PDHonline Course L155 (5 PDH)

Data Models and Data processing in GIS

Instructor: Steve Ramroop, Ph.D.

2020

PDH Online | PDH Center

5272 Meadow Estates Drive
Fairfax, VA 22030-6658
Phone: 703-988-0088
www.PDHonline.com

Slide 1

Lecture 3 Content

- Geographic Information Systems (GIS)

Data Models, Data Structure and Data Management

This lecture is the continuation of the GIS topic identified in the course description which is Data Models, Data Structure and Data Management.

Slide 2

Lecture Content

- **Spatial Data Structures**
  - – **complexity of reality**
  - – **vector structure representation**
    - ➢ **spaghetti data structure**
    - ➢ **topological data structure**

This slide presents the content of lecture 3. The content of this lecture presents the important consideration of selecting suitable data for a given GIS application. It also presents one of the two common data structures/data formats used in GIS applications. In this lecture is shown common characteristics of the vector data structure; its various forms of creation; and it s various forms of storage.

Slide 3

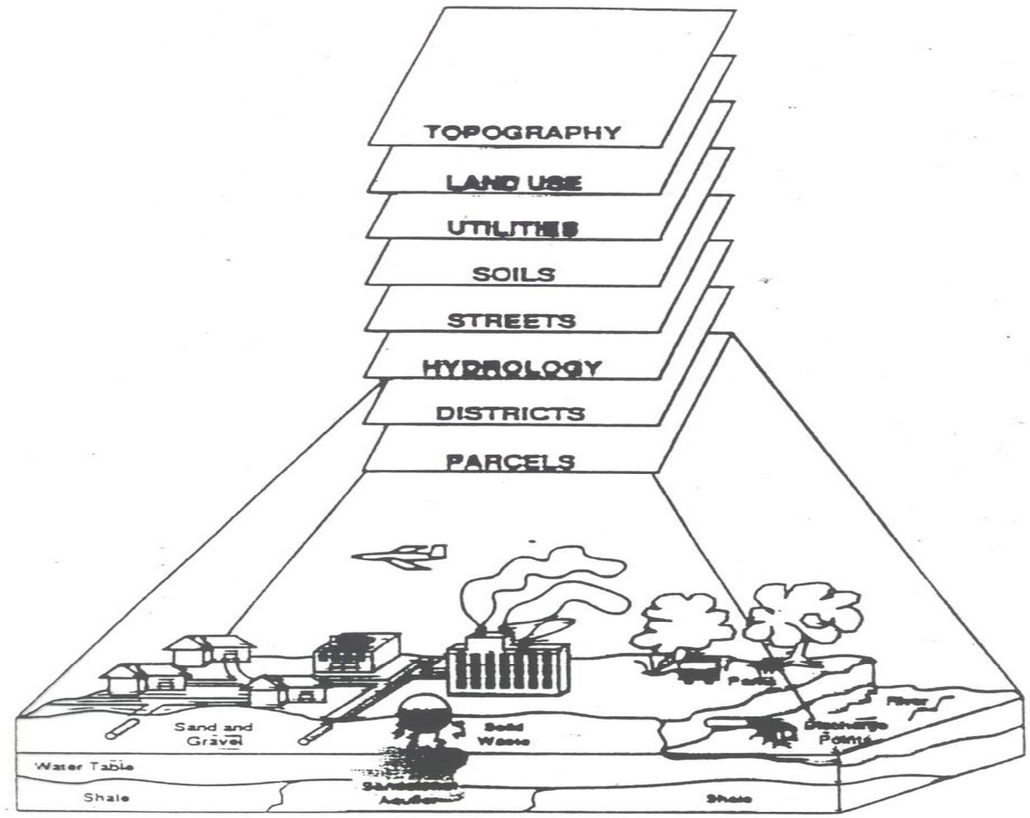
❑ Spatial Data Structures

- This is a representation of the relative location of features / phenomenon occurring in reality relative to one another
- The relative location of features is the strong point in GIS when compared to other database systems
- What data set is needed is a major consideration in a GIS application and in addition what is a suitable data structure

This slide gives some general statements regarding spatial data structures and what makes spatial data different from non-spatial data. What data is need for the intended application is a major consideration, but so too is the consideration of what type of data structures is suitable.

