AALIM MUHAMMED SALEGH COLLEGE OF ENGINEERING

DEPARTMENT OF COMUTER SCIENCE AND ENGINEERING

R2017 - SEMESTER V

OPEN ELECTIVE I:
OCE552 - GEOGRAPHICAL INFORMATION SYSTEM

UNIT III - DATA INPUT AND TOPOLOGY

SYLLABUS

UNIT III DATA INPUT AND TOPOLOGY 9

Scanner - Raster Data Input – Raster Data File Formats – Vector Data Input – Digitiser – Topology - Adjacency, connectivity and containment – Topological Consistency rules – Attribute Data linking – ODBC – GPS - Concept GPS based mapping.

Table of Contents

1. Introduction

1.1 Sources of data for GIS

- Problems can arise when some of the data is drawn from large scale mapping and other data is drawn from much smaller scale mapping.
- In this case great care has to be taken that conclusions are not drawn on the basis of the less reliable data.

There are several methods used for entering spatial data into a GIS, including:

1. manual digitising and scanning of analogue maps
2. image data input and conversion to a GIS
3. direct data entry including global positioning systems (GPS)
4. transfer of data from existing digital sources

- At each stage of data input there should be data verification should occur to ensure that the resulting database is as error free as possible.
- Data encoding is the process of getting data into the computer. It is a process that isfundamental to almost every GIS project.

For example:

- An archaeologist may encode aerial photographs of ancient remains to integrate with newly collected field data.
- A planner may digitize outlines of new buildings and plot these on existing topographical data.
- An ecologist may add new remotely sensed data to a GIS to examine changes in habitats.
- A historian may scan historical maps to create a virtual city from the past.
- A utility company may encode changes in pipeline data to record changes and upgrades to their pipe network.

Once in a GIS, data almost always need to be corrected and manipulated to ensure that they can be structured according to the required data model.

Problems that may have to be addressed at this stage of a GIS project include:

- the re-projection of data from different map sources to a common projection;
- the generalization of complex data to provide a simpler data set; or
- the matching and joining of adjacent map sheets once the data are in digital form.

- This unit looks in detail at the range of methods available to get data into a GIS.

- These include keyboard entry, digitizing, scanning and electronic data transfer.
- Then, methods of data editing and manipulation are reviewed, including re-projection, transformation and edge matching.

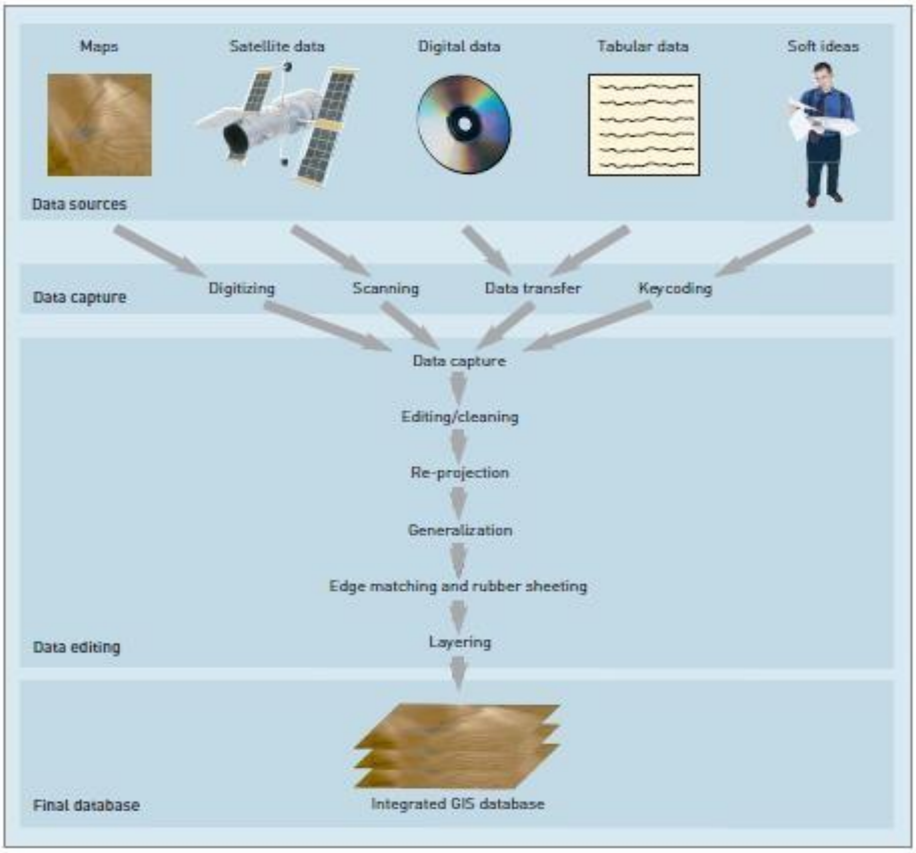


Figure: The Data Stream

- The whole process of data encoding and editing is often called the 'data stream'.
- Analogue data are normally in paper form, and include paper maps, tables of statistics and hard-copy (printed) aerial photographs.
- These data all need to be converted to digital form before use in a GIS, thus the data encoding and correction procedures are longer than those for digital data.
- Digital data are already in computer-readable formats and are supplied on CD-ROM or across a computer network.
- Map data, aerial photographs, satellite imagery, data from databases and automatic data collection devices (such as data loggers and GPS) areall available in digital form.

****

2. SCANNER

2.1 Introduction to Scanner

- Scanning coverts paper maps into digital format by capturing features as individual cells, or pixels, producing an automated image.
- Maps are generally considered the backbone of any GIS activity.
- But many a time paper maps are not easily available in a form that can be readily used by the computers.
- Most of the paper maps had been prepared on the basis of old conventional surveys. New maps can be produced using improved technologies but this requires time as it increases the volume of work.
- Thus, we have to resort to the available maps. These paper maps have to be first converted into a digital format usable by the computer.
- This is a critical step as the quality of the analog document must be preserved in the transition to the computer domain.

- The technology used for this kind of conversions is known as scanning and the instrument used for this kind of operation is known as a scanner.
- A scanner can be thought of as an electronic input device that converts analog information of a document like a map, photograph or an overlay into a digital format that can be used by the computer.
- Scanning automatically captures map features, text, and symbols as individual cells, or pixels, and produces an automated image.

2.2 Working of a Scanner

- The most important component inside a scanner is the scanner head which can move along the length of the scanner.
- The scanner head contains either a charged-couple device (CCD) sensor or a contact image (CIS) sensor.
- A CCD consists of a number of photosensitive cells or pixels packed together on a chip. The most advanced large format scanners use CCD's with 8000 pixels per chip for providing a very good image quality.
- While scanning a bright white light from the scanner strikes the image to be scanned and is reflected onto the photosensitive surface of the sensor placed on the scanner head.
- Each pixel transfers a gray tone value (values given to the different shades of black in the image ranging from 0 (black) – 255 (white) i.e. 256 values to the scan board (software).
- The software interprets the value in terms of 0 (Black) or 1 (white), thereby, forming a monochrome image of the scanned portion.
- As the head moves ahead, it scans the image in tiny strips and the sensor continues to store the information in a sequential fashion. The software running the scanner pierces

together the information from the sensor into a digital form of the image. This type of scanning is known as one pass scanning.

- Scanning a colour image is slightly different in which the scanner head has to scan the same image for three different colours i.e. red, green, blue. In older colour scanners, this was accomplished by scanning the same area three times over for the three different colours.
- This type of scanner is known as three-pass scanner. However, most of the colour scanners now scan in one pass scanning all the three colours in one go by using colour filters.
- In principle, a colour CCD works in the same way as a monochrome CCD. But in this each colour is constructed by mixing red, green and blue.
- Thus, a 24-bit RGB CCD presents each pixel by 24 bits of information. Usually, a scanner using these three colours (in full 24 RGB mode) can create up to 16.8 million colours.

2.3 Types of Scanners

- *Hand-held scanners* although portable, can only scan images up to about four inches wide.
- They require a very steady hand for moving the scan head over the document. They are useful for scanning small logos or signatures and are virtually of no use for scanning maps and photographs.
- The most commonly used scanner is a *flatbed scanner* also known as desktop scanner.It has a glass plate on which the picture or the document is placed.
- The scanner head placed beneath the glass plate moves across the picture and the result is a good quality scanned image.
- For scanning large maps or top sheets wide format flatbed scanners can be used.

Figure: Types of Scanners

- Then there are the *drum scanners* which are mostly used by the printing professionals. In this type of scanner, the image or the document is placed on a glass cylinder that rotates at very high speeds around a centrally located sensor containing photo-multiplier tube instead of a CCD to scan.
- Prior to the advances in the field of sheet fed scanners, the drum scanners were extensively used for scanning maps and other documents.

****

3. RASTER DATA INPUT

3.1 An Overview of GIS Data Input

Representing Data with Raster:

- Area is covered by grid with (usually) equal –sized, square cells.
- Attributes are recorded by assigning each cell a single value based on the majority feature (attribute) in the cell, such as land use type.
- Image data is a special case of raster data in which the "attribute" is a reflectance value from the geomagnetic spectrum cells in image data often called pixels (picture elements).
- Raster data divides the entire study area into a regular grid of cells.
- Each cell contains a single value.
- Raster data can be imagined as collection of cells organized like a matrix.

What is raster data input?

