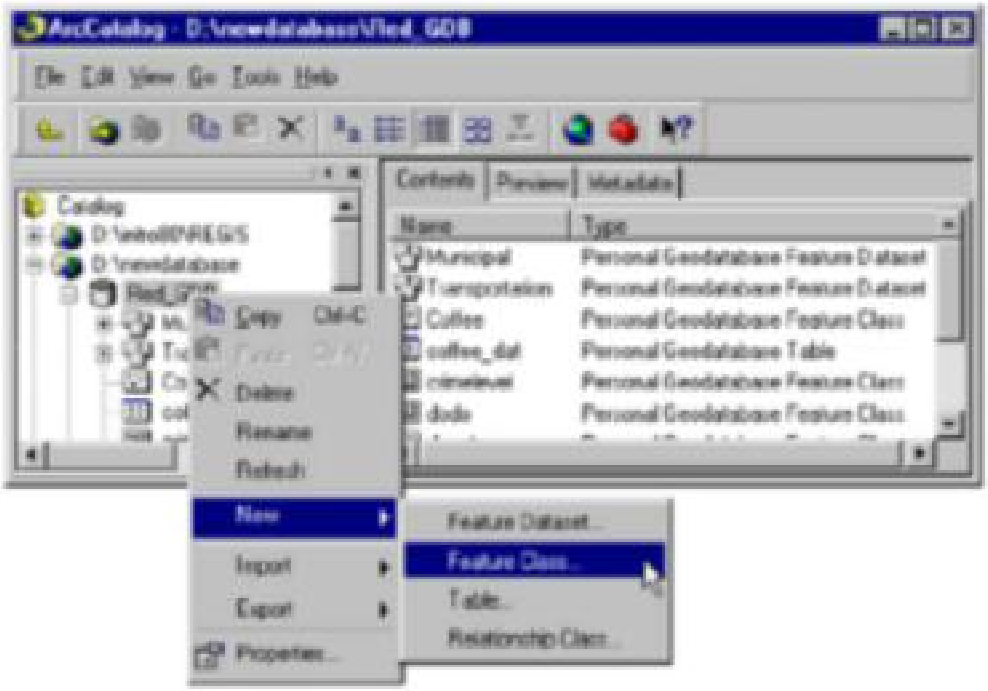


Fig 6.9 Creating a new feature class in ArcCatalog

When creating a feature class, you must define the geometry field's properties, such as its spatial reference and the geometry type: point, line, or polygon. Then, add fields for the attribute data by either defining them in the dialog, or importing the schema from an existing table.

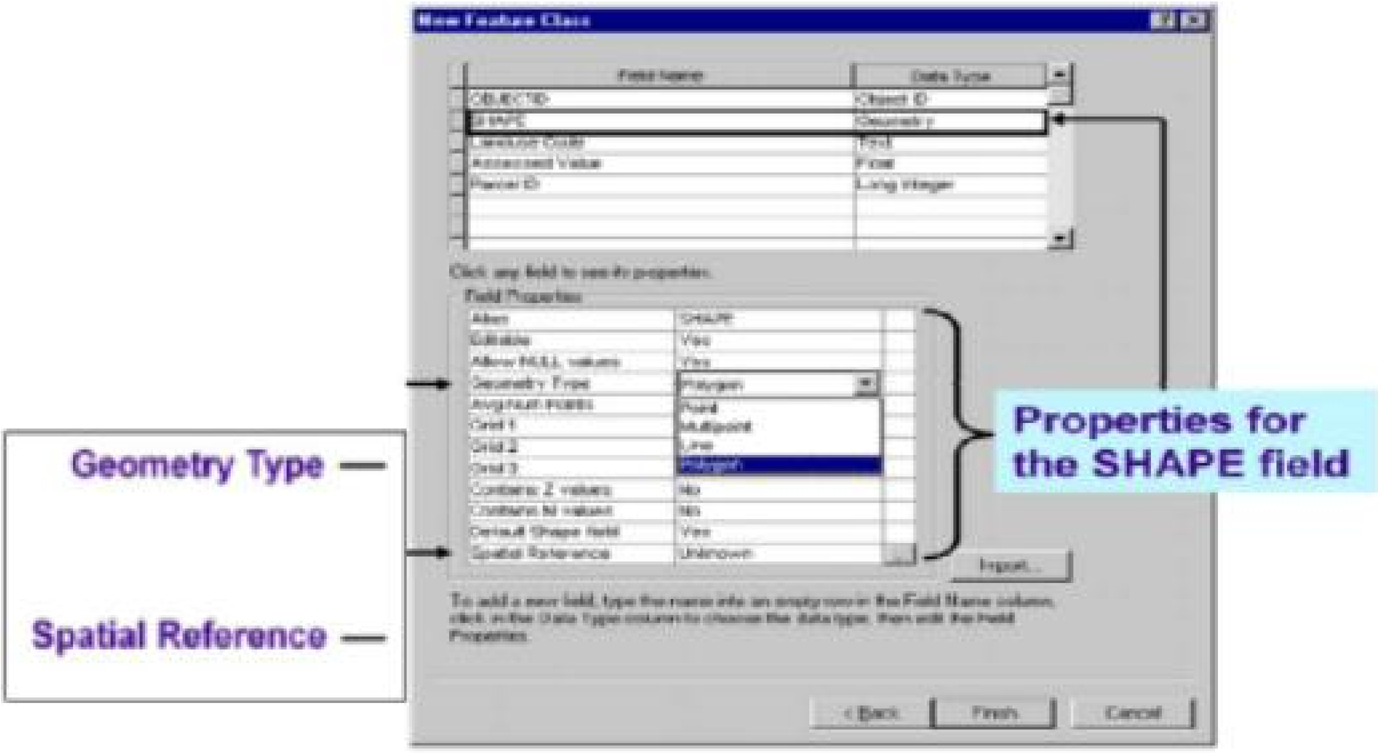
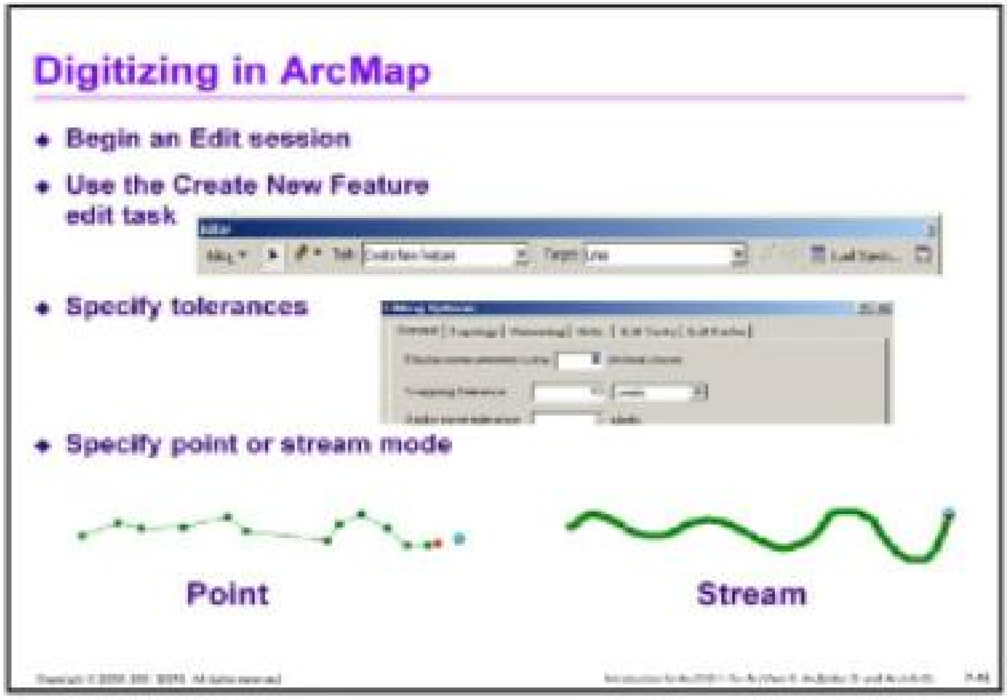


Fig 6.10 New Feature Class

(8) Digitizing in ArcMap

When adding new features to a feature class, follow these steps:

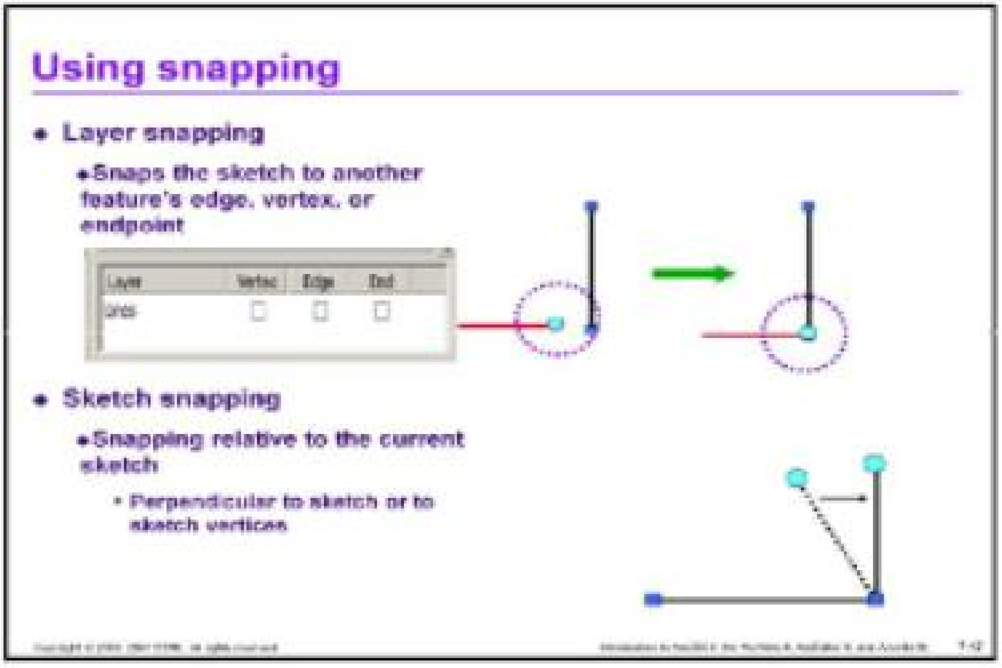
1. Begin an editing session and click the Create New Feature edit task on the Editor toolbar.
2. Specify an appropriate tolerance for snapping your pointer to other features.
3. Specify the mode for adding vertices. Point mode requires you to click the mouse for each vertex. Stream tolerance requires you to click the mouse to enter only the first vertex, then adds vertices at every interval of the stream tolerance.
4. Specify the stream tolerance—the minimum distance between vertices when adding features using stream mode.



(9) Snapping

You have probably heard the expression "Close only counts in horse shoes and hand grenades", but close can also count when you are editing in ArcMap, if you are using your snapping tolerance correctly. Snapping automatically moves your pointer to the location or orientation (parallel or perpendicular) of a feature. This prevents you from having to click the exact coordinate of a feature in order to connect the feature you are editing.

The snapping tolerance is the distance the pointer must be to a feature before it snaps to that location. You can express the snapping tolerance in map units or pixels.



Using snapping

There are two general varieties of snapping in ArcMap.

•Layer snapping: Moves your pointer to the location of a feature's edge, vertex, or endpoint. You can decide which, if any, of these options will be available for snapping.