Slide 4

Complexity of reality



The real world consists of many geographies

<u>Notes on complexity of reality:</u>

Reality has a multitude of data sets. The choice of data sets is dependent upon the GIS application being developed, the budget of the GIS project, and the accuracy of the required final result(s). The more data sets collected the greater is the requirement on the computer hardware needed to store and manipulate the data sets. From the previous figure eight data sets, (Topography, Land Use, ….. Districts, and Parcels) were selected from reality, however, there are other data sets that exists. For example: Air pollution; Sub-surface data; Airspace; etc. are not included. A balance is required to select the optimal data sets which is suitable and appropriate to the intended GIS application

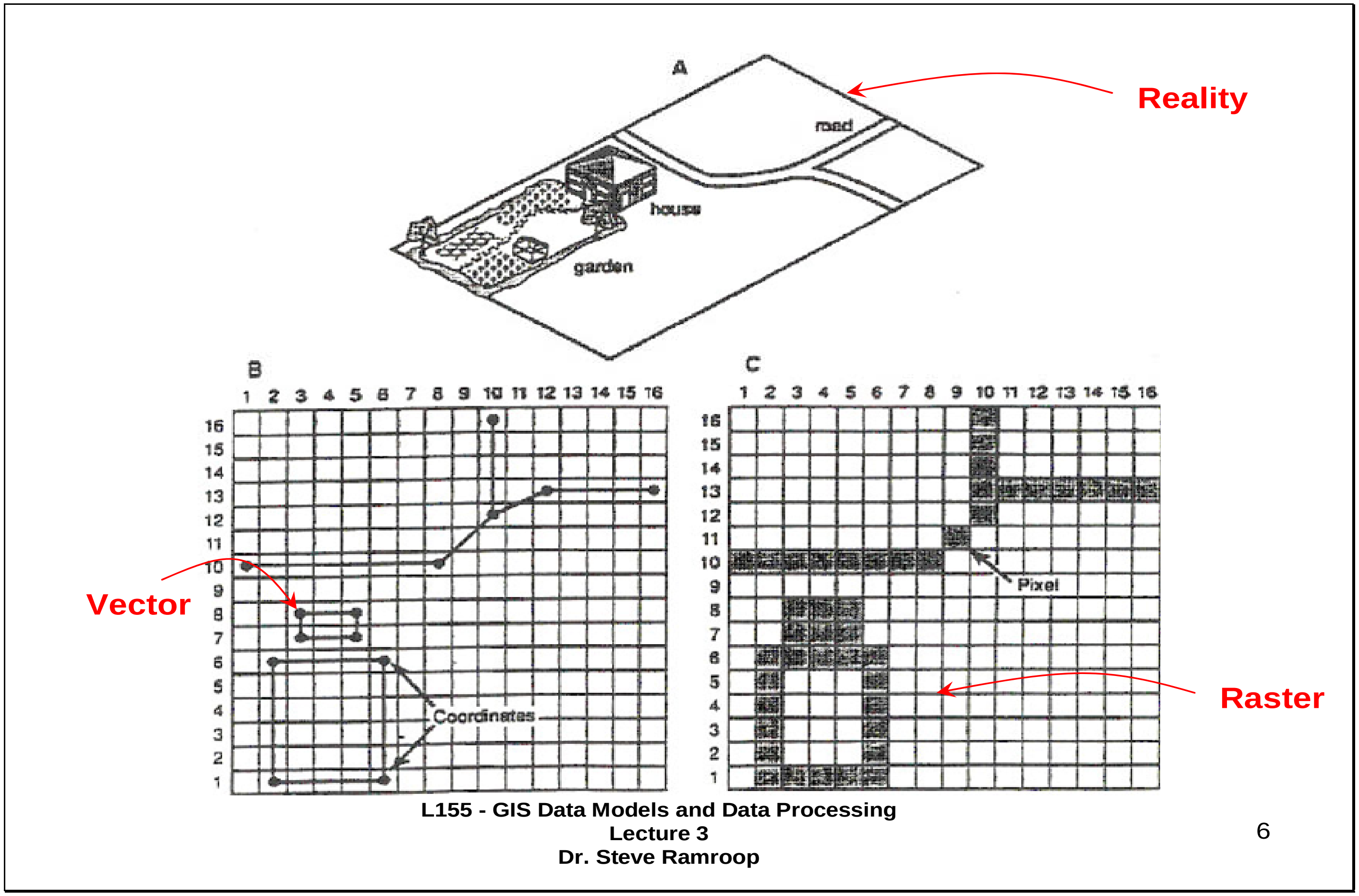
Slide 5

❑ Representing objects from the real world in a GIS:

- ➢ **The central element of a GIS is the spatial database consisting of an integrated collection of data about spatial objects and their attributes**
- ➢ **The spatial database is a realization of a model, used to describe reality**
- ➢ **Two data structures are commonly recognized:**
  - – **Vector Structure**
  - – **Raster Structure**

In order to represent features from reality two structures are identified: Vector and raster. They are the two data structures which are used in all GIS software.