- Raster data (also known as grid data) represents the fourth type of feature: surfaces.
- Raster data is cell-based and this data category also includes aerial and satellite imagery.
- There are two types of raster data: continuous and discrete.
- An example of discrete raster data is population density

What are the types of raster data?

There are three types of raster data that can be stored in a geodatabase:

a) raster datasets,
b) raster catalogs, and
c) raster as attributes.

Raster datasets are single images that are stored in the database.

3.2 Raster Data Input Collection

Raster Data Input Collection can be most expensive GIS activity
Many diverse sources
Two broad types of collection

a) Data capture (direct collection)
b) Data transfer

Two broad capture methods

a) Primary (direct measurement)
b) Secondary (indirect derivation)

Input Collection options

a) Keyboard entry
b) Manual Digitizing (example: tablet, on screen)
c) Scanning

Raster Data Collection Techniques

Primary Raster Data Capture

- Capture specifically for GIS use

Raster – remote sensing

- Example SPOT and IKONOS satellites and aerial photography, echo sounding at sea
- Passive and active sensors

Resolution is key consideration

a) Spatial
b) Spectral, Acoustic
c) Temporal

	Field/Raster
Primary	Digital remote sensing images
	Digital aerial photographs
Secondary	Scanned maps
	DEMs from maps

- Data collected for other purposes, then converted for use in GIS

<u>Raster conversion</u>

- Scanning of maps, aerial photographs, documents, etc.
- Important scanning parameters are spatial and spectral (bit depth) resolution

****

4. RASTER GIS FILE FORMATS

4.1 Introduction to GIS FILE

- Raster data is made up of pixels (also referred to as grid cells). They are usually regularly-spaced and square but they don't have to be.
- Rasters have pixel that are associated with a value (continuous) or class (discrete).

4.2 List of GIS File Formats

The lists of GIS File Formats are described in the following Table:

Extension	File Type	Description

ERDAS Imagine (IMG)	*.IMG*	ERDAS Imagine IMG files is a proprietary file format developed by Hexagon Geospatial. IMG files are commonly used for raster data to store single and multiple bands of satellite data. IMG files use a hierarchical format (HFA) that are optional to store basic information about the file. For example, this can include file information, ground control points and sensor type. Each raster layer as part of an IMG file contains information about its data values. For example, this includes projection, statistics, attributes, pyramids and whether or not it's a continuous or discrete type of raster.
American Standard Code forInformation Interchange ASCII Grid	*.ASC*	ASCII uses a set of numbers (including floats) between 0 and 255 for information storage and processing. They also contain header information with a set of keywords. In their native form, ASCII text files store GIS data in a delimited format. This could be comma, space or tab-delimited format. Going from non-spatial to spatial data, you can run a conversion process tool like ASCII to raster.
GeoTIFF	*.TIF* *.TIFF* *.OVR*	The GeoTIFF has become an industry image standard file for GIS and satellite remote sensing applications. GeoTIFFs may be accompanied by other files: ▪ TFW is the world file that is required to give your raster geolocation. ▪ XML optionally accompany GeoTIFFs and are your metadata. ▪ AUX auxiliary files store projections and other information. ▪ OVR pyramid files improves performance for raster display.

IDRISI Raster	*.RST* *.RDC*	IDRISI assigns RST extensions to all raster layers. They consist of numeric grid cell values as integers, real numbers, bytes and RGB24. The raster documentation file (RDC) is a companion text file for RST files. They assign the number of columns and rows to RST files. Further to this, they record the file type, coordinate system, reference units and positional error.
Envi RAW Raster	*.BIL* *.BIP* *.BSQ*	Band Interleaved files are a raster storage extension for single/multi-band aerial and satellite imagery. ▪ Band Interleaved for Line (BIL) stores pixel information based on rows for all bands in an image. ▪ Whereas Band interleaved by pixel (BIP) assigns pixel values for each band by rows. ▪ Finally, Band sequential format (BSQ) stores separate bands by rows. BIL files consist of a header file (HDR) that describes the number of columns, rows, bands, bit depth and layout in an image.
Esri Grid		Grid files are a proprietary format developed by Esri. Grids have no extension and are unique because they can hold attribute data in a raster file. But the catch is that you can only add attributes to integer grids. Attributes are stored in a value attribute tables (VAT) – one record for each unique value in the grid, and the count representing the number of cells. The two types of Esri Grid files are integer and floating point grids. Land cover would be an example of a discrete grid. Each class has a unique integer cell value. Elevation data is an example of a floating point grid. Each cell represents an elevation floating value.

****

5. VECTOR GIS FILE FORMATS

5.1 Introduction to GIS File Formats

- Vector data is not made up of grids of pixels.
- Instead, vector graphics are comprised of vertices and paths.
- The three basic symbol types for vector data are points, lines and polygons (areas).

5.2 Vector File Formats

The lists of Vector File Formats are described in the following Table:

Extension	File Type	Description
Esri Shapefile	*.SHP, .DBF, .SHX*	The shapefile is BY FAR the most common geospatial file type you'll encounter. All commercial and open source accept shapefile as a GIS format. It's so ubiquitous that it's become the industry standard. But you'll need a complete set of three files that are mandatory to make up a shapefile. The three required files are: • SHP is the feature geometry. • SHX is the shape index position. • DBF is the attribute data.
Geographic JavaScript ObjectNotation (GeoJSON)	*.GEOJSON .JSON*	The GeoJSON format is mostly for web-based mapping. GeoJSON stores coordinates as text in JavaScript Object Notation (JSON) form. This includes vector points, lines and polygons as well as tabular information. GeoJSON store objects within curly braces {} and in general have less markup overhead (compared to GML). GeoJSON has straightforward syntax that you can modify in any text editor. Webmaps browsers understand JavaScript so by default GeoJSON is a common web format.

		But JavaScript only understands binary objects. Fortunately, JavaScript can convert JSON to binary.
Geography Markup Language(GML)	*.GML*	GML allows for the use of geographic coordinates extension of XML. And eXtensible Markup Language (XML) is both human-readable and machine-readable. GML stores geographic entities (features) in the form of text. Similar to GeoJSON, GML can be updated in any text editor. Each feature has a list of properties, geometry (points, lines, curves, surfaces and polygons) and spatial reference system. There is generally more overhead when compare GML with GeoJSON. This is because GML results in more data for the same amount of information.

Google Keyhole MarkupLanguage (KML/KMZ)	*.KML* *.KMZ*	KML stands for Keyhole Markup Language. This GIS format is XML-based and is primarily used for Google Earth. KML was developed by Keyhole Inc which was later acquired by Google. KMZ (KML-Zipped) replaced KML as being the default Google Earth geospatial format because it is a compressed version of the file. KML/KMZ became an international standardof the Open Geospatial Consortium in 2008. The longitude, latitude components (decimal degrees) are as defined by the World Geodetic System of 1984 (WGS84). The vertical component (altitude) is measured in meters from the WGS84 EGM96 Geoid vertical datum.
GPS eXchange Format (GPX)	*.GPX*	GPS Exchange format is an XML schema that describes waypoints, tracks and routes captured from a GPS receiver. Because GPX isan exchange format, you can openly transfer GPS data from one program to another based on its description properties. The minimum requirement for GPX are latitude and longitude coordinates. In addition,GPX files optionally stores location propertiesincluding time, elevation and geoid height as tags.
IDRISI Vector	.VCT .VDC	IDRISI vector data files have a VCT extension along with an associated vector documentation file with a VDC extension. VCT format are limited to points, lines, polygons, text and photos. Upon the creationof an IDRISI vector file, it automatically creates a documentation file for building metadata. Attributes are stored directly in the vector files. But you can optionally use independent data tables and value files.

****

5. DIGITIZING

5.1 Introduction to Digitizing

- Digitizing in GIS is the process of converting geographic data either from a hardcopy or a scanned image into vector data by tracing the features.
- During the digitzing process, features from the traced map or image are captured as coordinates in either point, line, or polygon format.

5.2 Types of Digitizing in GIS:

1) **Manual Digitizing.** Manual Digitizing is done by digitizing tablet**.**
2) **Heads-up Digitizing.** Heads-up Digitizing is similar to manual digitizing**.**
3) **Interactive Tracing Method.** The interactive tracing method is an advanced technique that has evolved from Heads-up digitizing..
4) **Automatic Digitizing.**

1) Manual Digitizing

- Manual Digitizing is done by digitizing tablet.
- The digitizer manually traces all the lines from the hardcopy map (eg.Toposheet), and parallelly.
- The digital maps are created on the computer.
- It is only less time consuming but also has high accuracy when comparing with other digitizing methods.
- The most common method of encoding spatial features from paper maps is **manual digitizing.**

- It is an appropriate technique when selected features are required from a paper map.
- Manual digitizing requires a digitizing table that is linked to a computer workstation.
- The digitizing table is essentially a large flat tablet, the surface of which is underlain by a very fine mesh of wires.
- Attached to the digitizer via a cable is a cursor (puck) that can be moved freely over the surface of the table. Buttons on the cursor allow the user to send instructions to the computer.
- The position of the cursor on the table is registered by reference to its position above the wire mesh.