•Sketch snapping: When adding a sketch, this option will snap your pointer to a location perpendicular to the previous sketch segment. The snapping can be overridden if you move your pointer far enough. Setting the proper tolerance for snapping is sometimes a trial and error process. You should try the default tolerance first, then change it in small intervals until you settle on one that is appropriate to the data and your needs.

6/8 DIGITIZING ERRORS :

There some basic types of digitizing errors :

1- *Extra arcs*
2- *Missing arcs*
3- *Dangling nodes*
4- *Pseudo nodes*
5- *Missing labels*
6- *Extra labels*

Post-digitizing tasks

- *Sometimes the computer screen that is being used to display the map being digitized has a lower resolution than the digitizer. In this case, it may be difficult to see objects that have been digitized very close together, until one zooms in to see whether they are separate lines.*
- *Zooming also helps in performing three other tasks that are necessary in creating an accurate digitized layer. They are:*
- *error identification and elimination*
- *editing*
- *labeling of features* [1]

[1] Each digitized point, line segment or area edge must have a label or identifier associated with it. These feature identifiers represent a category of a feature (e.g. "1" for highways, "4" for dirt roads) and can either be added using an EDIT or LABEL menu in the post-digitizing phase, or during digitizing.

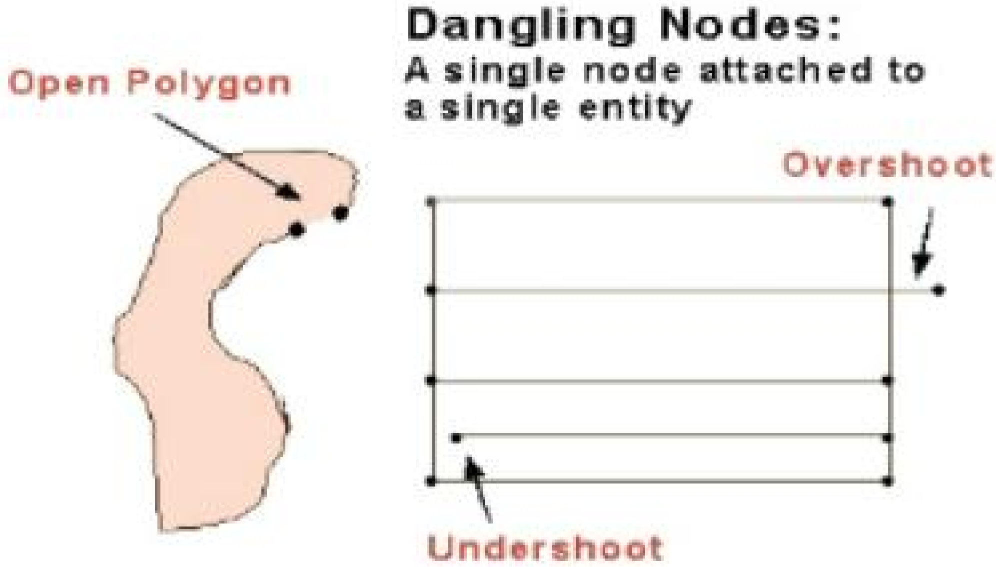
6/9 *Relationship between digitizing and editing*

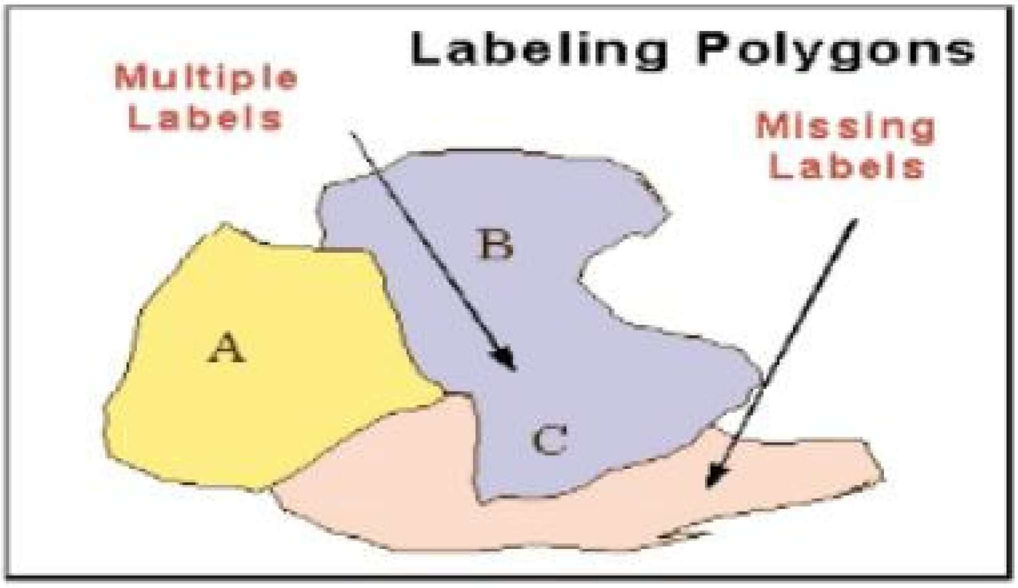
- digitizing and editing are complementary activities
  - poor digitizing leads to much need for editing
  - good digitizing can avoid most need for editing
  - both can be very labor-intensive

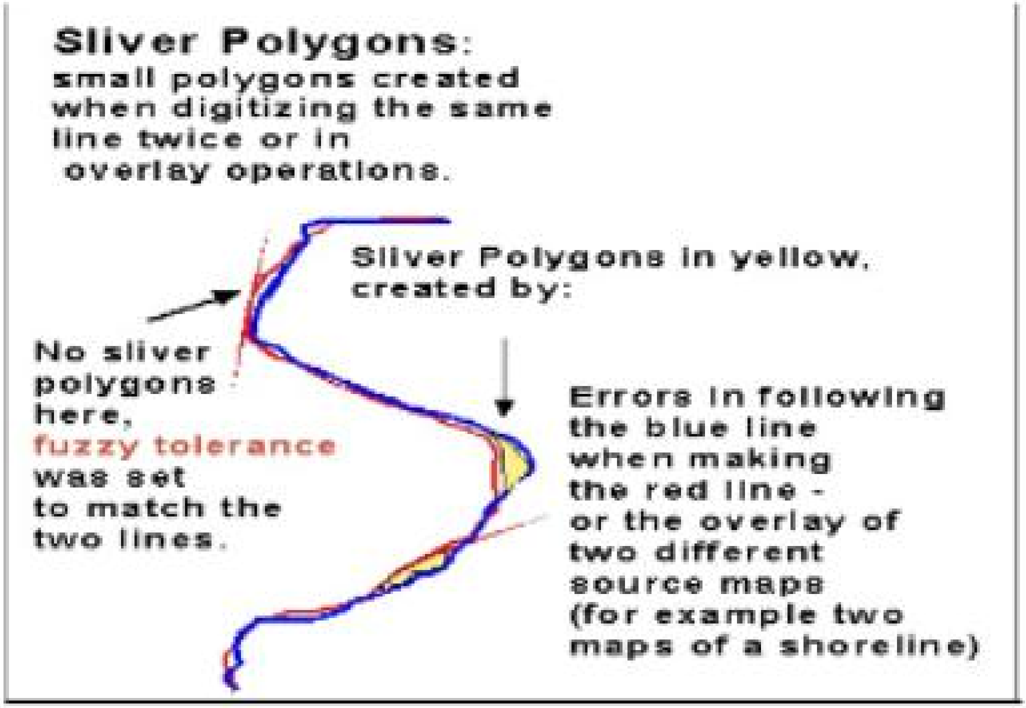
6/9 Editing of digitizing errors:

1. review the digitized maps, checking for errors.
2. remove duplicate lines where they occur.
3. snap nodes for polygons that should be closed.
4. snap lines to the relevant nodes where under- or overshooting occurs

Example of some digitizing errors

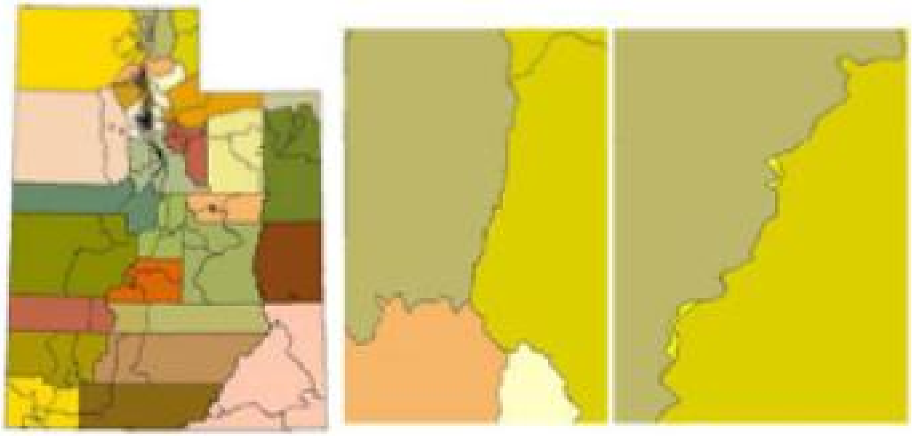



Labeling Polygons
Multiple Labels
Missing Labels
B
A
C


Sliver Polygons:
small polygons created when digitizing the same line twice or in overlay operations.
Sliver Polygons in yellow, created by:
No sliver polygons here, fuzzy tolerance was set to match the two lines.
Errors in following the blue line when making the red line - or the overlay of two different source maps (for example two maps of a shoreline)

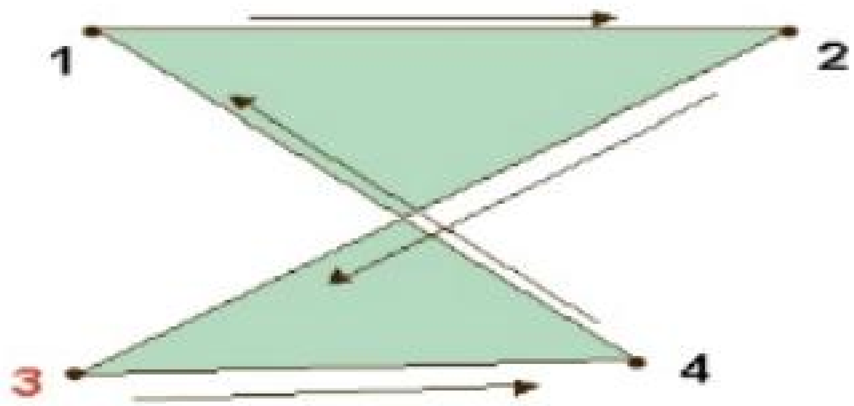
Sliver polygons

·When two arcs are added next to each other, you may have some "unintended" polygons added in



