

4.0 **DIGITIZATION, EDITING AND STRUCTURING OF MAP DATA**

- ❑ The process of digitizing existing maps is a transformation from one (analog) form of information to another (digital) form. Data input is the operation of encoding the data and writing them to the database.
- ❑ Two aspects of the data need to be considered separately for GIS, these are first the positional or geographic data necessary to define where the graphic or cartographic features occur, and second the associated attributes that record what the cartographic features represent.

4.1 **Entering the Spatial Data:**

- ❑ There is no single method of entering the spatial data to a GIS rather; there are several mutually compatible methods that can be used singly or in combination.
- ❑ The choice of method is governed largely by application, the available budget and the type of data being input.
- ❑ **The types of data encountered are:**
 - i) Existing maps
 - ii) Aerial photos
 - iii) RS data from satellite
 - iv) Point sample data
 - v) Data from census or surveys

❑ **Methods of Input Data in GIS:**

1. **Manual Input to Vector System:**

- The source data are entered as point, line and areas.
- The coordinates of the data are obtained from the reference grid overlaid on map. They can then be simply typed into a file or input to a Programme. E.g. sample coordinates points.

2. **Digitization:**

- A digitizer is an electronic device consisting of a tablet upon which the map or drawing can be placed.

- The coordinates of each point are entered into the computer through digitizing pad or puck. A pair of X, Y coordinates for a single point, a series of X, Y pairs for a line and a series of X, Y pairs with a common point for polygon or area.
- It is a laborious and time consuming process and hence mostly used for digitization of small area.

3. Semi- Automatic Digitization:

- An alternative to manual digitization process of entering spatial data in GIS is through semi-automatic digitization.
- Here, the initial position of the point is entered manually by the user for each linear feature to be digitized. The moment the initial position is entered the rest of the line is digitized automatically.
- But user has to keep track of the line that gets digitized automatically for error check. They may also need to be guided manually when they encounter junctions between two or more linear features.
- This method is found to be very helpful in generation of database for very large area but available in very few types of software like Arc/Info.

4. Automatic Scanning:

- Scanners are used to create a digital representation of maps in the form of 2D array of pixels or cells or grids. Each pixel is associated with a numerical value called cell or pixel value.
- The scanned image is entered directly as raster data into GIS system. They can be converted into vector format by digitizing each feature on the screen-using mouse. For each feature a separate layer is created and stored in vector format.
- The raster scanner is based on either drum or flat bed design.

5. Spatial data already in digital format:

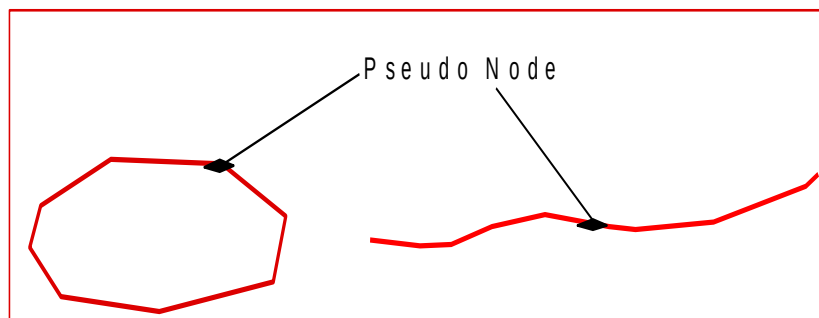
- Sensors and scanners on satellites and aeroplane generate digital images of the terrain. These images are generated in raster format composed of pixel associated with a value.

- Depending upon the spectral resolution of the sensor, the images are generated in different spectral bands. For each pixel, the value will be different in different bands depending upon the behavior of the object in that wavelength.
- These data being digital in nature can be used directly in GIS. Or analyzed in image processing softwares and results can be brought directly in the GIS where it can be combined with other data.