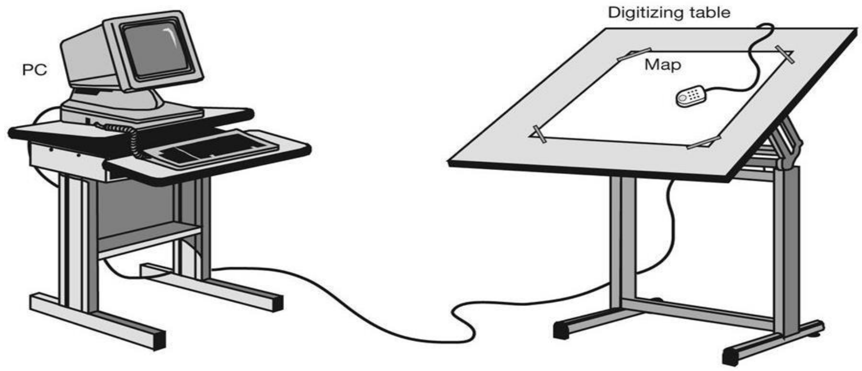


Figure: Manual Digitizer

- Heads up digitizing (also referred to as on-screen digitizing) is the method of tracing geographic features from another dataset (usually an aerial, satellite image, or scanned image of a map) directly on the computer screen.
- Automated digitizing involves using image processing software that contains pattern recognition technology to generated vectors.

The procedure followed when digitizing a paper map using a manual digitizer has the following five stages:

a) ***Registration:*** The map to be digitized is fixed firmly to the table top with sticky tape. Five or more control points are identified (usually the four corners of the map sheet and one or more grid intersections in the middle). The geographic co-ordinates of the control points are noted and their locations digitized by positioning the cross-hairs on the cursor exactly over them and pressing the 'digitize' button on the cursor. This sends the co-ordinates of a point on the table to the computer and stores them in a file as 'digitizer co-ordinates'.

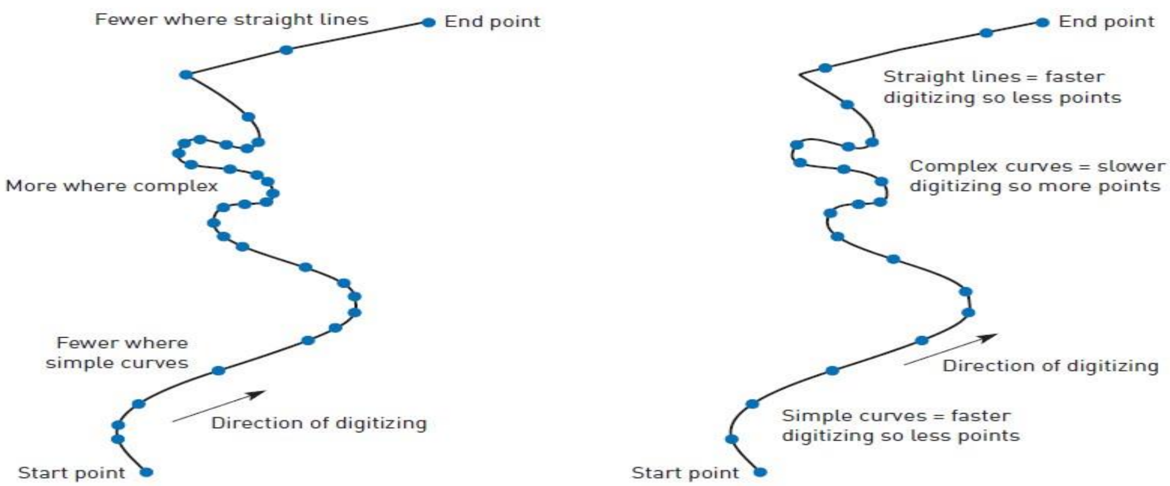
b) ***Digitizing point features:*** Point features, for example spot heights, hotel locations or meteorological stations, are recorded as a single digitized point. A unique code number or identifier is added so that attribute information may be attached later. For instance, the hotel with ID number '1' would later be identified as 'Mountain View'.

c) ***Digitizing line features****:* Line features (such as roads or rivers) are digitized as a seriesof points that the software will join with straight line segments. In some GIS packageslines are referred to as arcs, and their start and end points as nodes. This gives rise to the term arc–node topology, used to describe a method of structuring line features.

d) ***Digitizing area (polygon) features****:* Area features or polygons, for example forested areas or administrative boundaries, are digitized as a series of points linked togetherby line segments in the same way as line features. Here it is important that the start and end points join to form a complete area. Polygons can be digitized as a series of individual lines, which are later joined to form areas. In this case it is important that each line segment is digitized only once.

e) ***Adding attribute information:*** Attribute data may be added to digitized polygon features by linking them to a centroid (or seed point) in each polygon. These are either digitized manually (after digitizing the polygon boundaries) or created automatically once the polygons have been encoded. Using a unique identifier or code number, attribute data can then be linked to the polygon centroids of appropriate polygons. In this way, the forest stand may have data relating to tree species, tree ages, tree numbers and timber volume attached to a point within the polygon.

- Manual digitizers may be used in one of two modes: point mode or stream mode.
- In point mode the user begins digitizing each line segment with a start node, records each change in direction of the line with a digitized point and finishes the segment with an end node.
- Thus, a straight line can be digitized with just two points, the start and end nodes.
- For more complex lines, a greater number of points are required between the start and end nodes. Smooth curves are problematic since they require an infinite number of points to record their true shape.

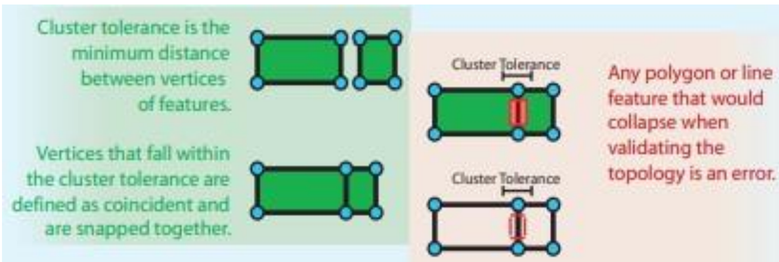


(a) *Point mode - person digitizing decides where to place each individual point such as to most accuratelyrepresent the line within the accepted tolerances of the digitizer.*

(b) *Stream mode – person digitizing decides on time or distance interval between the digitizing hardwareregistering each point as the person digitizing moves the cursor along the line.*

- In stream mode the digitizer is set up to record points according to a stated time interval or on a distance basis.

- Once the user has recorded the start of a line the digitizer might be set to record a point automatically every 0.5 seconds and the user must move the cursor along the line to record its shape.
- An end node is required to stop the digitizer recording further points.
- The speed at which the cursor is moved along the line determines the number of points recorded.
- Thus, where the line is more complex and the cursor needs to be moved more slowly and with more care, a greater number of points will be recorded.



- Conversely, where the line is straight, the cursor can be moved more quickly and fewer points are recorded.
- The choice between point mode and stream mode digitizing is largely a matter of personal preference.
- Stream mode digitizing requires more skill than point mode digitizing, and for an experienced user may be a faster method.
- Stream mode will usually generate more points, and hence larger files, than point mode.

<u>2) Heads-up Digitizing</u>

- Heads-up Digitizing is similar to manual digitizing.
- In the manual digitizing process, it digitizes in hardcopy, but in this method, it scans the map directly and displays it on the desktop screen.

<u>3) Interactive Tracing Method</u>

- The interactive tracing method is an advanced technique that has evolved from Heads-up digitizing.
- It is quite excellent in terms of accuracy and speed.

<u>4) Automatic Digitizing</u>

- Automatic Digitizing is the process of converting raster to vector in an automated method using pattern recognition and image processing techniques.
- In this technique, the computer traces all the features on the map; it gives high accuracy with low time consumption. It allows customization and improved quality of images. This process is also known as Vectorisation.

****

6. TOPOLOGY

6.1 Introduction to Topology

- In spatial technology, the word “topology” is all about the building blocks of geometry.
- Like other ‘ology’ words (e.g. geology and biology) it has its root in the study of something – and in this case we are talking about the study of geometrical properties and spatial relationships between constituent parts of a shape or a set of related shapes.
- Topology considers anything from the individual feature to the relationship between features. Individual features can be in the form of points, lines or polygons.
- Furthermore a polygon can be broken down into its constituent nodes/vertices, segments/edges and faces (see schematic below).
- Topology usually deals with the way in which features interact with adjacent or connecting features.
- Topology is the mathematical representation of the physical relationships that exists between the geographical elements.

What is Topology?

- Topology is a collection of rules that, coupled with a set of editing tools and techniques, enable the geodatabase to more accurately model geometric relationships.
- A topology is stored in a geodatabase as one or more relationships that define how the features in one or more feature classes share geometry.

6.2 Topology in GIS

- In geodatabases, a topology is a set of rules that defines how point, line, and polygon features share coincident geometry.
- Topology describes the means whereby lines, borders, and points meet up, intersect, and cross.
- This includes how street centerlines and census blocks share common geometry, and adjacent soil polygons share their common boundaries.
- Another example could be how two counties that have a common boundary between them will share an edge, creating a spatial relationship.
- The features participating in a topology are still simple feature classes—rather than modifying the definition of the feature class, a topology serves as a description of how the features can be spatially related.

6.3 Why topology?

- Topology has long been a key GIS requirement for data management and integrity.
- In general, a topological data model manages spatial relationships by representing spatial objects (point, line, and area features) as an underlying graph of topological primitives—nodes, faces, and edges.
- These primitives, together with their relationships to one another and to the features whose boundaries they represent, are defined by representing the feature geometries in a planar graph of topological elements.