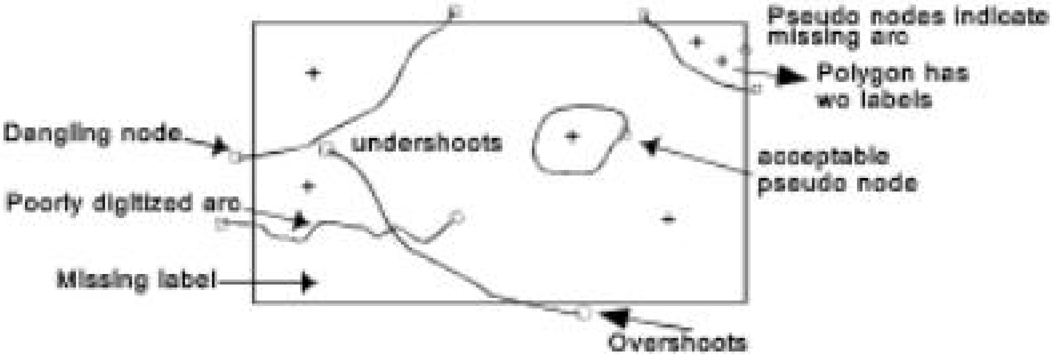
Weird Polygons

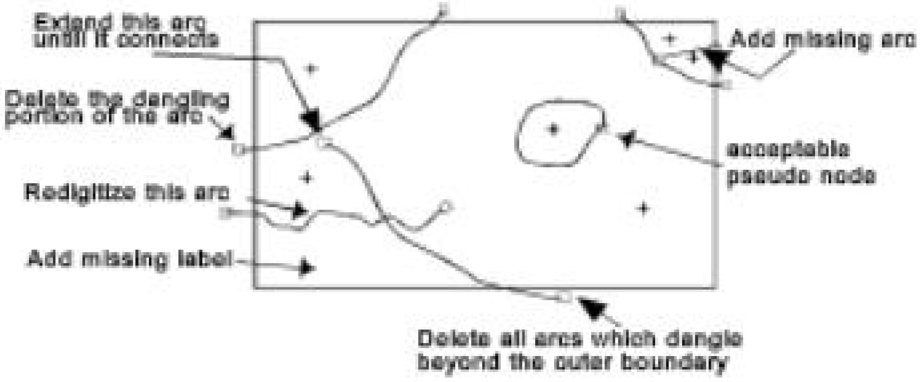
Polygons with missing or misordered nodes



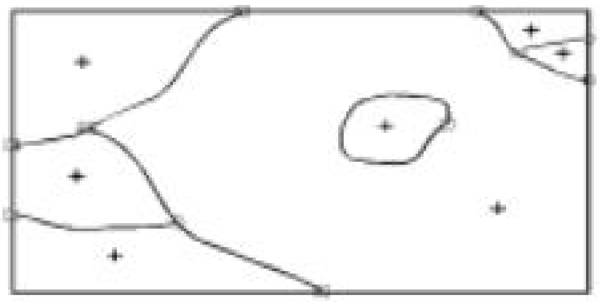
Example 1 :

Some errors as they interpreted by you after comparing them with the source map





Showing a set of errors marked to be fixed and how to fix them



The same set of Data after the spatial errors have been corrected.

Chapter 7
Building Topology

7.1 Topology Definition :

The Science of mathematics of relationships used to validate the geometry of vector entities, and for operations such as network tracing and tests of polygon adjacency.

The study of geometric properties that do not change when the forms are bent, stretched or undergo similar geometric transformations.

The primary purpose of topology is to define spatial relationships between features in one or more feature classes. Incorporating topology into your datasets allows you to better model the real world. The primary spatial relationships you want to model are adjacency, coincidence, and connectivity.

Adjacency

Adjacency allows you to identify which land owners or soil types share a common boundary with each other.

Coincidence

Coincidence lets you identify the bus routes on top of roads.

Connectivity

Connectivity allows you to identify a route to the airport, or connect streams to rivers, or follow a path from the water treatment plant to a house.

7.2 Importance of Topology in Vector Format

Thus far we have presented vector data simply as points, lines, and polygons with associated attributes. This provides for location and meaning. However, it is also important to understand the spatial relationship between spatial objects. This is called topology.

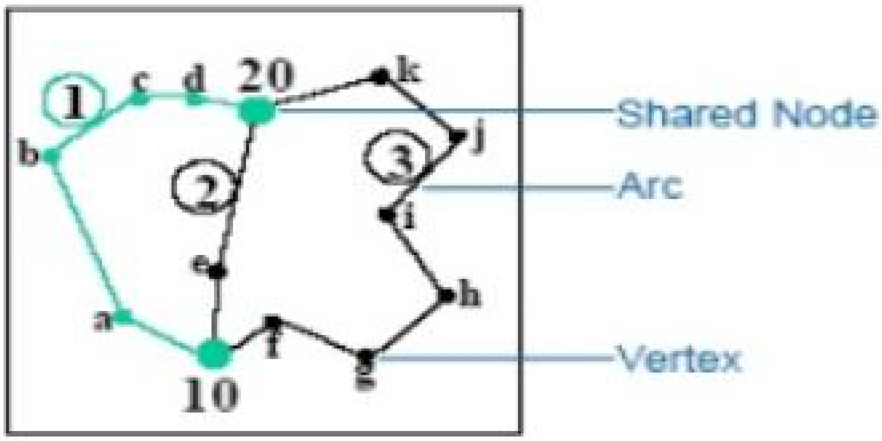
7.3 Models of Spatial Relationship between features :

There are three major topological concepts :

- *Arcs connect to each other at nodes (connectivity)*
- *Arcs that connect to surround an area define a polygon (area definition)*
- *Arcs have direction and left and right sides (contiguity)*

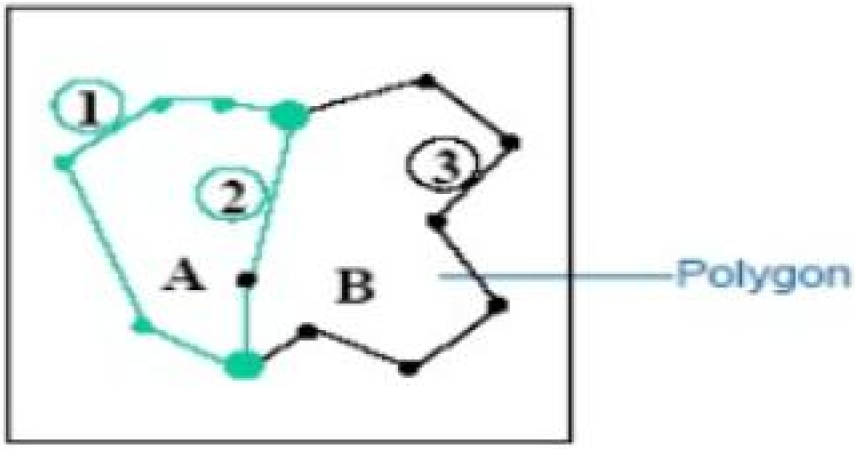
(1) Connectivity

Storing connectivity by recording the nodes that mark the end points of arcs (lines) is useful for modeling and tracing flows in linear arcs. Arcs that share a node are connected. This is called Arc-Node topology



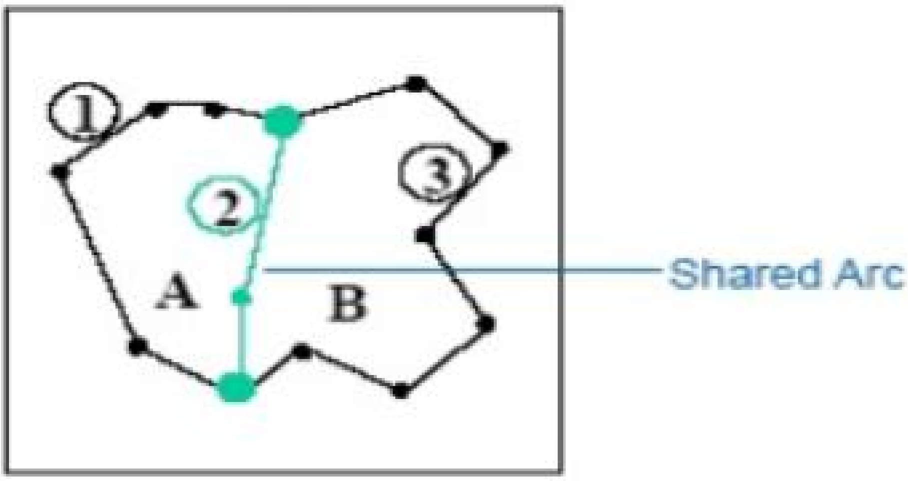
(2) Area Definition

Coverages define areas by keeping a list of connected arcs that form the boundaries of each polygon. This is called Polygon-Arc topology.

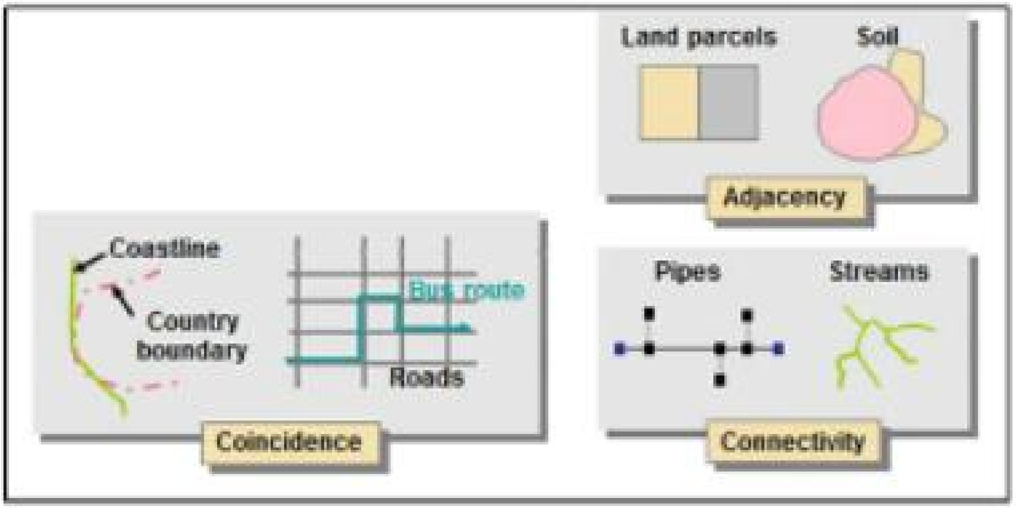


(3) Contiguity

Coverages store contiguity by keeping a list of the polygons on the left and right side of each arc.



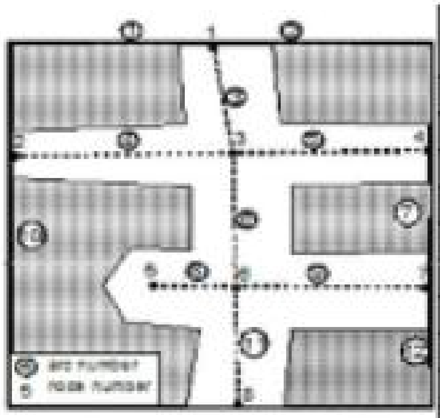
Some models of Spatial relationship between features :



Example 1:

- Arc-node topology (connectivity)
- Arcs are used to represent street centre lines. Nodes are located at street intersections. By definition the points (x, y pairs) along the arc, called *vertices*, define the shape of the arc. The ending points of

the arc are called *nodes*. Each arc has two nodes; a from-node and to-node. Arcs can join only at their end points, or nodes. In the above illustration, arcs 6, 3, 4, 8 all join at node 3. The computer now knows that it is possible to travel along arc 3 and turn onto arc 8 because they share a common node 3, but it is not possible to turn directly from arc 3 onto arc 9 because are 3 and arc 9 do not share a common node.