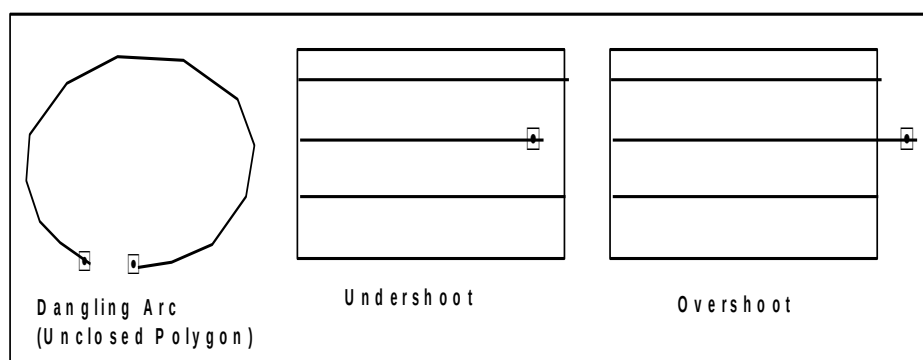
4.2 Error Detection and Correction:

- All digital data can be assumed to contain some amount of errors. These arise because of the inaccuracy of the source data or by operators mistake.
- As a general rule, if an error can be detected at the time of data entering into GIS, it will be easier to correct it then as compared to later time.
- Arc/Info has special characteristics to identify digitizing errors. It marks potential node errors with special symbols. However this facility is not found in other GIS softwares and that is why Arc/Info is considered to be the best for GIS data base creation.

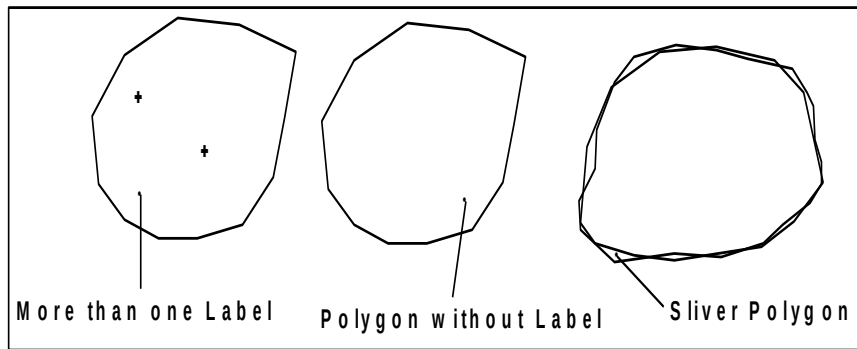
Pseudo Nodes: drawn with a diamond shaped symbol where a single line connects with itself or where only two arcs intersect. Pseudo node is not necessarily an error or a problem.



Dangling Nodes: represented by a square and shows unconnected arcs.



Polygon having more than one label or no label.



Sliver polygon – small-unwanted polygon created when one arc digitized twice.

Correcting Errors: Correcting errors are one of the most important steps in database creation. Unless errors are corrected, area calculation and any analysis and subsequent maps will be not valid.

Errors

Missing Arc (s) -

Missing label point -

More than one label -

A gap between 2 arcs or unclosed polygon-

An overshoot -

Incorrect user id value -

What should be marked?

draw them in

mark position and unique id

identify which one to delete

indicate which arc to be extended
or which node to move

indicate whether it should be
deleted

mark the correct value

Tolerances to be checked:

Arc/Info uses tolerances, expressed in digitizer unit, for coverage automation and update steps such as coverage registration, feature napping, coordinate spacing and so on.

RMS Error:

It is a measure of Tic registration accuracy during digitization and coverage transformation. Arc/info automatically calculates the RMS error when tics are used to register a map on the digitizer and during transform operations.

- The RMS value represents the amount of errors between original and new coordinate locations calculated by transformation process.

- The lower the RMS error, the more accurate the digitization or transformation will be.
- To maintain highly accurate geographic data the RMS error should be kept under 0.004 inches.