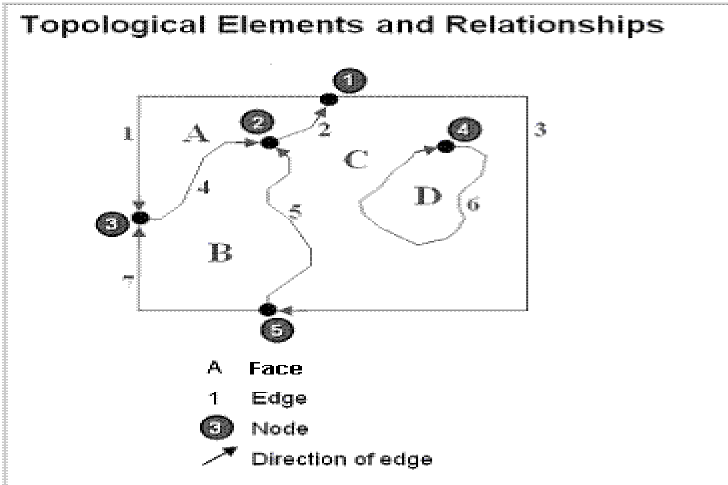
- Topology has long been a key GIS requirement for data management and integrity.
- In general, a topological data model manages spatial relationships by representing spatial objects (point, line, and area features) as an underlying graph of topological primitives—nodes, faces, and edges.
- These primitives, together with their relationships to one another and to the features whose boundaries they represent, are defined by representing the feature geometries in a planar graph of topological elements.

- Topology is useful in GIS because many spatial modeling operations don’t require coordinates, only topological information.
- For example, to find an optimal path between two points requires a list of the arcs that connect to each other and the cost to traverse each arc in each direction. Coordinates are only needed for drawing the path after it is calculated.

6.4 Elements of Topology

- Topology is fundamentally used to ensure data quality of the spatial relationships and to aid in data compilation.
- Topology is also used for analyzing spatial relationships in many situations, such as dissolving the boundaries between adjacent polygons with the same attribute values or traversing a network of the elements in a topology graph.

- Topology can also be used to model how the geometry from a number of feature classes can be integrated. Some refer to this as vertical integration of feature classes.



6.5 Features and topological elements

- A layer of polygons can be described and used in the following ways:
  1) As collections of geographic features (points, lines, and polygons).
  2) As a graph of topological elements (nodes, edges, faces, and their relationships).
  3) This means that there are two alternatives for working with features—one in which features are defined by their coordinates and another in which features are represented as an ordered graph of their topological elements.

In a geodatabase, the following properties are defined for each topology:

1) **The name of the topology to be created**.
- The cluster tolerance used in topological processing operations.
- The cluster tolerance is often a term used to refer to two tolerances: the x,y tolerance and the z-tolerance.
- The default value for the cluster tolerance is 10 times the coordinate resolution.

2) **List of feature classes.** First you need a list of the feature classes that will participate in a topology. All must be in the same coordinate system and organized into the same feature dataset.
3) **The relative accuracy rank of the coordinates in each feature class.**
If some feature classes are more accurate than others, you will want to assign a higher coordinate rank. This will be used in topological validation and integration. Coordinates of a lower accuracy will be moved to the locations of more accurate coordinates when they fall within the cluster tolerance of one another. Features with the highest accuracy should receive a value of 1, less accurate feature classes a value of 2, even less accurate feature classes a value of 3, and so on.

4) **<u>A list of topology rules for how features share geometry.</u>**

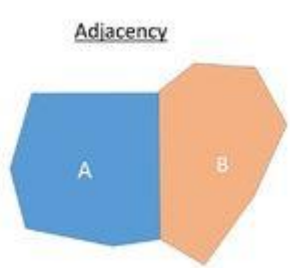
- Features can share geometry within a topology.
- Here are some examples among adjacent features:
  - Area features can share boundaries (polygon topology).
  - Line features can share endpoints (edge-node topology).
- In addition, shared geometry can be managed between feature classes using a geodatabase topology,
- For example:
  - Line features can share segments with other line features.
  - Area features can be coincident with other area features.
- For example, parcels can nest within blocks.
- Line features can share endpoint vertices with other point features (node topology).
- Point features can be coincident with line features (point events).

****

7. ADACENCY, CONNECTIVITY AND CONTAINMENT

1) ADJEACENCY

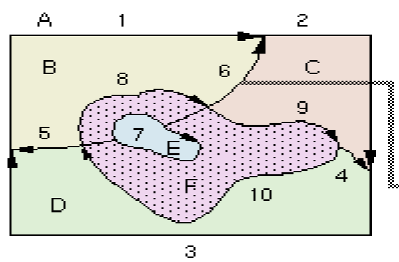
- Adjacent features can easily be found by intersecting target polygons with other polygons in the same map and identifying the points of intersection of polygons that touch boundaries or overlap.
- The geometric intersections of adjacent features are calculated on the fly by comparing the vertices of adjacent features rather than looking up adjacent features in a table.



Two adjacent polygons sharing a common border

Contiguity:

- Two geographic features that share a boundary are called adjacent. Contiguity is the topological concept that allows the vector data model to determine adjacency.
- Polygon topology defines contiguity. Polygons are contiguous to each other if they share a common arc. This is the basis for many neighbor and overlay operations.
- Recall that the from-node and to-node define an arc.
- This indicates an arc's direction so the polygons on its left and right sides can be determined. Left-right topology refers to the polygons on the left and right sides of an arc.
- In the below example, polygon B is on the left of arc 6, and polygon C is on the right. Thus we know that polygons B and C are adjacent.



Left-Right Topology

Arc	Left Polygon	Right Polygon
1	A	B
2	A	C
3	A	D
4	C	D
5	D	B
6	B	C
7	F	E
8	B	F
9	C	F
10	D	F

Figure: Left-Right topology example

- Notice that the label for polygon A is outside the boundary of the area.
- This polygonis called the external, or universe, polygon and represents the world outside the study area.
- The universe polygon ensures that each arc always has a left and right side defined.

Computing Adjacency Lists

Although analytical operations that require adjacency information can be performed in ArcView GIS through the interface, performance requirements many necessitate building a table to store adjacency information.

Topological data structures provide

- an automated way to handle digitizing and editing errors and artifacts;
- reduce data storage for polygons because boundaries between adjacent polygons are stored only once; and
- enable advanced spatial analyses such as adjacency, connectivity, and containment.

The topological structure supports three major topological concepts:

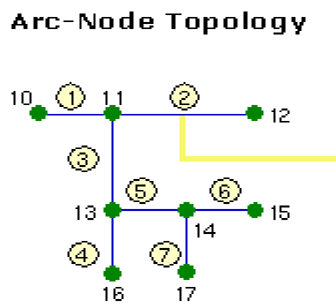
1) Connectivity: Arcs connect to each other at nodes.
2) Area definition: Arcs that connect to surround an area define a polygon.
3) Contiguity: Arcs have direction and left and right sides.

2) CONNECTIVITY

- Connectivity is defined through arc-node topology.
- This is the basis for many network tracing and path finding operations. Connectivity allows you to identify a route to the airport, connect streams to rivers, or follow a path from the water treatment plant to a house.

Example:

- In the arc-node data structure, an arc is defined by two endpoints:
  - the from-node indicating where the arc begins and a to-node indicating where it ends. This is called arc-nodetopology.



Arc-Node List

Arc	From Node	To Node
1	10	11
2	11	12
3	11	13
4	13	16
5	13	14
6	14	15
7	14	17

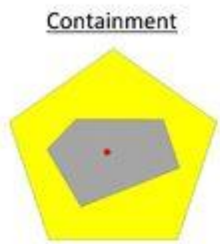
Figure: Arc-Node topology example

- Arc-node topology is supported through an arc-node list. The list identifies the from- and to-nodes for each arc. Connected arcs are determined by searching through the list for common node numbers.
- In the above example, it is possible to determine that arcs 1, 2, and 3 all intersect because they share node 11.
- The computer can determine that it is possible to travel along arc 1 and turn onto arc 3 because they share a common node (11), but it's not possible to turn directly from arc 1 onto arc 5 because they don't share a common node.

3) CONTAINMENT

- Many of the geographic features that may be represented cover a distinguishable area on the surface of the earth, such as lakes, parcels of land, and census tracts.
- An area is represented in the vector model by one or more boundaries defining a

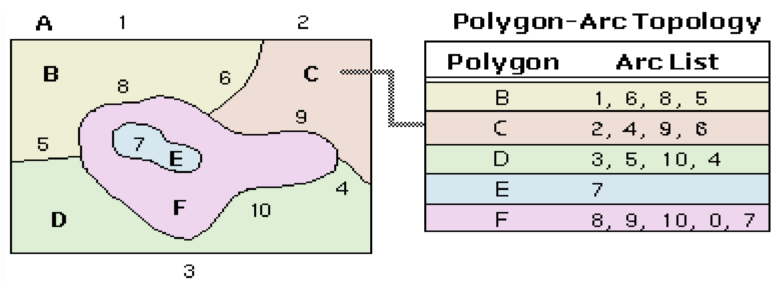
polygon. Although this sounds counterintuitive, consider a lake with an island in the middle.