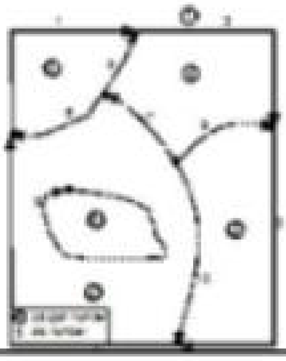
Arc-node topology (connectivity)

Arc	From-node	To-node
1	2	1
2	1	4
3	1	3
4	2	3
5	4	3
6	3	6
7	4	7
8	5	6
9	6	7
10	2	8
11	6	8
12	8	7

Example 2:

Area definition (Polygon – arc Topology)

Polygons are represented by a series of x, y coordinates that connect to enclose an area. Some system stores polygons in this format. ARC/INFO however, stores the arcs defining the polygon rather than a closed set of x, y pairs. A list of arcs that make up each polygon is stored and used to construct the polygon when necessary. In the above illustration, arcs 4, 6, 7, 10 comprise polygon 2 (the 0 before the 8 indicates that this arc creates an island in polygon 2).

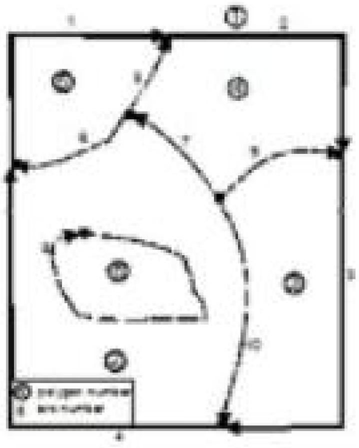


Polygon No.	Arc No.
٢	[illegible]،٤،٦،٧،١٠
٣	٨
٤	٩ ،١٠ ،٣
٥	٤ ،٢ ،٥ ،٧
٦	٦ ،٥ ،١
	٨

Example 3 :

Contiguity (Left-right Topology)

Because every arc has direction (a from-node and a to-node); ARC/INFO maintains a list of polygons on the left and right sides of each arc. Thus any polygons sharing a common arc are adjacent. In the above illustration, polygon 2 is on the left of the arc 6, and polygon 5 is on the right. Notice that the label of polygon 1 is outside the boundary of the area. This polygon is called the *external* or *universe* polygon, and represents the area outside all the polygons in the map.



Arc No..	L Poly.	R Poly
١	٥	١
٢	٤	١
٣	٣	١
٤	٢	١
٥	٤	٥
٦	٥	٢
٧	٤	٢
٨	٦	٢
٩	٣	٤
١٠	٢	٣

7/4 Arc/INFO coverages :

In Arc/INFO, all spatial information has topological information attached to it

- Containment, connectivity, adjacency

to correctly derive this, all coverages are subject to planar enforcement

- All intersections of arcs must be joined by nodes
- All polygons must be closed

Labeling polygons :

A polygon is formed of one or more arcs

The computer uses an algorithm to build the arcs into a higher level object (polygon). The build process can be processor intensive, so it is only done when the user asks for it .

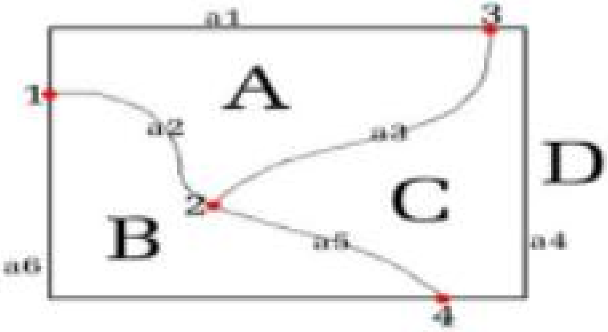
the computer wishes to attach the information to the polygon, but attaching an attribute to a variable number of arcs is messy . A single label point is added to the polygon.

That label point is the object upon which all attributes get attached.

Feature No.	Road-type	Surface	Width	Lanes	Name
1	1	Asphalt	40	4	Galaa St
2	1	Asphalt	40	4	Gehan St
3	1	Asphalt	40	4	Bahr St
4	2	Gravel	20	2	Bostan St
5	4	Gravel	20	2	Ferdous St
6	3	concrete	30	3	Mubarak st

Record

Item



Polygon Topology		Node Topology		Arc Topology	
Polygon	Arcs	Node	Arcs	Arc	Left & Right Polygons
				a1	A D
A	a1, a2, a3	1	a1, a2, a6	a2	A B
B	a2, a5, a6	2	a2, a3, a5	a3	A C
C	a3, a4, a5	3	a1, a3, a4	a4	C D
D	a1, a4, a6	4	a4, a5, a6	a5	B C
				a6	B D

7/5 Spaghetti versus Topological data

A) 'Simple' spaghetti data

Vector data that has been created without topology is referred to as 'spaghetti' data for reasons you can imagine (strings of unconnected lines). This is easier to create, but if to be used for GIS, one pays for lack of topology later: Individual features may appear the same, for example:

- Points: have x and y coordinates.
- Lines (arcs): Strings of x, y vertices.
- Polygons: Closed set of coordinates.
- But there is NO spatial relationship between these features:
- Arcs may not necessarily join and Polygons may not close to form areas.
- Intersections may not have nodes where two arcs cross.
- Adjacent digitized polygons may overlap
- Arcs may consist of many broken segments.

B) Topological data

Creating topologically correct data takes longer, but enables GIS queries and analysis.

1- Points: are polygons of zero area and length.

2- Lines (arcs): start and end at nodes.

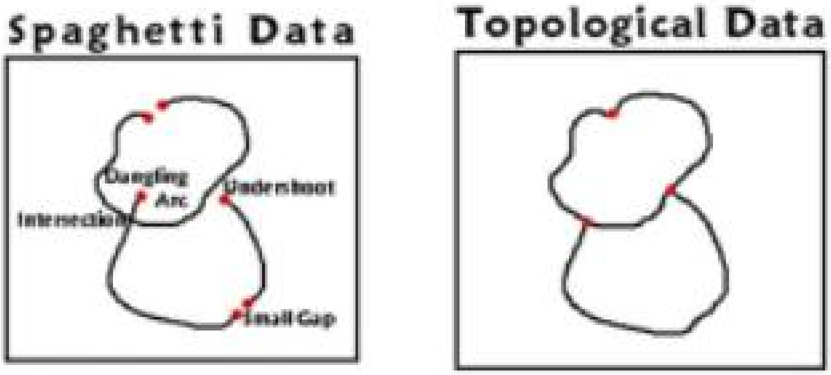
3- Polygons: given by sets of connected arcs and an interior label point.

4- Shared polygon arcs result in:

a) Lower total number of arcs in a database.

b) Adjacent polygons do not enclose overlap wedges or slivers.

c) Cleaner map output (more evident when you zoom in or magnify).



7/6 Creation of Topology: 'Clean & Build'

	EXAMPLE	DEFINITION	ACCEPTABLE
NORMAL NODES		At an intersection of 3 or more arcs	Always
DANGLING NODES		At the end of an arc	Arcs (Not polygons) e.g. roads, streams
PSEUDO NODES		Between 2 arcs	Island polygons, attribute change

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

Chapter 8
Working with Tables and Attribute Data

A table contains formatted descriptive information. In ArcGIS, the information in a table is generally associated with spatial data, such as a feature attribute table, but can also be independent of any spatial data (e.g., nonspatial statistical data). For the purposes of this course, you will consider only tables that contain information related to spatial datasets.

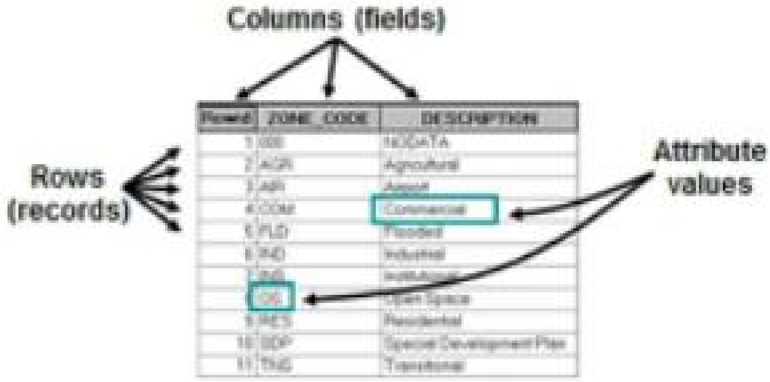
The feature attribute table contains descriptive information about the features in a feature class. To open a feature attribute table in ArcMap, right-click the layer, then click Open Attribute Table. In ArcCatalog, you highlight the table and view it using the Preview tab.

The feature attribute table consists of fields (also known as columns or items). Each field represents one type of descriptive information. Each row (also known as a record) contains the attributes of one feature in the dataset.

8.1 Understanding table anatomy

Each table has the same basic format: an array of rows and columns. The intersection of a row and a column represents a specific attribute for a single feature.

Some tables, like a feature class's default attribute table, have a preset number of columns. For instance, a polygon coverage has four standard columns: area, perimeter, <coverage>#, and <coverage>-id. Other tables are completely defined by the user. The INFO table in the graphic has two user-added columns: ZONE_CODE and DESCRIPTION. ArcGIS automatically adds the first column for displaying the record number. For a feature class, this column is the FID (Feature ID); for tables in the geodatabase, it is the OID (Object ID); for INFO tables, it is called the Rowid. These



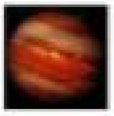
Each table has the same basic format: an array of rows and columns. The intersection of a row and a column represents a specific attribute for a single feature.

Some tables, like a feature class's default attribute table, have a preset number of columns. For instance, a polygon coverage has four standard columns: area, perimeter, <coverage>#, and <coverage>-id. Other tables are completely defined by the user. The INFO table in the graphic has two user-added columns: ZONE_CODE and DESCRIPTION. ArcGIS automatically adds the first column for displaying the record number. For a feature class, this column is the FID (Feature ID); for tables in the geodatabase, it is the OID (Object ID); for INFO tables, it is called the Rowid. These columns (FID, OID, or Rowid) are not physically stored as part of the table, but you can perform queries on them as you would any other column.

8.2 Tabular data field types

Tables are capable of storing date, number, and text values, but most tabular formats have several different field types to store this information. Choosing the best field type for the values to be stored is an important consideration. In addition, the available field types can vary between tabular formats. Supported formats in ArcCatalog include short integer, long integer, float, double, text, date, and blob. Consult the online documentation for more information about these column types.