Slide 6

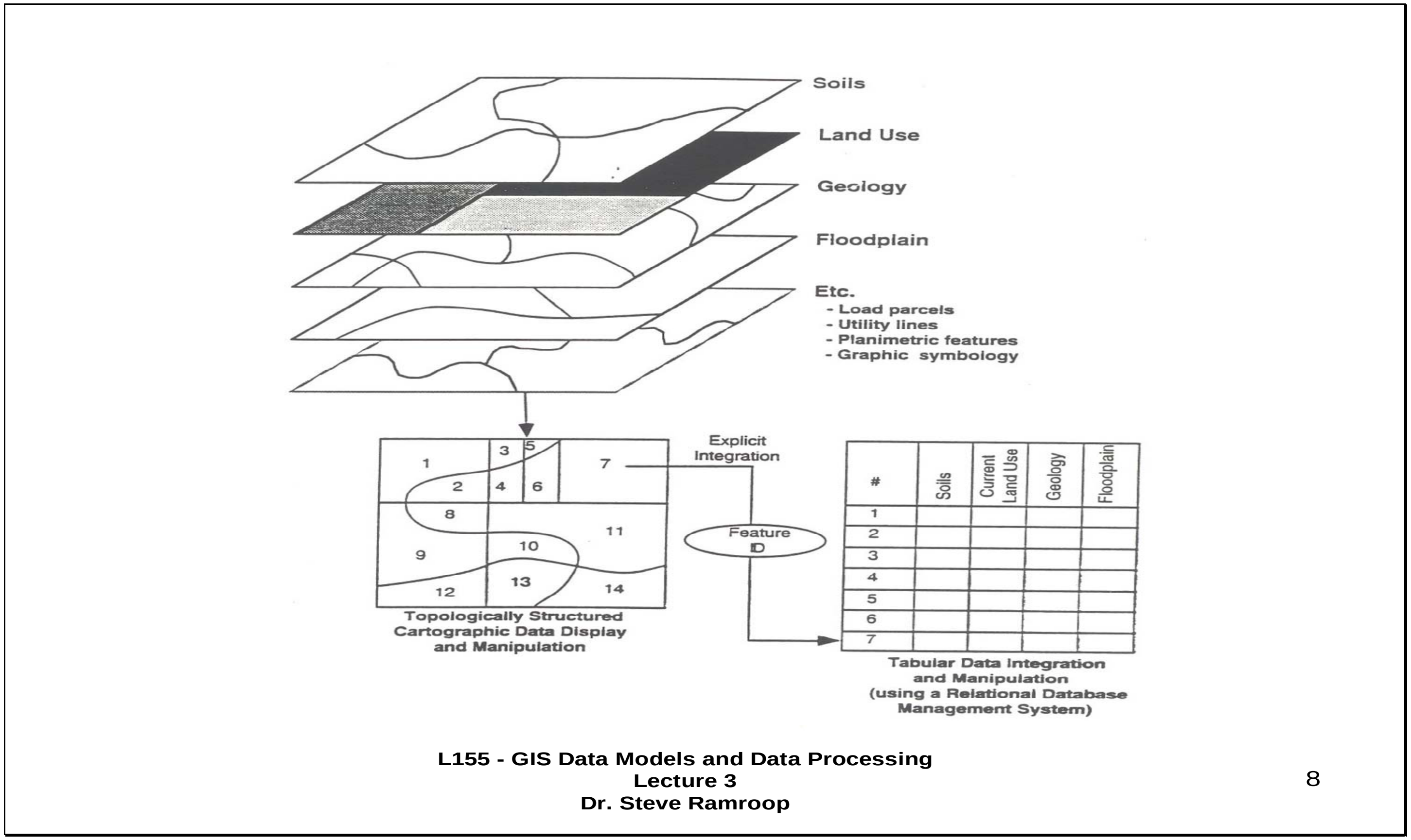


This slide presents the representation of the two types of data structures. From reality the same entities were selected for the intended GIS application. The house, garden and road data sets are required. Using vector data structure, coordinates are collected and stored; while the raster data structure pixels are collected and stored.

Slide 7

This slide identifies the three requirements of all spatial databases. They are: coordinate system; attribute data; and the topology in the data sets.

Slide 8