A point-in-polygon and a polygon-in-polygon

Example:

- The lake actually has two boundaries: one that defines its outer edge and the island that defines its inner edge. In the terminology of the vector model, an island defines an inner boundary (or hole) of a polygon.
- The arc-node structure represents polygons as an ordered list of arcs rather than a closed loop of x,y coordinates.
- This is called polygon-arc topology. In the illustration below, polygon F is made up of arcs 8, 9, 10, and 7 (the 0 before the 7 indicates that this arc creates an island in the polygon).



Polygon	Arc List
B	1, 6, 8, 5
C	2, 4, 9, 6
D	3, 5, 10, 4
E	7
F	8, 9, 10, 0, 7

Figure: Polygon-Arc topology example

- Each arc appears in two polygons (in the above example, arc 6 appears in the list for polygons B and C).
- Since the polygon is simply the list of arcs defining its boundary, arc coordinates are stored only once, thereby reducing the amount of data and ensuring that the boundaries of adjacent polygons don't overlap.

****

8. TOPOLOGICAL CONSISTENCY RULES

8.1 Introduction Topological rules

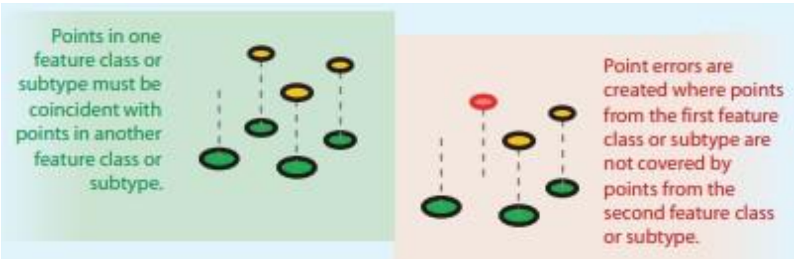
Topology Rules:

- There are many topology rules you can implement in your geodatabase, depending on the spatial relationships that are most important for your organization to maintain.
- You should carefully plan the spatial relationships you will enforce on your features. Some topology rules govern the relationships of features within a given feature class, while others govern the relationships between features in two different feature classes or subtypes.
- Topology rules can be defined between sub types of features in one or another feature class. This could be used, for example, to require street features to be connected to other street features at both ends, except in the case of streets belonging to the cul-de-sac or dead-end subtypes.
- Many topology rules can be imposed on features in a geodatabase.
- A well-designed geodatabase will have only those topology rules that define key spatial relationships needed by an organization. Most topology violations have fixes that you can use to correct errors.

8.2 Topology rules based on points

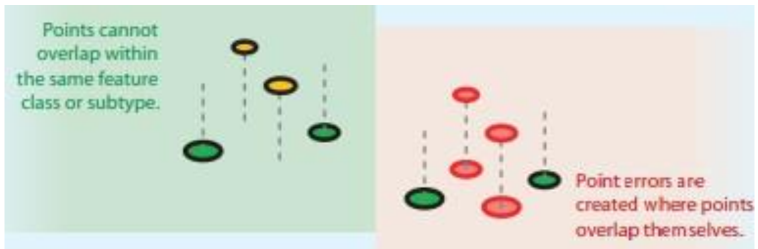
Must Coincide With:

- Requires that points in one feature class (or subtype) be coincident with points in another feature class (or subtype).
- This is useful for cases where points must be covered by other points, such as transformers must coincide with power poles in electric distribution networks and observation points must coincide with stations.



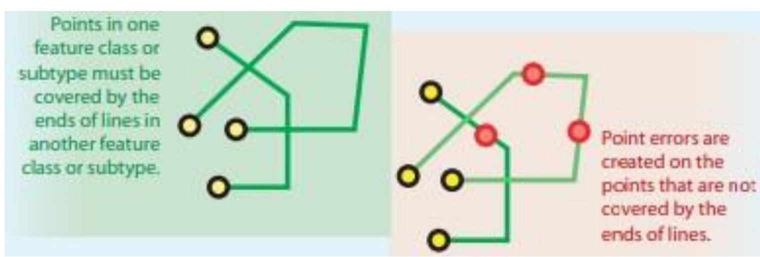
Must Be Disjoint:

- Requires that points be separated spatially from other points in the same feature class (or subtype). Any points that overlap are errors.
- This is useful for ensuring that points are not coincident or duplicated within the same feature class, such as in layers of cities, parcel lot ID points, wells, or streetlamp poles.



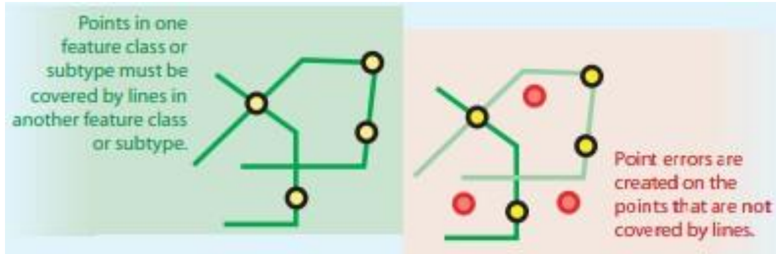
Must Be Covered By Endpoint of:

- Requires that points in one feature class must be covered by the endpoints of lines in another feature class.
- This rule is similar to the line rule Endpoint Must Be Covered By except that, in cases where the rule is violated, it is the point feature that is marked as an error rather than the line.
- Boundary corner markers might be constrained to be covered by the endpoints of boundary lines.



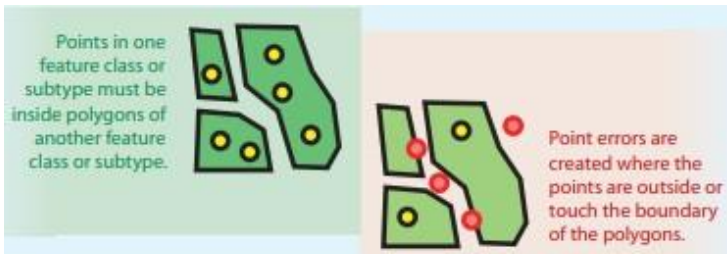
Point Must Be Covered By Line:

Requires that points in one feature class be covered by lines in another feature class. Itdoes not constrain the covering portion of the line to be an endpoint. This rule is useful for points that fall along a set of lines, such as highway signs along highways.



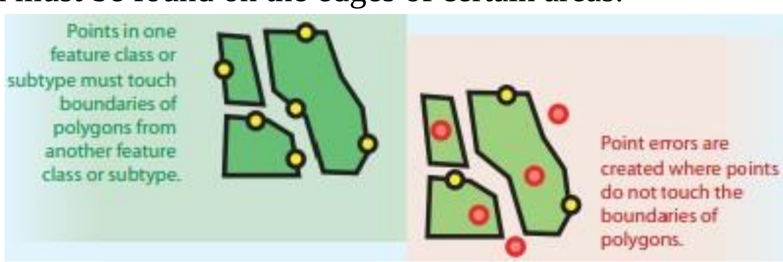
Must Be Properly Inside Polygons:

Requires that points fall within area features. This is useful when the point features are related to polygons, such as wells and well pads or address points and parcels.



Must Be Covered By Boundary of:

Requires that points fall on the boundaries of area features. This is useful when the point features help support the boundary system, such as boundary markers, which must be found on the edges of certain areas.

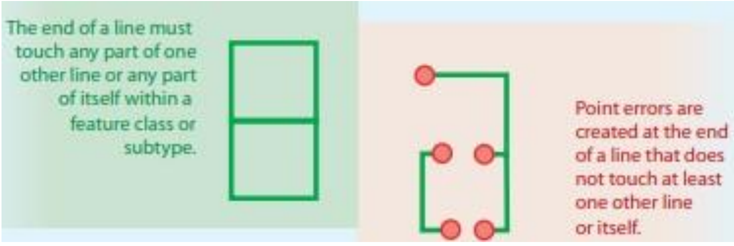


8.3 Topology rules based on Lines

Must Not Have Dangles:

- Requires that a line feature must touch lines from the same feature class (or subtype) at both endpoints.
- An endpoint that is not connected to another line is called a dangle.
- This rule is used when line features must form closed loops, such as when they are defining the boundaries of polygon features.

- It may also be used in cases where lines typically connect to other lines, as with streets. In this case, exceptions can be used where the rule is occasionally violated, as with cul-de-sac or dead-end street segments.

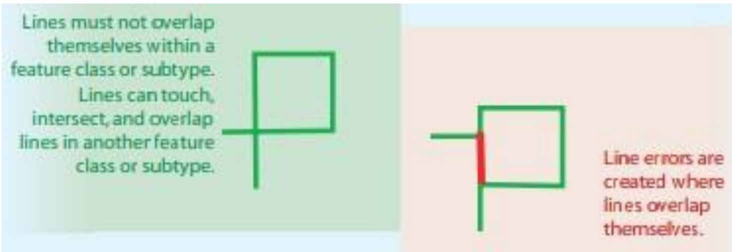


Must Not Overlap:

- Requires that lines not overlap with lines in the same feature class (or subtype).
- This rule is used where line segments should not be duplicated, for example, in a stream feature class. Lines can cross or intersect but cannot share segments.