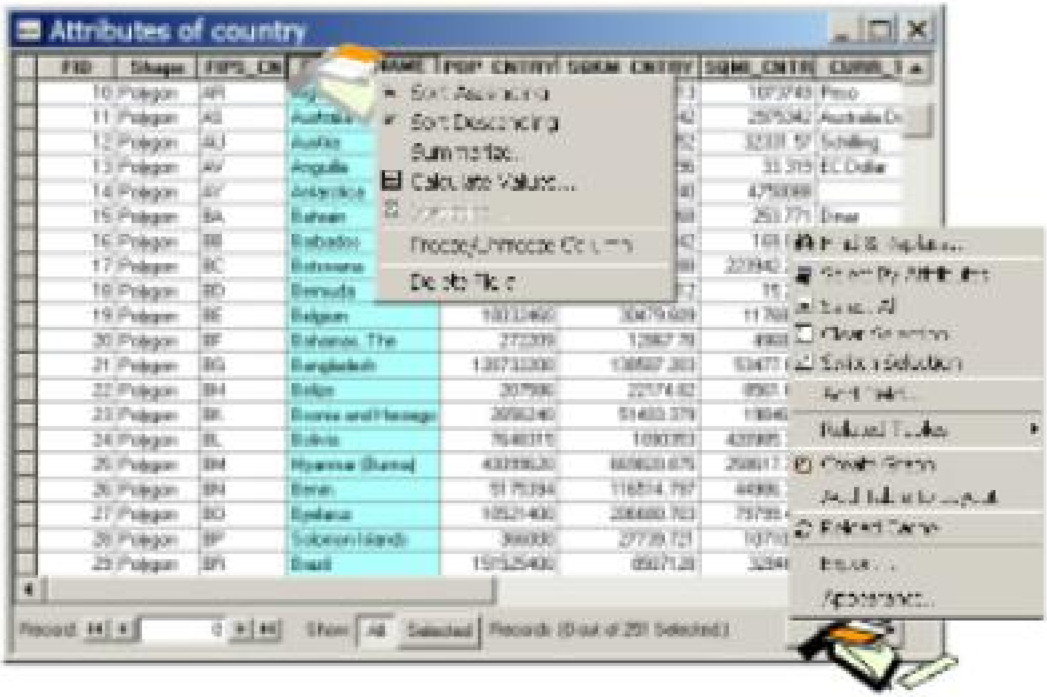
Name: Jupiter
Moons: 16
Diameter: 142,984 km
Date of Comet Shoemaker-Levy impact: 7/16/1994
Rotation period: 9.8 hr



Text	Date	Short	Long	BLOB	Float
Jupiter	7/16/1994	16	142984		9.8

8.3 Table manipulation :

Many operations, such as sort, find, select, and freeze, can be performed on the tables in ArcMap and ArcCatalog. In ArcCatalog, you can create new tables as well as create and delete fields for existing tables. In ArcMap, you can edit the table record values.

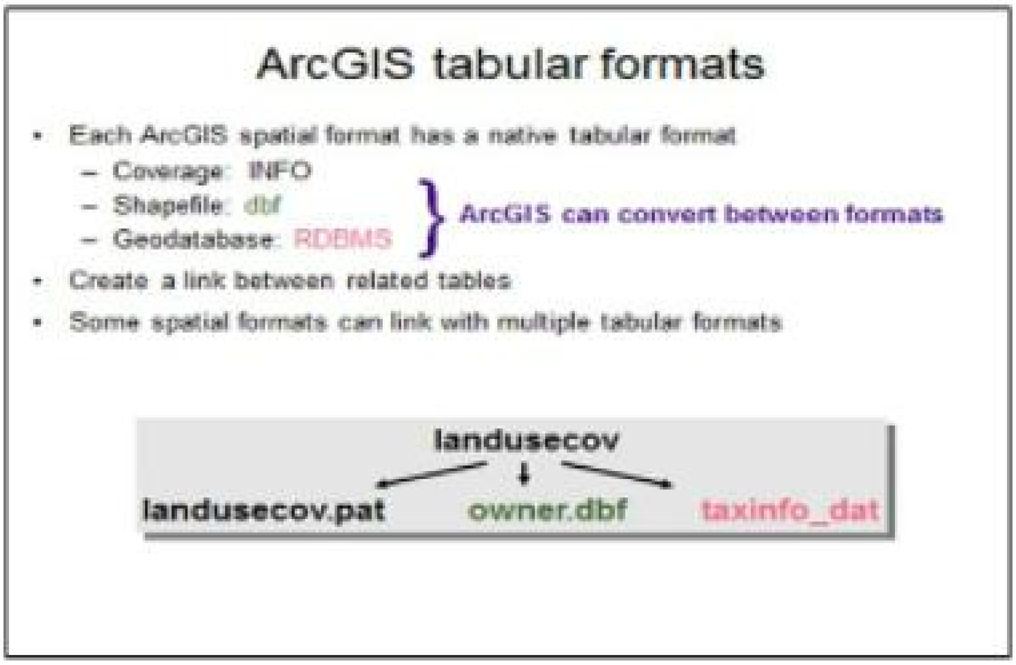


8/4 ArcGIS tabular formats

ArcGIS supports the use of multiple formats for the storage and management of tabular information. Each of ArcInfo's primary spatial formats has its own native format. Coverages use INFO formatted tables; shapefiles store their attributes in dBase (dbf) format; geodatabases rely on the format of their supporting RDBMS (e.g., Access or Oracle).

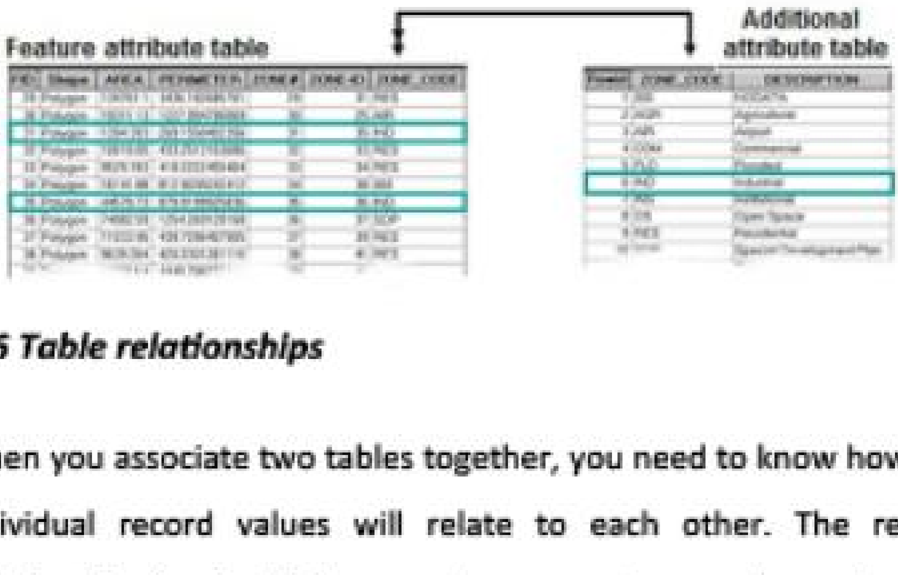
Some formats, such as the coverage, can link to independent tables regardless of their format.

Deciding on the proper format to store attribute information is an important part of database design and can affect how efficiently you access tabular information. To facilitate sharing of data in different formats, ArcGIS contains tools to convert between the various tabular formats.



8/5 Associating tables

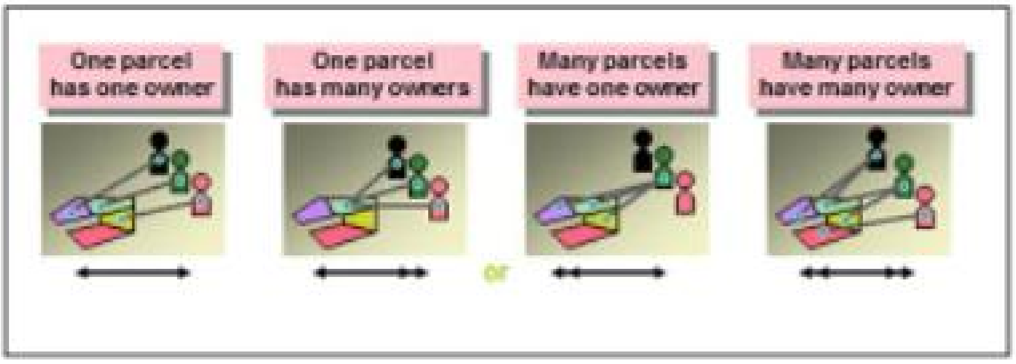
Two tables can be connected, or associated, if there is a similar field in each table containing common values. In the example, the tables are linked by the common field called ZONE_CODE. Once the tables are connected, the description of what each zone code means (from the DESCRIPTION field) can be accessed from the feature attribute table by looking up the value in the associated table. Often features have many attributes, so most database design guidelines promote organizing your database into multiple tables—each focused on a specific topic—instead of one large table containing all the necessary fields. This scheme prevents duplicate information in the database, because you store the information only once, in one table. Tables can be connected so that when you need information that isn't in the current table, you can access it from an associated table.



8/6 Table relationships

When you associate two tables together, you need to know how the individual record values will relate to each other. The record relationships (cardinality) are: one to one, one to many (many to one), and many to many.

Knowing which relationship type you have will prevent potential record omission errors. For instance, if you have a one-to-many association and you connect the tables as if they were a one-to-one association, you will omit needed information from the connected table because one-to-one record searches stop looking for more matches after they find the first match. You can avoid these types of problems by becoming familiar with tabular database management strategies—strategies that will also make you a better GIS user. Read the sections about tabular management in the online documentation.



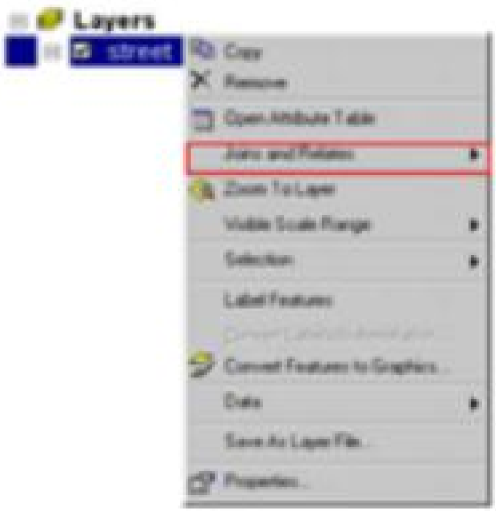
8/6/1 Joins and relates

ArcMap provides two methods to associate data stored in tables with geographic features: joins and relates. When you join two tables, you append the attributes from one onto the other based on a field common to both tables. When you relate tables, you define a relationship between the two tables—also based on a common field—but don't append the attributes of one to the other. Instead, you can access the related data when necessary.

You will join two tables when the data in the tables has a one-to-one or a many-to-one relationship (e.g., you have a layer showing store locations, and you want to join a table of the latest monthly sales figures to it).

You will relate two tables when the data in the tables has a one-to-many or many-to-many relationship (e.g., your map displays a parcel database, and you have a table of owners; a parcel may have more than one owner, and an owner may own more than one parcel).

Joins and relates are reconnected whenever you open the map. This way, if the underlying data in your tables changes, it will be reflected in the join or relate. When you're through using a join or relate, you can remove it.