This slide shows the result of overlaying a number of data sets. The resulting map is a combination of polygons formed from the series of polygons of the various data sets. The final output map will have unique numbers for each resulting polygon, with each polygon containing the cumulative information in its attribute data tables.

Slide 9

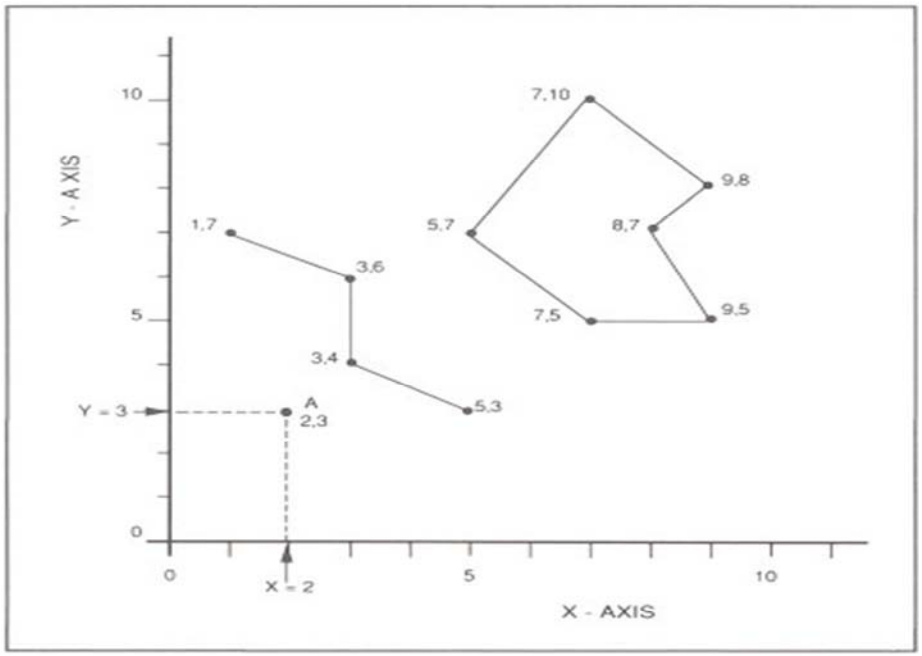
A. VECTOR STRUCTURE REPRESENTATION

– **The fundamental cartographic primitive is a point**
– **Objects are created by connecting points with straight lines or arcs of circles**
– **Vector structure is used to draw areas, lines and points to represent real world objects**
– **it does not fill space**
– **TELLS WHERE EVERYTHING OCCURS**