Must Not Self-Overlap:

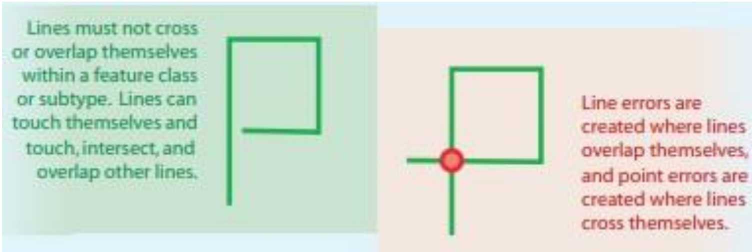
- Requires that line features not overlap themselves. They can cross or touch themselves but must not have coincident segments.
- This rule is useful for features, such as streets, where segments might touch in a loop but where the same street should not follow thesame course twice.



Must Not Self-Intersect:

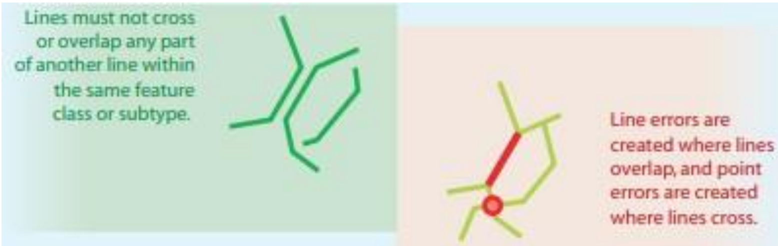
- Requires that line features not cross or overlap themselves.

- This rule is useful for lines, such as contour lines, that cannot cross themselves.



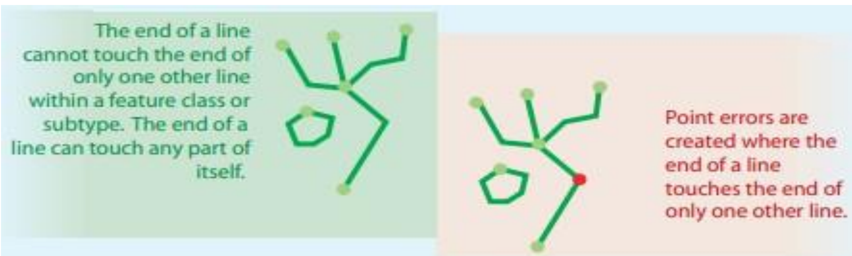
Must Not Intersect :

- Requires that a line in one feature class (or subtype) must only touch other lines of thesame feature class (or subtype) at endpoints.
- Any line segment in which features overlap or any intersection not at an endpoint is an error. This rule is useful where lines must only be connected at endpoints, such as in the case of plot lines, which must split (only connect to theendpoints of) back lot lines and cannot overlap each other.



Must Not Have Pseudo Nodes:

- Requires that a line connect to at least two other lines at each endpoint.
- Lines that connect to one other line (or to themselves) are said to have pseudo nodes.
- This rule is used where line features must form closed loops, such as when they define the boundaries of polygons or when line features logically must connect to two other line features at each end, as with segments in a stream network, with exceptions being marked for the originating ends of first-order streams.



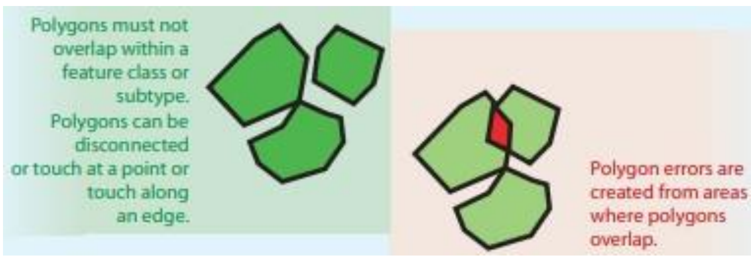
Must Be Larger Than Cluster Tolerance:

- Requires that a feature does not collapse during a validate process.
- This rule is mandatory for a topology and applies to all line and polygon feature classes.
- In instances where this rule is violated, the original geometry is left unchanged.

8.4 Topology rules based on Polygons

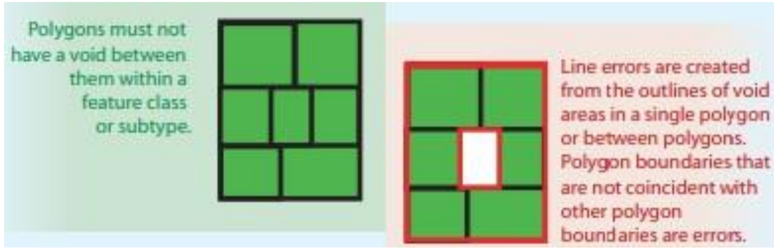
Must Not Overlap:

- Requires that the interior of polygons not overlap.
- The polygons can share edges or vertices. This rule is used when an area cannot belong to two or more polygons.
- It is useful for modeling administrative boundaries, such as ZIP Codes or voting districts, and mutually exclusive area classifications, such as land cover or landform type.



Must Not Have Gaps:

- This rule requires that there are no voids within a single polygon or between adjacent polygons.
- All polygons must form a continuous surface. An error will always exist on the perimeter of the surface.
- You can either ignore this error or mark it as an exception. Use this rule on data that must completely cover an area.
- For example, soil polygons cannot include gaps or form voids—they must cover an entire area.

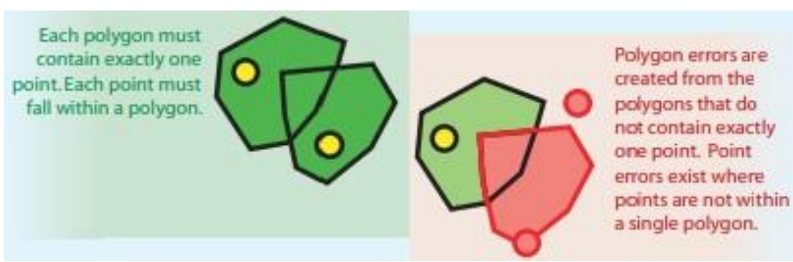


Contains Point:

Requires that a polygon in one feature class contain at least one point from another feature class. Points must be within the polygon, not on the boundary. This is useful when every polygon should have at least one associated point, such as when parcels must have an address point.

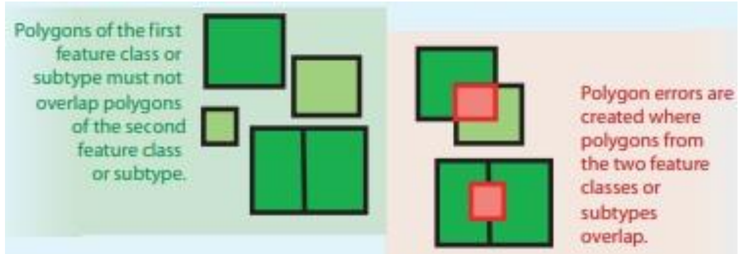
Contains One Point:

- Requires that each polygon contains one point feature and that each point feature falls within a single polygon.
- This is used when there must be a one-to-one correspondence between features of a polygon feature class and features of a point feature class, such as administrative boundaries and their capital cities.
- Each point must be properly inside exactly one polygon and each polygon must properly contain exactly one point.
- Points must be within the polygon, not on the boundary.



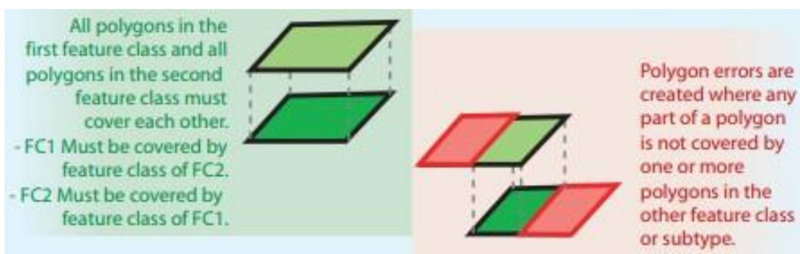
Must Not Overlap With:

- Requires that the interior of polygons in one feature class (or subtype) must not overlap with the interior of polygons in another feature class (or subtype).
- Polygons of the two feature classes can share edges or vertices or be completely disjointed.
- This rule is used when an area cannot belong to two separate feature classes.
- It is useful for combining two mutually exclusive systems of area classification, such as zoning and water body type, where areas defined within the zoning class cannot also be defined in the water body class and vice versa.



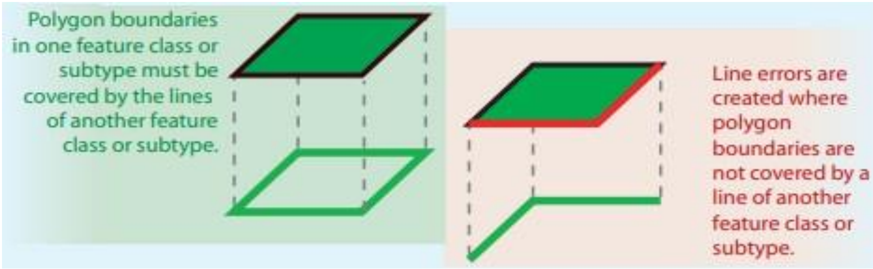
Must Cover Each Other:

- Requires that the polygons of one feature class (or subtype) must share all of theirarea with the polygons of another feature class (or subtype). Polygons may share edges or vertices.
- Any area defined in either feature class that is not shared with the other is an error.
- This rule is used when two systems of classification are used for the same geographic area, and any given point defined in one system must also be defined in the other.
- One such caseoccurs with nested hierarchical datasets, such as census blocks and block groups or small watersheds and large drainage basins.
- The rule can also be applied to non-hierarchically related polygon feature classes, such as soil type and slope class.