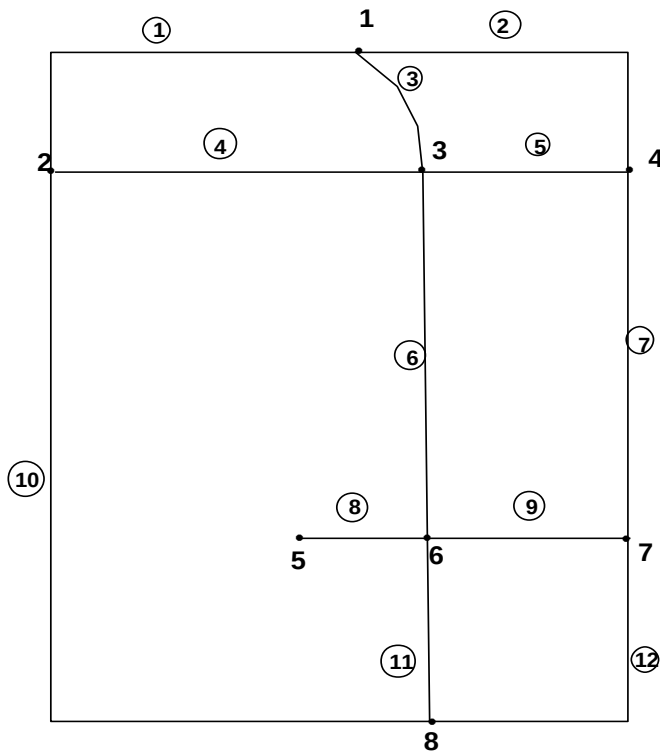
4.3 **Topology- Structuring of Map Data:**

- In a GIS, topology is used to represent the spatial relationships that exist between geographic data.
- Spatial relationships are the associations between geographic data based on their relative locations to one another.
- Topology is a mathematical representation of the physical relationships that exists between the geographical elements. The three major topological concepts of GIS are –
 - a) Arcs connect to each other at nodes (connectivity)
 - b) Arcs that connect to surround an arc define a polygon (Containment or area definition)
 - c) Arcs have direction and left and right sides (Contiguity)

Connectivity:

- The two points (two X, Y pairs) along the arc, called vertices, define the shape of the arc.
- The end point of arc called nodes. Each arc has two nodes – from- node and to- node. Arc joins only at nodes.
- By tracking all the arcs that meet at nodes, GIS understands which are connected.
- In the example, arcs 3,4,5 and 6 all join at node 3.
- With this information, the computer knows that it is possible to travel along arc 5 and turn onto arc 3 because they share a common node 3. But it is not possible to turn directly from arc 5 onto arc 9 because arc 5 and arc 9 do not share a common node. This is called connectivity and topology is **called – Arc-Node Topology.**

Connectivity – Arc Node Topology



• 1, 2 etc - Node Number

① ② etc - Arc Number

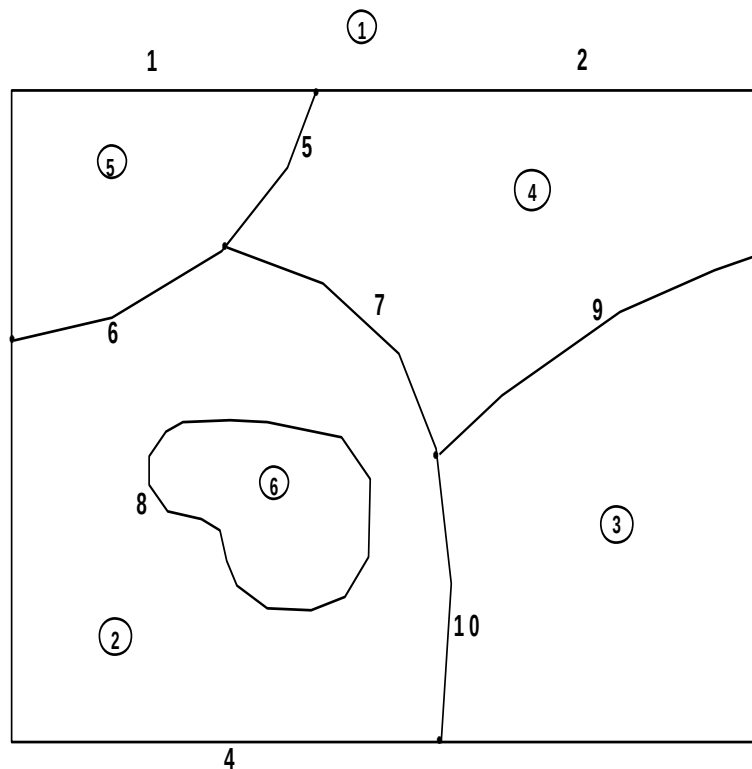
Arc #	From-Node	To-Node
1	2	1
2	1	4
3	1	3
4	2	3
5	4	3
6	3	6
7	4	7
8	5	6
9	6	7
10	2	8
11	6	8
12	8	7