This slide gives some characteristics of the Vector data structure. The basic primitive is the point. A series of points represents a line. A line that starts and ends with the same point represents a polygon. Note the phrase that a vector structure “Tells where everything occurs”…..if features of importance to the intended GIS application that relates to the phenomena being mapped is there, then such features will be stored in the GIS database.

Slide 10

Representing points, lines, and polygons as XY coordinate strings



L155 - GIS Data Models and Data Processing
Lecture 3
Dr. Steve Ramroop

10

This slide shows a simplistic representation of point, line and polygons using the vector structure. It is what is done when plotting data using basic mathematics. The following slides presents some of the common methods used to create data sets having a vector structure.

Slide 11

How to create vector data?

1) Digitizing

- main method of data input
- utilizes the existing hardcopy maps as a basis for digitizing
- Three methods of digitizing (all of which have detailed methodologies) :
  - a) Manual digitization
    - utilizes the existing maps to digitize
    - requires georeferencing of information (ground control)
    - point or stream mode of digitization is possible

Digitizing is the most common form of vector data structure creation. This is a popular method when GIS first came into the professional market. There was a rush to get the hardcopy maps into digital data. This was done using digitizing tables connected to computers. There are three methods of digitizing as presented.

Slide 12

b) Scan digitization

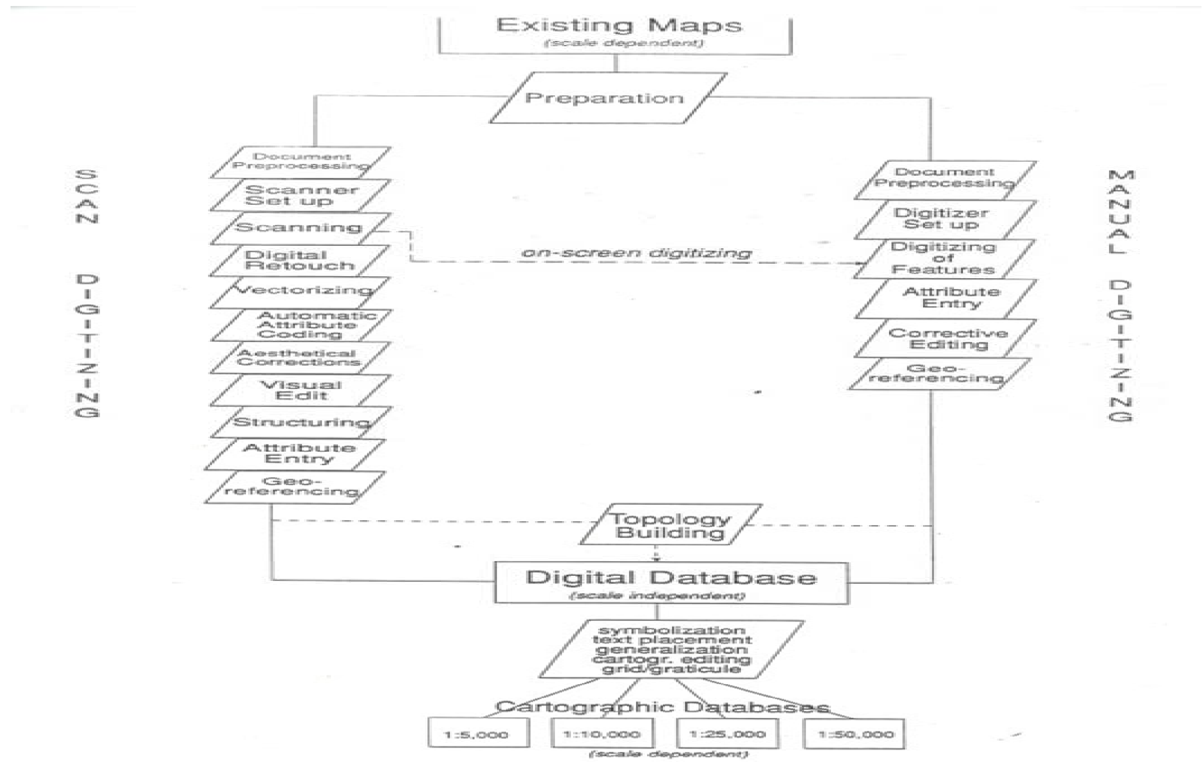
- makes use of a scanner
- produces raster data initially, then converted (via a vectorization process) to vector data
- certain inaccuracies are involved (discussed in an earlier lecture)