Area Boundary Must Be Covered By Boundary of:

- Requires that boundaries of polygon features in one feature class (or subtype) be covered by boundaries of polygon features in another feature class (or subtype).
- This is usefulwhen polygon features in one feature class, such as subdivisions, are composed of multiple polygons in another class, such as parcels, and the shared boundaries must be aligned.



****

9. ATTRIBUTE DATA LINKING

9.1 Introduction to Attribute Data

- There are two types of GIS data: spatial data (coordinate and projection information for spatial features) and attribute data. Attribute data is additional information appended in tabular format linked with spatial features.

9.2 Attribute Data Link

- The attribute data is linked with spatial data through unique id (i.e. feature ID). The spatial data contains information about where and attribute data can contain information about what, where, and why.
- Attribute data provides characteristics about spatial data.

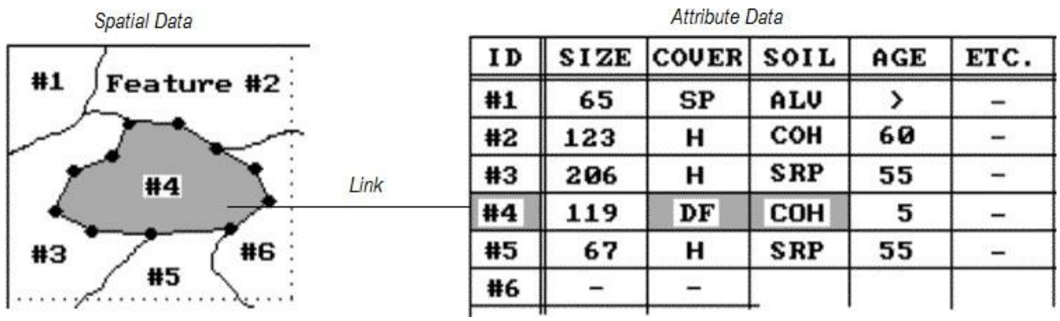


Figure: Attribute data and Spatial data linking

9.2.1 Joins:

- When our data was all in a single table, we could easily retrieve a particular row from
- that table.

- But if the data we are looking for is available in two or more tables then joins can be used to retrieve those data.
- Join is used to fetch data from two or more tables, which is joined to appear as single set of data. It is used for combining column from two or more tables by using values common to both tables.

There are several types of JOINs: INNER, LEFT OUTER and RIGHT OUTER; theyall do slightly different things, but the basic theory behind them all is the same.

a) Inner Join:

- An INNER JOIN returns a result set that contains the common elements of the tables,
- i.e. the intersection where they match on the joined condition.
- An INNER JOIN focuses on the commonality between two tables. When using an INNER JOIN, there must be at least some matching data between two (or more) tables that are being compared. INNER JOINs are the most frequently used JOIN operation.

location

fid	city	state	country
101	Chennai	TN	IND
102	Coimbatore	TN	IND
103	Nellore	AP	IND
104	Vellore	TN	IND
105	Tirupathi	AP	IND

city details

city	population	area	rainfall
Chennai	122345	9205.56	55.32
Coimbatore	122344	1234.36	120.36
Vellore	354212	1002.36	75.64
Kanchepuram	564321	4561.89	80.64
Nellore	123456	6455.45	98.65

Inner join

fid	city	state	country	city	populatio	area	rainfall
101	Chennai	TN	IND	Chennai	122345	9205.56	55.32
102	Coimbatore	TN	IND	Coimbatore	122344	1234.36	120.36
103	Nellore	AP	IND	Nellore	123456	6455.45	98.65
104	Vellore	TN	IND	Vellore	354212	1002.36	75.64

b) Left Outer Join:

- A LEFT JOIN or a LEFT OUTER JOIN takes all the rows from one table, defined as the left table, and joins it with a second table.
- A LEFT JOIN will always include the rows from the LEFT table, even if there are no matching rows in the table it is JOINed with.

location

fid	city	state	country
101	Chennai	TN	IND
102	Coimbatore	TN	IND
103	Nellore	AP	IND
104	Vellore	TN	IND
105	Tirupathi	AP	IND

city details

city	population	area	rainfall
Chennai	122345	9205.56	55.32
Coimbatore	122344	1234.36	120.36
Vellore	354212	1002.36	75.64
Kanchepuram	564321	4561.89	80.64
Nellore	123456	6455.45	98.65

Left outer join

fid	city	state	country	city	populatio	area	rainfall
101	Chennai	TN	IND	Chennai	122345	9205.56	55.32
102	Coimbatore	TN	IND	Coimbatore	122344	1234.36	120.36
103	Nellore	AP	IND	Nellore	123456	6455.45	98.65
104	Vellore	TN	IND	Vellore	354212	1002.36	75.64
105	Tirupathi	AP	IND	NIL	NIL	NIL	NIL

c) Right Outer Join:

- A RIGHT OUTER JOIN is similar to a LEFT OUTER JOIN except that the roles between the two tables are reversed, and all the rows on the second table are included along with any matching rows from the first table
- That is, A RIGHT JOIN will always include the rows from the RIGHT table, even if there are no matching rows in the table it is JOINed with.

location

fid	city	state	country
101	Chennai	TN	IND
102	Coimbatore	TN	IND
103	Nellore	AP	IND
104	Vellore	TN	IND
105	Tirupathi	AP	IND

city details

city	population	area	rainfall
Chennai	122345	9205.56	55.32
Coimbatore	122344	1234.36	120.36
Vellore	354212	1002.36	75.64
Kanchepuram	564321	4561.89	80.64
Nellore	123456	6455.45	98.65

fid	city	state	country	city	populatio	area	rainfall
101	Chennai	TN	IND	Chennai	122345	9205.56	55.32
102	Coimbatore	TN	IND	Coimbatore	122344	1234.36	120.36
104	Vellore	TN	IND	Vellore	354212	1002.36	75.64
103	Nellore	AP	IND	Nellore	123456	6455.45	98.65
NIL	NIL	NIL	NIL	Kanchepuram	564321	4561.89	80.64

9.2.3 Relates

- Relates can help us to discover specific information within our data. A relate (also called a table relate) is a property of a layer.
- We can create a table relate so that we can query and select features in one layer and see all the related features in another layer or table.
- Unlike joining tables, relating tables simply defines a relationship between two tables. The associated data isn't appended to the layer's attribute table like it is with a join. Instead, we can access the related data through selected features or records in your layer or table.

Relation Class:

- A relationship class is an object in a geo-database that stores information about a relationship between two feature classes, between a feature class and a non-spatial table, or between two non-spatial tables.
- Both participants in a relationship class must be stored in the same geo-database.
- A relationship class stores information about associations among features and records in a geo-database and can help ensure your data's integrity.
- Relates that are added to a layer or table in a map are essentially the same as simple relationship classes defined in a geo- database, except that they are saved with the map instead of in a geo-database.

****

10. OPEN DATABASE CONENCTIVITY (ODBC)

10.1 Open Database Connectivity (ODBC)

- An Open Database Connectivity (ODBC) is an interface that allows applications to access data in database management systems (DBMS) using SQL as a standard for accessing the data.
- ODBC permits maximum interoperability, which means a single application can access different DBMS.
- Application end users can then add ODBC database drivers to link the application to their choice of DBMS.
- Application end users can then add ODBC database drivers to link the application to their choice of DBMS.

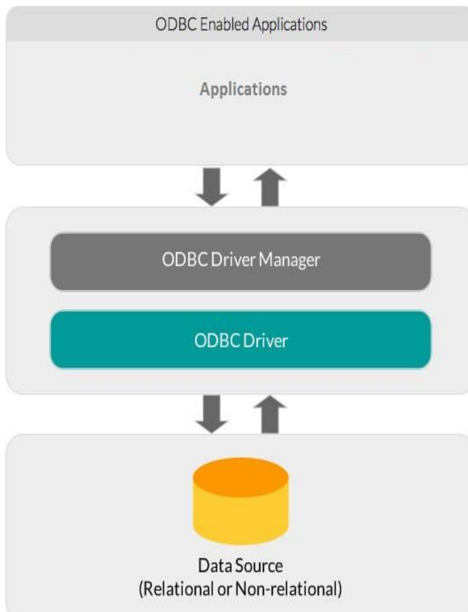


Figure: Architecture of ODBC

- The ODBC solution for accessing data led to ODBC database drivers, which are dynamic-link libraries on Windows and shared objects on Linux/UNIX.
- These drivers allow an application to gain access to one or more data sources.
- ODBC provides a standard interface to allow application developers and vendors of database drivers to exchange data between applications and data sources.

10.2 ODBC Driver Manager

- The ODBC Driver Manager loads and unloads ODBC drivers on behalf of an application.
- The Windows platform comes with a default Driver Manager, while non- windows

platforms have the choice to use an open source ODBC Driver Manager like unixODBC and iODBC.

- The ODBC Driver Manager processes ODBC function calls, or passes them to an ODBC driver and resolves ODBC version conflicts.