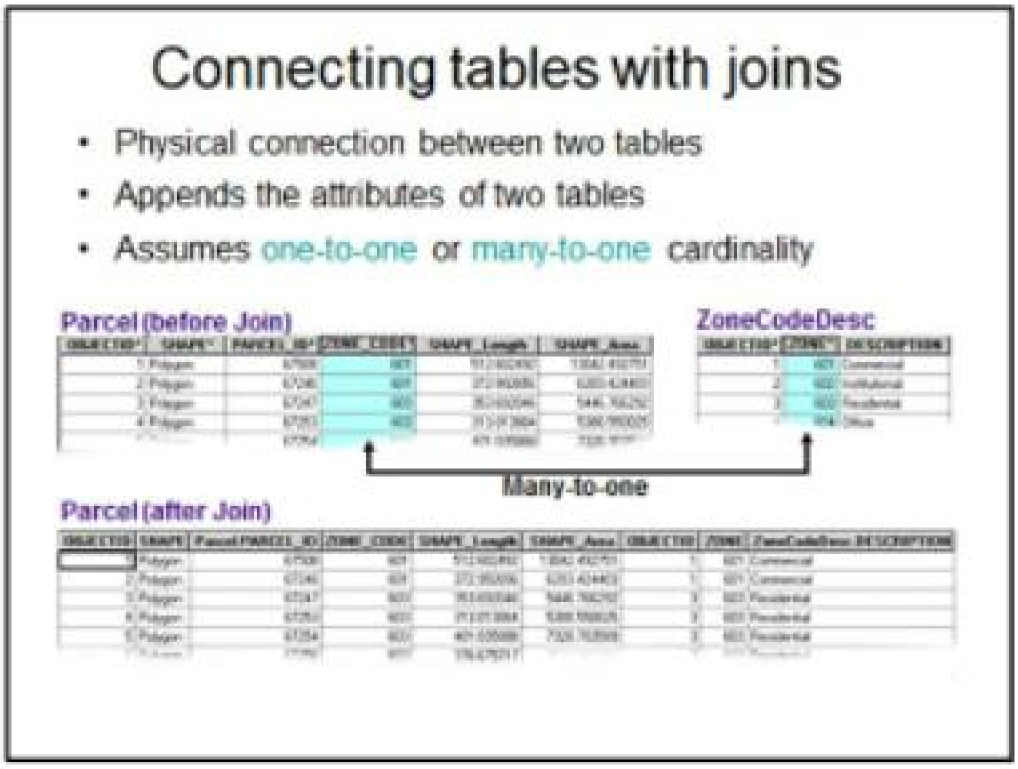


8/6/2 Connecting tables with joins

You can connect two tables in ArcMap using a join. Joins work with shapefiles, coverages, and geodatabase files. Once the tables are connected, you can query, symbolize, or analyze the new table based on the joined values.

Tables connected by an ArcMap join are not permanently connected. The tables are dynamically linked together in ArcMap and you can remove or add them whenever you want. When two tables are joined, the names of the common fields don't need to be identical, but the fields must be the same type (e.g., text, date, float).

Table joins are designed for one-to-one or many-to-one relationships. For other cardinalities you should use a relate instead of a join. If you use join with one-to-many or many-to-many cardinality, you will omit all records after the first match for each primary key value.

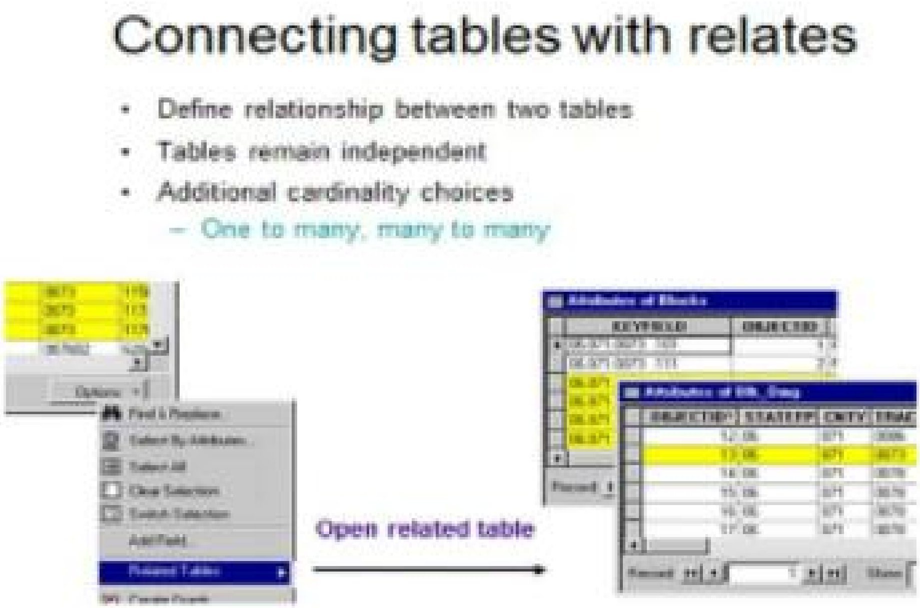


8/6/3 Connecting tables with relates

Another way to connect tables in ArcMap is by creating a relate. Like joining tables, relating tables defines a relationship between two tables and is also based upon a common field.

Unlike joining tables, a relate doesn't append the fields of one table to the other. Instead, the two tables remain as independent tables in ArcMap. ArcMap knows the two tables are connected and you can access data in the related tables when you need it.

You relate tables instead of joining them when there is a one-to-many or many-to-many relationship between the tables, or when you need to maintain the information in the related table independent of the attribute table.



8/7 Graphs :

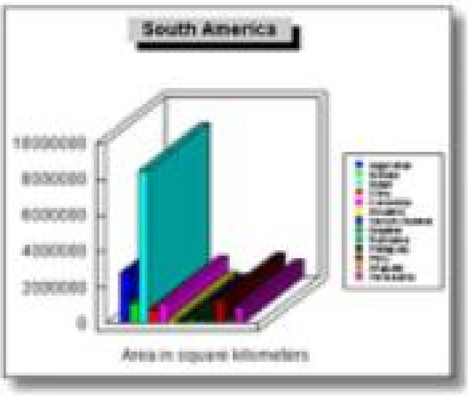
By displaying data values graphically, graphs simplify the often difficult task of interpreting the large amount of quantitative (numerical) attribute data associated with layers.

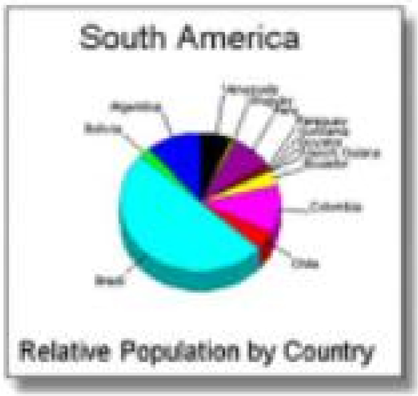
You can represent your data and analysis results using many styles of graphs including two- and three-dimensional graphs. ArcGIS uses graphics server software that provides a variety of chart types so you can represent your data in the clearest and most efficient manner.

Values for ArcGIS graphs come directly from feature attribute tables. Some graphs are better than others at presenting certain kinds of information. Carefully consider the information you want to present before choosing a graph style.

You can control most visual aspects of the graph in order to create an effective display of your data. For example, you can add titles, label axes, change the color of graph markers, or change the color and font of the chart's text.

Once you've created a graph, you can add it to a map in ArcMap's Layout view. When placed on the layout, a graph becomes a graphic element that you can size and position as desired.

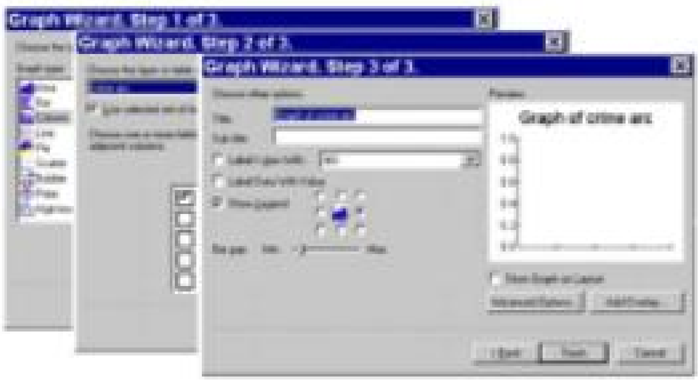




Graph Creation :

The graphing tool in ArcMap can be activated in the Tools menu using the Graphs option. You can then create, manage, or load a graph. A series of wizard panels guides you through the process of creating a graph.

Graphs present information about map features—and the relationship(s) between them—in an attractive, easy-to-understand graphic. They may show trends and/or patterns that are not easily visible in the attribute tables of the map features. They can show additional information about the features on the map or show the same information in a different way. The information displayed on a graph comes directly from the attribute information stored with your geographic data. Once a graph is created, adding it to your map is easy.



To add additional attributes for each feature, you will:

a) Create a new data file to hold the attributes

b) Add the attribute values to the data file

c) Join the data file to the feature attribute table for the coverage

Step 1: Creating a new data file to hold the attributes:

As mentioned above, attribute information is stored in a tabular database file called *Feature Attribute Table*. For each geographic feature (point, line, polygon) there is one entry record, in the file. For each record, there are a number of kinds of information, or items.

Step 2: Add the attribute values to the newly created data file:

If the attribute values you are on the list, you can type them directly into the data file on the computer. The attribute values are already in a file on the computer, you may be able to put them directly into the data file without having to retype them.

Step 3: Relate or join the attributes to the feature attribute table

Once the attribute values are added to the data file, you can attach them to the feature attribute table for the coverage using a common item as a key.

Since the records in the feature attribute table can be linked to corresponding records in the new data file, the new attributes will

also be associated with the features. Any two tables can be connected if they share a common attribute.

Parcel No.	Owner		Parcel No.	Zoning	Area
11-100 11-002 11-300	Hassan Ibrahim Mohamed		11-100 11-002 11-300	Residential Residential Commercial	15,900 12,100 19,200

Whereas a *Relate* temporarily connects two attribute tables,

a *relational join relates* and merges two attribute tables using their common item (Parcel No.).

Parcel No.	Owner	Zoning	Area
11-100 11-002 11-300	Hassan Ibrahim Mohamed	Residential Residential Commercial	15,900 12,100 19,200

Chapter 9
GIS Query

A GIS must provide utilities for finding specific features based on location and attribute value. One common types GIS query is to determine what exists at a particular location. In this type of query the user understand where the feature of interest are. But wants to know characteristics are associated with them. This can be accomplished with GIS because the spatial features are linked to the descriptive characteristics.

Another type of GIS query is to determine the locations that satisfy certain conditions. In this case the user knows what characteristics are important and wants to find out where the features are that have those characteristics.

The most common query tools are :

a) Examining GIS Data are identify, find, measure, map tips, hyperlink
b) Selection tools.

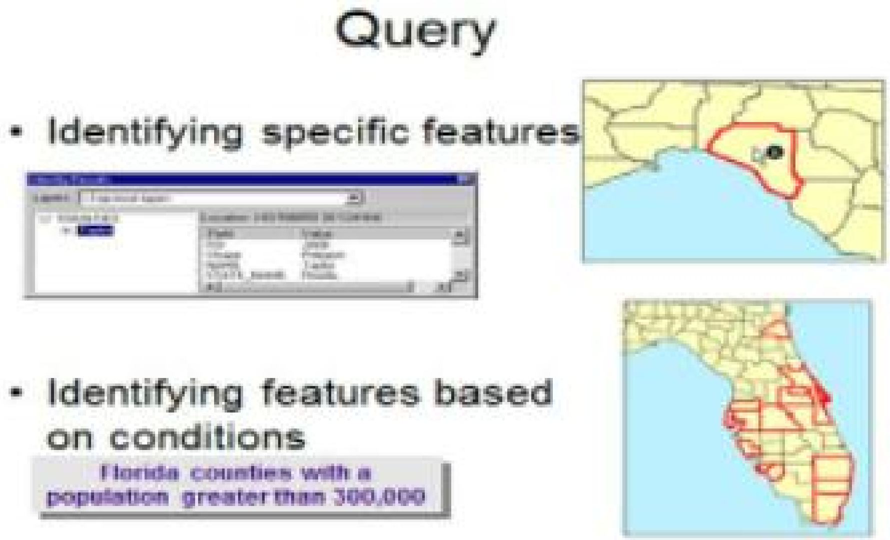
9/1 Types of GIS Query :

9/1/1 Identifying specific features

One common type of GIS query is to determine what exists at a particular location. In this type of query, the user understands where the features of interest are, but wants to know what characteristics are associated with them. This can be accomplished with GIS because the spatial features are linked to the descriptive characteristics.