c) On-screen digitization

- uses the scanned raster image as a back-drop from which the features are manually digitized
- requires a computer environment

These are 2 of the 3 methods of digitizing methods. They are self explanatory.

Slide 13

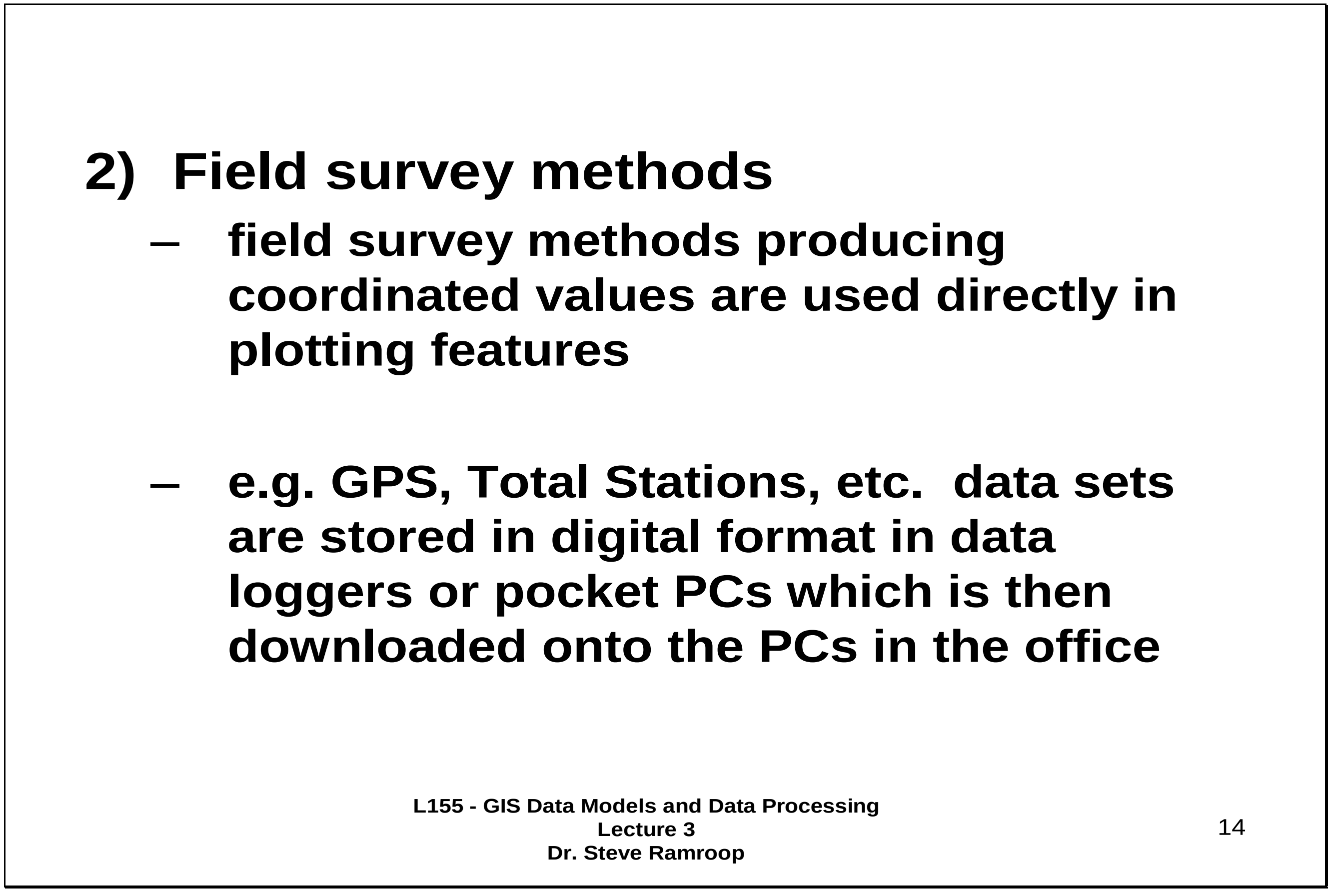


L155 - GIS Data Models and Data Processing
Lecture 3
Dr. Steve Ramroop

13

This slide compares scan digitizing to manual digitizing. Note that scan digitizing (even though it is initially faster), requires more post-processing than manual digitizing. Which is better (scan or manual)? The choice of method depends on the intended GIS application and the resources available for the intended GIS application.

Slide 14



The second method of creating data having a vector structure is through field survey methods. X and Y coordinates are collected directly out in the field and then downloaded in the office after field work. Such data are point data that is used to define the points, draw lines and polygons.

Slide 15

- **<u>Vector data Structures</u>**
  1. **Spaghetti Data Structure**