Arc #	X, Y Pair
1	55, 57, 87
2	87, 117, 11
3	87, ... 9
4	55, 9 5
5	
6	
7	
8	
9	
10	
11	
12	

Containment:

- Containment is nothing but area definition. Polygons are represented as a series of X, Y coordinates that connect to enclose an area.
- Arc/Info stores the arc defining the polygon rather than closed set of X, Y pairs.
- In the example, arcs 4, 6, 7 and 8 comprise polygon 2.
- Though an arc may appear in the list of arcs for more than one polygon, each is stored only once.
- This reduces the amount of data in the database and also ensures that the boundaries of adjacent polygon do not overlap. This is called containment and the topology is **called- Polygon – Arc Topology**.

Containment – Polygon Arc Topology



1 2 3 - Arc Number

① ② - Polygon Number

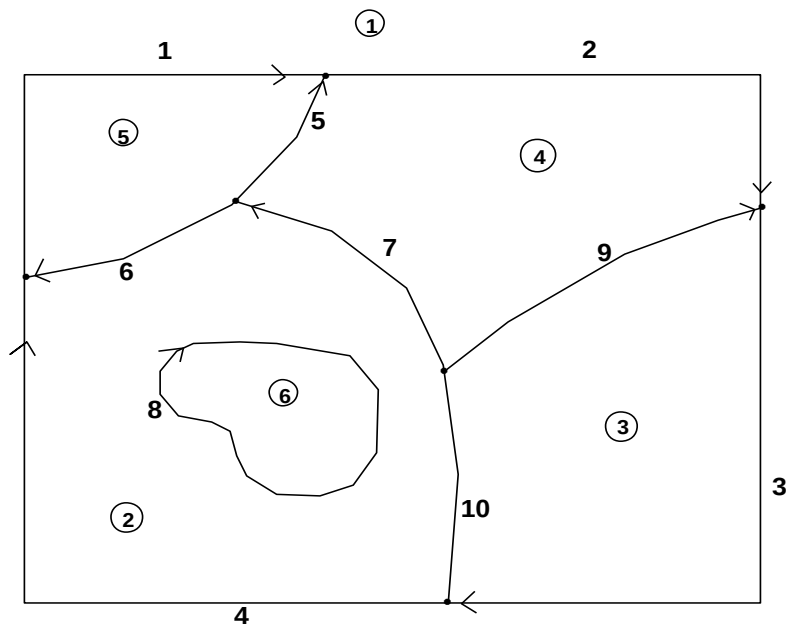
Poly #	Line #
2	4, 6, 7, 10, 0, 8
3	3, 10, 9
4	7, 5, 2, 9
5	1, 5, 6
6	8

Arc #	X, Y Pair
1	55, 57, 87
2	87, 117, 11
3	87, ...9
4	55, 95
5	
6	
7	
8	
9	
10	
11	
12	

Contiguity:

- Contiguity defines the direction of each arc and neighborhoodness because each arc has directions – f-node and t-node.
- GIS maintains a list of polygons on the left and right sides of each arc. Polygons sharing a common arc are adjacent.
- In the example, arc 9 is having polygon 4 at the left side and polygon 3 at the right side. This is called contiguity and the topology is **called Left-Right Topology.**

Contiguity – Left Right Topology



1 2 3 - Arc Number

① ② - Polygon Number

Arc #	L-Poly	R-Poly
1	1	5
2	1	4
3	1	3
4	1	2
5	5	4
6	2	5
7	2	4
8	2	6
9	4	3
10	3	2

Arc #	X, Y Pair
1	55,57,87
2	87,117,11
3	87,.....9
4	55,9 5
5	
6	
7	
8	
9	
10	
11	
12	