9/1/2 Identifying features based on conditions

Another type of GIS query is to determine the locations that satisfy certain conditions. In this case the user knows what characteristics are important and wants to find out where the features are that have those characteristics.

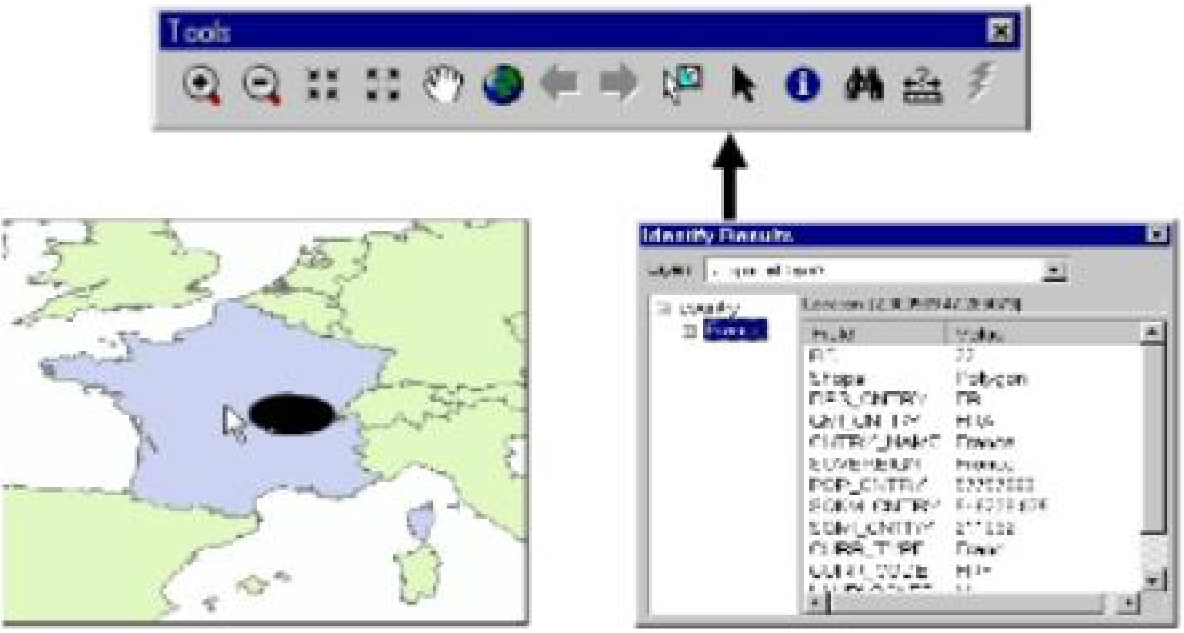


9/2 The Identify Features tool

This tool allows you to display the attributes for any feature you click on with your pointer.

9/2/1 Identifying Features

Popup attributes for a specific feature



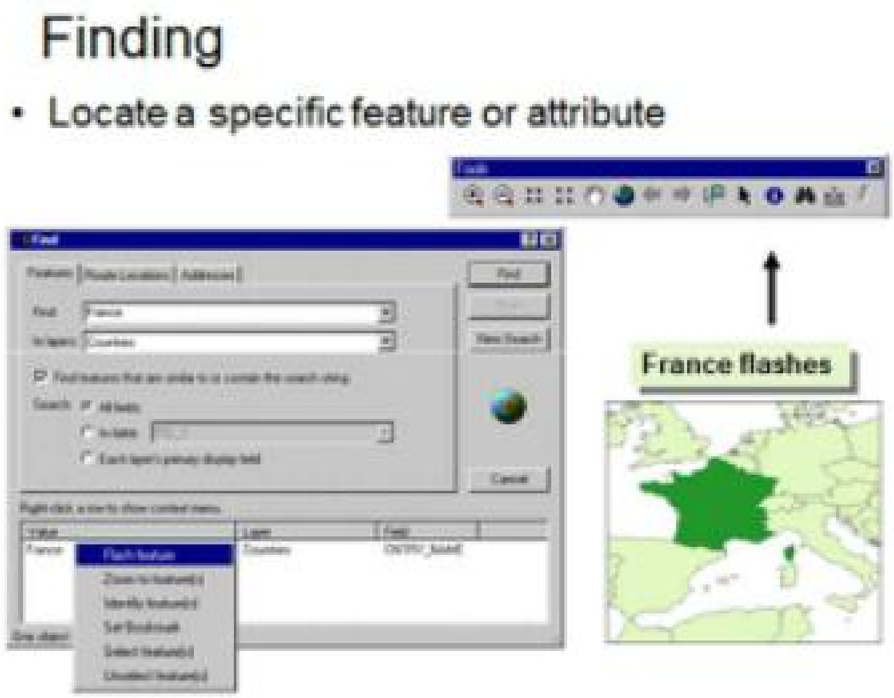
Identifying Features

9/2/2 The Find button

Clicking this button brings up the Find window on the screen. Type in the string that you want to search for. Notice that you can search in all the layers in your map, or just in a specific layer. You can also choose to search all the fields (attributes), specific fields, or each layer's primary field for the occurrence of the string you typed earlier.

The string can be typed in lowercase, uppercase, or a combination of both.

After the search is over, the window expands to reveal the findings. You will get the layer(s) and the field(s) in which the string was found in the feature attribute table(s). If you right-click the value, a context menu appears with options to flash the feature, zoom to it, identify it, set a bookmark, and select or unselect the feature.



9/2/3 Measuring :

Use the Measure tool to find distances on the map. The Measure tool allows you to draw a line on the map using your mouse. To start, click at a location, and as you move your mouse, a thick line will appear on the map. A single click will add a line segment, while a double click will end the line. The Status bar will report the length of each segment as it is created and the total length of the line when you are finished.

All measurements use a pure Cartesian coordinate system, so use such measurements with caution. You should expect some errors due to many factors (some personal and others induced by projections) when using this tool for small-scale renditions.

9/2/4 Map tips and hyperlinks

If you have map tips set for a layer, when you move the mouse pointer over a feature in the layer, a rectangular box containing textual information appears.

The map tip text comes from a field in the attribute table of that layer. You have to set which field you want attribute values to be reported from when using the map tips.

You can display Web pages accessed over the Internet and documents (such as a text file or image), or run a macro (script). You can dynamically create hyperlinks as you browse your map, or you can store hyperlinks with your data in an attribute field.

When you click on a feature, ArcMap determines which program is needed to display the hyperlink. If you specify a Web address, ArcMap launches your default Web browser and displays the page. If you specify a different type of document (e.g., a text document), ArcMap will display it using its native program (such as Notepad or another text editor).

If you are creating maps that people will access interactively, or if you want to explore your data before you do analysis, map tips and hyperlinks are useful ways to present more information about the map's features.

Map tips and hyperlinks

- Display property of a layer
- Map tips
  - Pointer location displays specific attribute

- Hyperlinks
  - Document
  - URL
  - Macro

9/3 Working with the selection tools :

9/3/1 Why do you need a selection?

You may have several reasons why you would make a selection, including:

- Using the selected set for further analysis
- Using the selected set to select other features
- Editing the selected set
- Creating a new layer from the selected set (Working with a subset is faster than including the entire set.)
- Calculating statistics for the selected set
- Creating a report

- Exporting the selected set to a separate file
- Converting features to a graphic format

9/3/2 Available selection tools

There are four ways to select graphics in ArcMap: interactively, by attribute, by location, and by graphics.

9/3/2/1 Interactive selection method

This option offers different selection methods, including "create a new selection," "add to current selection," "remove from current selection," and "select from current selection." Options further refine the interactive selection methods, which include selecting features that are spatially or completely within the box or graphic, selecting features that are completely within the box or graphic, and selecting features that the box or graphic are completely within.

9/3/2/2 Select by attributes

With this option you can write a selection statement. The search is done on the records in the feature attribute table based on the selection criteria typed in the selection statement. Because all features are linked to their respective records in the feature attribute table, ArcMap will be able to select the features based on their attributes.

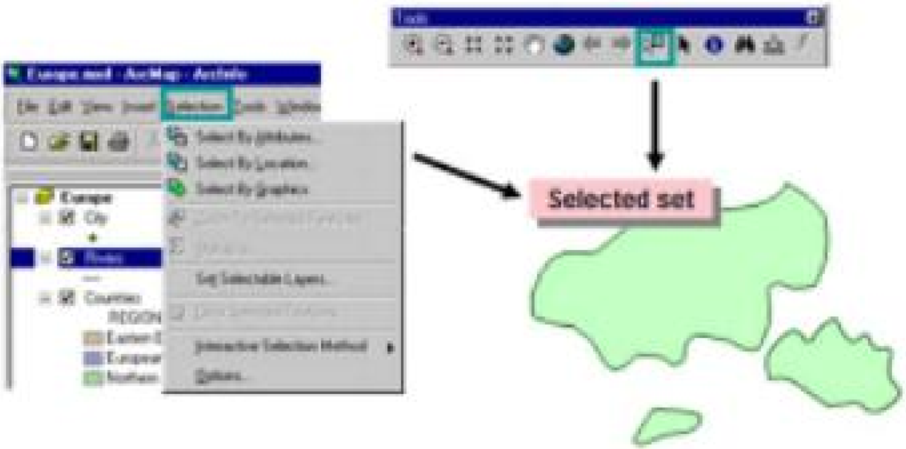
9/3/2/3 Select by location

Here, features from a certain layer can be selected by features from another. This is considered a spatial query tool.

9/3/2/4 Select by graphics

Features are selected based on their relationship to a graphic or graphics you create through the Draw toolbar (e.g., a polygon or line). The same selection options apply when you select by graphics as when you use the interactive selection method.

Other features in this selection menu include controlling which layers will be considered in the selection, zooming to the selected features, and clearing the selection.



9/3/3 Set Selectable Layers :

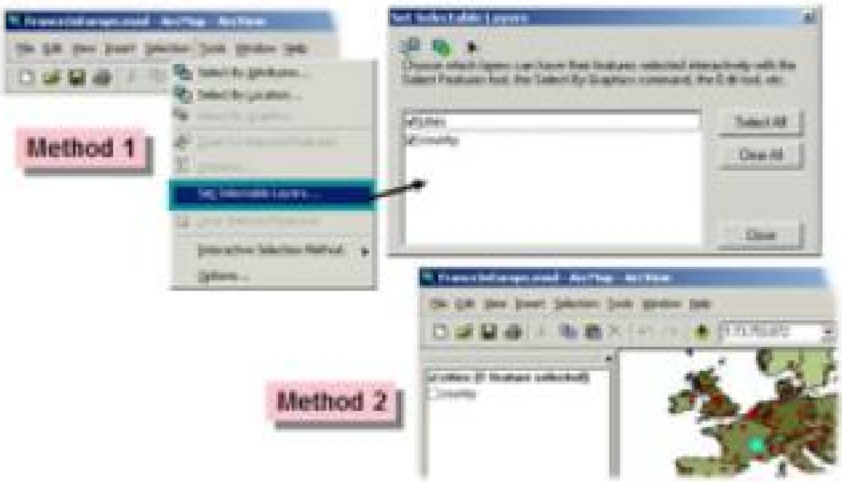
While making spatial selections, you have the option of turning on or off the layers that are going to be involved in the selection operation. This can be done in one of two ways.