  2. **Topological Structure**

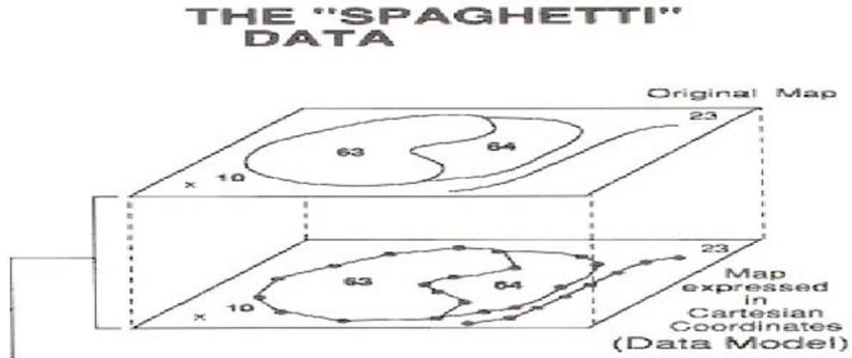
1. Spaghetti Data Structure

- **A collection of coordinate strings with no inherent structure hence the name “spaghetti”**
- **The data structure is really the map expressed in Cartesian Coordinates**
- **Although the spatial features are recorded, the spatial relationships between these features are not encoded**

  e.g. adjacent features from one another
- **Very inefficient for spatial analyses since it must be computed**
- **Data is stored redundantly**
- **The following figure presents an example of a spaghetti structure**

There are two types of vector data structures. These are the structure types which are used to store vector data. The characteristics of the spaghetti data structure are identified on this slide.

Slide 16



Data Structure

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

This slide gives a graphic example of how the vector data structure is stored using the spaghetti data structure. Notice along the shared boundary between the parcels, the coordinates for the boundary are stored redundantly.