10.3 ODBC Driver

- The ODBC driver processes ODBC function calls, submits SQL requests to a specificdata source and returns results to the application.
- The ODBC driver may also modify an application's request so that the request conforms to syntax supported by the associated database.
- A framework to easily build an ODBC drivers is available from Simba Technologies, as are ODBC drivers for many data sources, such as Salesforce, MongoDB, Spark and more.

The following are the steps involved in connecting application programs with the database using ODBC API:

- Load ODBC driver: The forName() method of Class class is used to register the driverclass. This method is used to dynamically load the driver class.
- Establish Connection: The getConnection() method of DriverManager class is used to establish connection with the database.
- Prepare and Execute SQL Statement: The createStatement() method of Connection interface is used to create statement. The executeQuery() and execute() method is used to execute queries to the database.
- Process the result: The executeQuery() method returns the object of ResultSet that canbe used to get all the records of a table.
- Close connection: The *close()* method is used to close the connection in order to free the allocated resource used by the connection.

The below java code is used for connecting with mysql database using ODBC application programming interface.

```
Class.forName("org.gjt.mm.mysql.Driver");
Connection conn=null;
conn = DriverManager.getConnection(connectionURL, "root", "admin");
String sql = "SELECT * FROM STUDENTS";
PreparedStatement  pst = connection.prepareStatement(sql);
conn.open();
ResultSet=pst.executeQuery();
conn.close();
```

Thus the ODBC connection is established with connectivity code as given above.

****

11. GLOBAL POSITIONING SYSTEM (GPS) CONCEPT

11.1 Introduction to Global Positioning System

- The Global Positioning System (GPS) is a U.S.-owned utility that provides users with positioning, navigation, and timing (PNT) services.
- This system consists of three segments: the space segment, the control segment, and the user segment.
- The U.S. Air Force develops, maintains, and operates the space and control segments.
- GPS technology was first used by the United States military in the 1960s and expanded into civilian use over the next few decades.
- Today, GPS receivers are included in many commercial products, such as automobiles, smartphones, exercise watches, and GIS devices.

11.2 Segments of GPS

There are three major segments of GPS.
They are:
a) Space Segment
b) Control Segment
c) User Segment

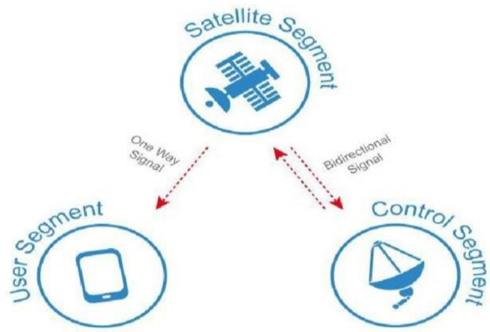


Figure: Three Segments of GPS

a) Space Segment:

- The GPS space segment consists of a constellation of satellites transmitting radio signals to users. The United States is committed to maintaining the availability of at least 24 operational GPS satellites, 95% of the time.
- To ensure this commitment, the Air Force has been flying 31 operational GPS satellites for the past few years. GPS satellites fly in medium Earth orbit (MEO) at an altitude of approximately 20,200 km (12,550 miles).
- Each satellite circles the Earth twice a day. The satellites in the GPS constellation are arranged into six equally-spaced orbital planes surrounding the Earth. Each plane contains four "slots" occupied by baseline satellites.
- This 24-slot arrangement ensures users can view at least four satellites from virtually any point on the planet.

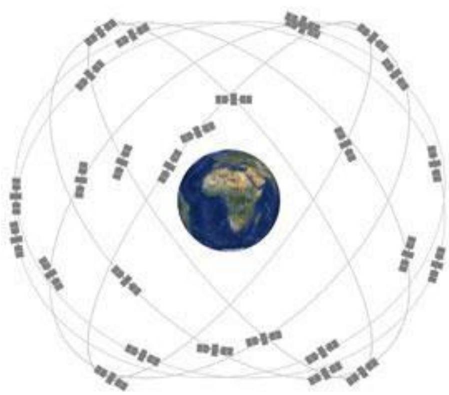


Figure: Constellation of satellites

- The Air Force normally flies more than 24 GPS satellites to maintain coverage whenever the baseline satellites are serviced or decommissioned.
- The extra satellites may increase GPS performance but are not considered part of the core constellation.
- In June 2011, the Air Force successfully completed a GPS constellation expansion known as the "Expandable 24" configuration.

- Three of the 24 slots were expanded, and six satellites were repositioned, so that three of the extra satellites became part of the constellation baseline.
- As a result, GPS now effectively operates as a 27-slot constellation with improved coverage in most parts of the world.

b) Control Segments

- The GPS control segment consists of a global network of ground facilities that track the GPS satellites, monitor their transmissions, perform analyses, and send commands and data to the constellation.
- The current Operational Control Segment (OCS) includes a master control station, an alternate master control station, 11 command and control antennas, and 16 monitoring sites.
- The locations of these facilities are shown in the map above.

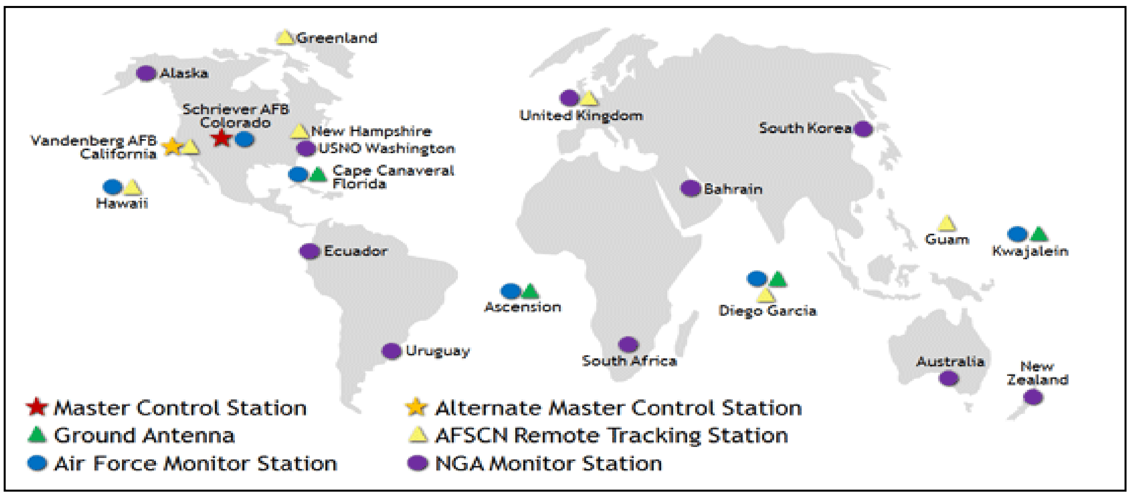


Figure: GPS Control Segments

- The GPS constellation delivers consistently high performance thanks to the dedicated efforts of its operators — the men and women of the U.S.
- Air Force's 2nd Space Operations Squadron (2SOPS) and the Air Force Reserve's 19th Space Operations Squadron (19SOPS)at Schriever Air Force Base, Colorado.

c) User Segments

- Like the Internet, GPS is an essential element of the global information infrastructure.
- The free, open, and dependable nature of GPS has led to the development of hundreds of applications affecting every aspect of modern life.
- GPS technology is now in everything fromcell phones and wristwatches to bulldozers, shipping containers, and ATM's.

****

13. GPS BASED MAPPING

13.1 Introduction GPS based Mapping

- The surveying and mapping community was one of the first to take advantage of GPS because it dramatically increased productivity and resulted in more accurate and reliable data.
- Today, GPS is a vital part of surveying and mapping activities around the world.
- When used by skilled professionals, GPS provides surveying and mapping data of the highest accuracy.

13.2 GPS Based Mapping

- GPS-based data collection is much faster than conventional surveying and mapping techniques, reducing the amount of equipment and labor required.
- A single surveyor can now accomplish in one day what once took an entire team weeks to do. GPS supports the accurate mapping and modeling of the physical world — from mountains and rivers to streets and buildings to utility lines and other resources.
- Features measured with GPS can be displayed on maps and in geographic information systems (GIS) that store, manipulate, and display geographically referenced data.
- Governments, scientific organizations, and commercial operations throughout the world use GPS and GIS technology to facilitate timely decisions and wise use of resources.
- Any organization or agency that requires accurate location information about its assets can benefit from the efficiency and productivity provided by GPS positioning.
- Unlike conventional techniques, GPS surveying is not bound by constraints such as line-of-sight visibility between survey stations.
- The stations can be deployed at greater distances from each other and can operate anywhere with a good view of the sky, rather than being confined to remote hilltops as previously required.
- GPS is especially useful in surveying coasts and waterways, where there are few land-based reference points.
- Survey vessels combine GPS positions with sonar depth soundings to make the nautical charts that alert mariners to changing water depths and underwater hazards.
- Bridge builders and offshore oil rigs also depend on GPS for accurate hydrographic surveys.
- Land surveyors and mappers can carry GPS systems in backpacks or mount them on vehicles to allow rapid, accurate data collection.

- Some of these systems communicate wirelessly with reference receivers to deliver continuous, real-time, centimeter-level accuracy and unprecedented productivity gains.
- To achieve the highest level of accuracy, most survey- grade receivers use two GPS radio frequencies: L1 and L2.
- Currently, there is no fully functional civilian signal at L2, so these receivers leverage a military L2 signal using "codeless" techniques.

****