Method 1

From the Selection menu, click Set Selectable Layers. Check the check boxes for any layers you want to include in your next selection.

Method 2

From the Tools menu, click Options and click on the TOC tab. Check the Selection check box in the TOC tab options panel, then click OK. Now you have a permanent Selection tab at the bottom of ArcMap's Table of Contents. When you click this tab, a list of layers with check boxes to their left appears. Now you can check which layers should or should not be included in your next selection.



9/3/4 Interactive selection method

Four methods are available:

9/3/4/1 Create New Selection

All features are available at the onset, and you select four countries: France, Poland, Bulgaria, and Greece.

9/3/4/2 Add to Current Selection

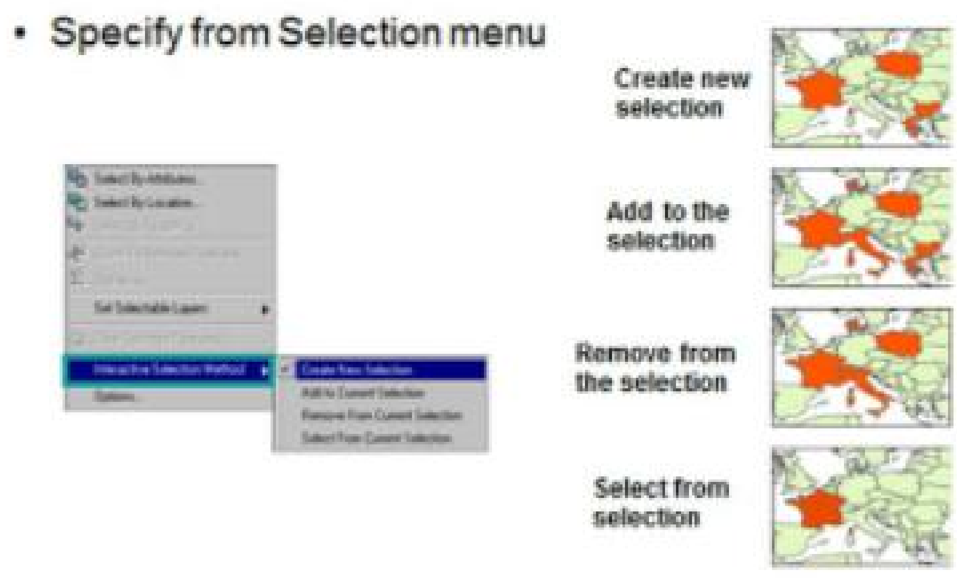
Now you want to add more countries to the first four selected. You select Denmark and Italy.

9/3/4/3 Remove From Current Selection

Here, you decide to remove Bulgaria and Greece from the current selection.

9/3/4/4 Select from Current Selection

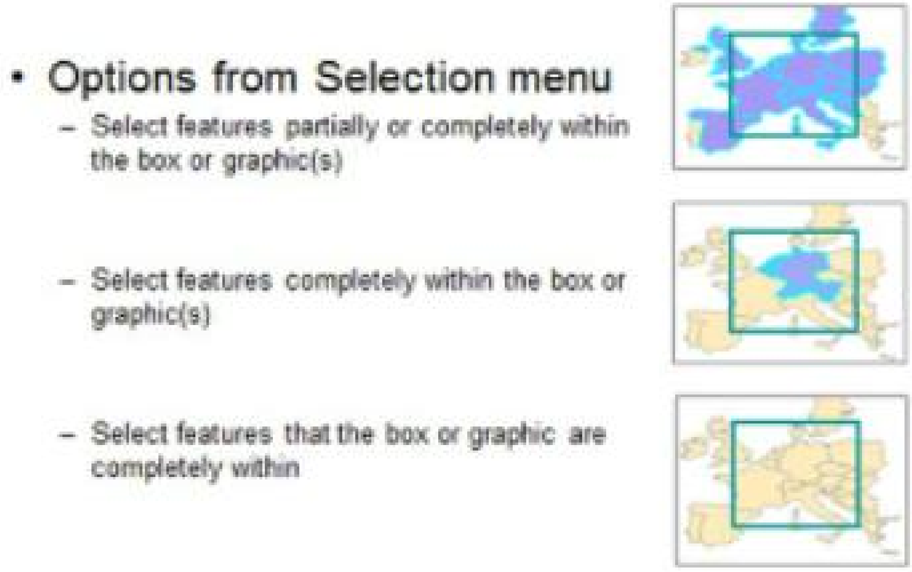
This time you decide to select only France from the current selection.



9/3/5 Interactive selection options

The Selection pulldown menu offers three additional selection options depending on how you want the features to be selected when you create a selection box:

1. Select features partially or completely within the box or graphic: Complete features are selected, whether they fall completely within the selection box or fall partially within the selection box.
2. Select features completely within the box or graphic: Complete features are selected only if they fall completely within the selection box.
3. Select features that the box or graphic are completely within: The selection box must fall completely within the feature to get selected. In the third example above, no feature was selected, because the box was simply too big to fit inside any country.



9/3/6 Attribute selection

ArcMap lets you select features using a "where clause" from Structured Query Language (SQL) in the Selection menu's Select By Attributes dialog. SQL is a powerful language you use to define one or more criteria by which you want to select features or rows. You define the criteria by creating expressions consisting of attributes, operators, and values.

For example, imagine you have a customer database and you want to find those customers who spent more than $50,000 last year and whose business type is Restaurant. You could select the customers with this expression: Sales > 50000 AND Business_type = 'Restaurant'.

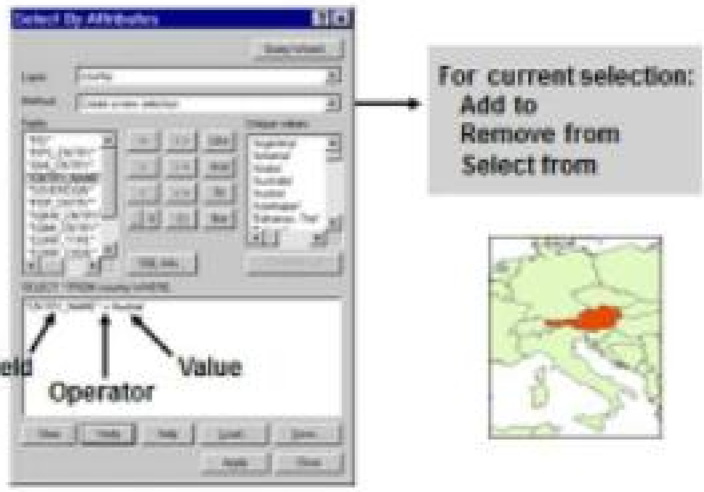
To create an attribute selection, you need to do the following:

1. From the Selection pulldown menu, click Select By Attributes.
2. On the Select By Attribute window, choose the layer from which you want to select features.
3. Write a selection statement (otherwise known as an SQL statement). You can type this yourself, or you can create it by clicking a field and an operator button, followed by a value. You can also write more complicated expressions using connectors, such as "And", "Or", and so on. Clicking the Verify button helps you ensure that the syntax of the expression is correct.
4. When you're convinced the syntax is correct, click OK.

   You can build expressions to select features directly from your map or to select records from atable. Selecting records in an attribute table also highlights features in the map, so you can see where the associated features are.

You can save selection expressions and reload them with the Save and Load buttons at the bottom of the Select By Attributes dialog. This saves time when you're working with complex query expressions.

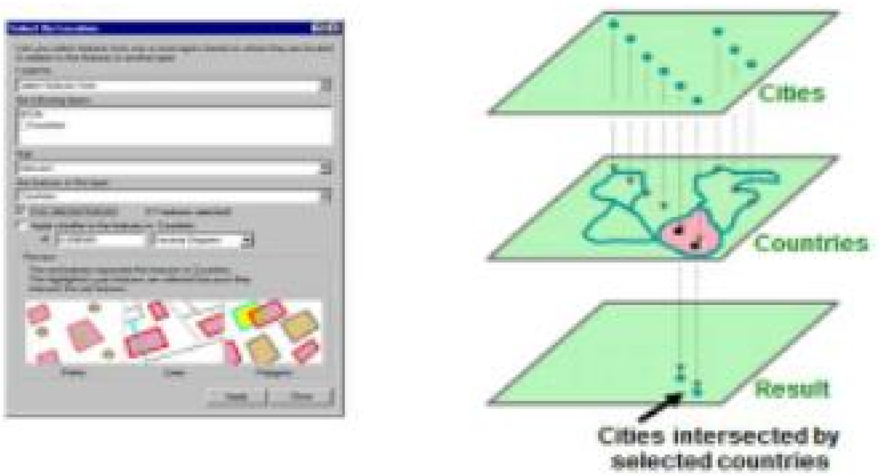
Simply load the expression back into the Select By Attributes dialog to easily regenerate a set of selected records.



Attribute Selection

9/3/7 Select by location (spatial query)

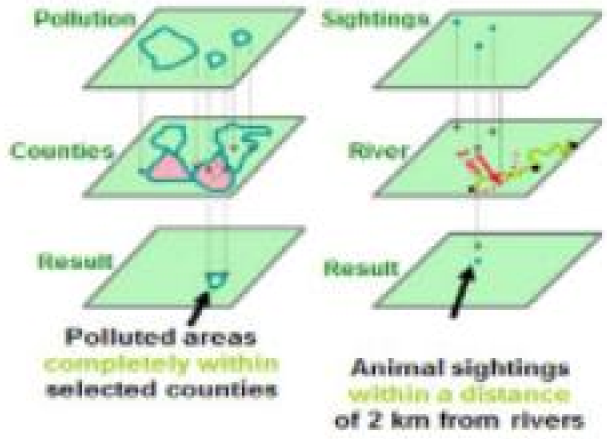
- Use features in one layer to select features in another.



Location selection methods:

- Select by location offers many selection methods
  - Intersects
  - Contains
  - Are contained by
  - Shares a line segment
  - Touch boundary
  - Within a distance
  - Are identical
  - Others

Location selection methods



With the Select By Location dialog box, you can select features based on their location relative to other features. Suppose you want to know how many homes were affected by a recent flood. Answering this question—and others like it—involves forming a spatial query. You want to find features based on where they are in relation to other features. For instance, if you mapped the flood boundary, you could then select all the homes that are within this area.

By combining queries, you can perform more complex searches. For example, suppose you want to find all the customers who live within a 20-mile radius of your store and who made a recent purchase so you can send them a promotional mailing. You would first select the customers within this radius (select by location) and then refine the selection by finding those customers who have made a purchase within the last six months according to a date-of-last-purchase attribute. You can use a variety of selection methods to select the point, line, or polygon features in one layer that are near or overlap the features in the same or another layer.

9/3/8 Select by graphics

You may use the tools on the Drawing toolbar to add graphics to the ArcMap display. For example, you may want to digitize a polygon around a group of islands that you wish to select. Once the graphic has been added to the display, the Select by Graphics option will be activated on the Selection menu.