Slide 17

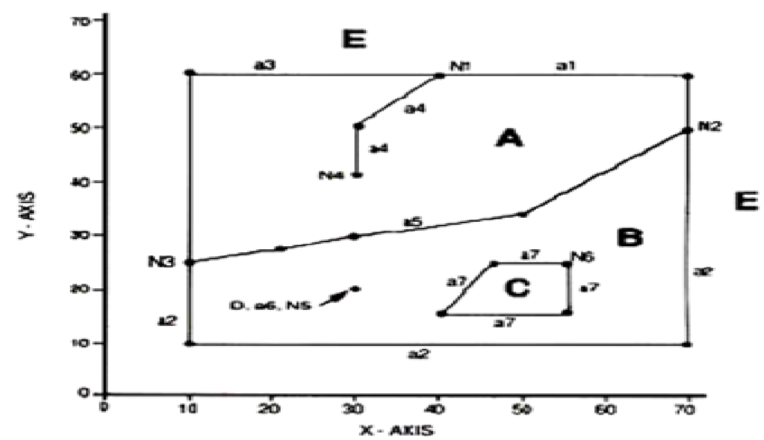
2) Topological Structure

- **most widely used method of encoding spatial relationships**
- **Also called Map-based topology**
- **"Topology" is the mathematical method used to define spatial relationships between points, lines and polygons**
- **Arc-Node data structure is a typical form of the topological structure**
  - **ARC - a series of points that start and end at a node**
  - **NODE - an intersection point where two or more arcs meet**
  - **POLYGON - comprise of a closed chain of arcs that represents the boundaries of the area**
- **topology is recorded in GIS using three data tables, one for each spatial element (Node, Lines, Polygon) and the fourth table store coordinate data**
- **the following figure presents an example of a topological structure**

This is the second vector data structure which is the most popular data structure. It is adopted in the ArcGIS software. Topology is defined by the software such that are, node, and polygon data tables are automatically generated when the GIS software builds topology.

Slide 18

Arc-Node Topological Structure



SPATIAL DATA ENCODING

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

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

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

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

This is an example of the arc, node, and polygon topology using the vector arc-node topological structure. Each table is defined by the GIS software.

The four tables are: Polygon Table; Node Table; Arc Topology; and Arc Coordinate Data.

For the Polygon Table, two columns are defined. One column for the Polygon IDs and the other column contain the arcs that define each polygon.

For the Node Topology Table, two columns are defined. One column for the Node IDs and the other column contain the arcs that define each node.

For the Arc Topology Table, five columns are defined. One column contains the Arc IDs. The second column has the start node ID while the third column has the end node ID. The fourth and the fifth column contain the left and right polygons when moving in the direction of the start and end node. Note that the direction along the arc is important for the arc topology table. Example ‘a1’ starts from ‘N1’ and ends at node ‘N2’. Traveling in that direction the left polygon is ‘E’ and the right polygon is ‘A’.

For the Arc Coordinate Data Table, four columns are defined. The first column contains the Arc IDs. The second column contains the start x and y coordinates. The third column contains any intermediate x and y coordinates. The fourth column contains the end x and y coordinate of the end node for the arc.

Slide 19

<u>The Four Tables in the topological structure:</u>

a) **Polygon Topology Table**
   - **shows the arcs that make up the boundaries of each polygon**
   - **polygons can have islands within them**
   - **a point can be considered a polygon with no area**

b) **Node Topology Table**
   - **each node is defined by the arcs to which it belongs**

c) **Arc Topology Table**
   - **defines the relationship of the nodes and polygons to the arcs**
   - **end points are distinguished by designating one node as the *"start"* or *"from"* node and the other as the *"end"* or *"to"* node**

d) **Arc Coordinate Data Table**
   - **defines the ground control coordinates of the start, end and intermediate coordinates of arcs and points**

This slide summarizes the characteristics of the four tables in the vector arc-node topology tables.

➢ Rule-based Topology

- **It is a topology that is implemented by using a set of rules (defined within a geodatabase) that describes the relationships between and among features**
- **It is defined as: ‘a set of rules defining the spatial relationships between features and the constraints placed on those spatial relationships’**
- **Example for point features:**
  - **Must be covered by boundary of polygon**
  - **Must be within a polygon**
  - **Must be covered by the end point of line……etc**

This slide summarizes the characteristics of the Rule-based topology also adopted by some GIS software vendors. It is another method used for storing vector topology.

Slide 21

… The End …