The Nature of Geographic Data

- Most features or phenomena occur as either:
  - discrete entities (objects/PLPs)
  - continuously varying phenomena (fields/CS)
  - some could go either way

GIS Data Representation

- How PLPs and CS are stored
- Two primary GIS data models
  - Raster
  - Vector
Data Structures and Data Models

- “Raster” and “Vector” refer to a particular data model.
- A data structure refers to a particular implementation of either the raster or vector model.

Topology

- The spatial relationships between things.
- An important distinction between data models and data structures.
- Can be stored or calculated “on the fly”.

Raster – Vector Data Models

- Numerous differences in terms of:
  1. 
  2. 
  3. 
  4. 
  5. 
  6. 
Database Models

- Also differences in how attribute data are stored

Raster Data Model

- Study area divided up into a regular array of discrete grid cells of uniform shape and size

Raster Representation of PLPs

- Figure 5.3: Representation of point, line, and area features; raster format on the left and vector format on the right.
**Sources of Raster Data**

- Common raster data sources include:
  - Imagery (satellite, aerial photography)
  - Classified imagery
  - DEMs, DSMs, DTMs, etc.
  - Scanned products (e.g. maps)
  - Interpolated or generated ‘surfaces’

**Thematic vs. Discrete**

- Thematic rasters - discrete representations
- Continuous rasters - continuous representations

**Accuracy of Raster Data**

- Determined by grid cell size
- Referred to as resolution
  - User defined
  - Sensor design
Projecting Raster Data

- Grid cells represent same location on the ground
- Georeferencing aligns raster layer to real world coordinate system

Advantages of Raster Model

- Advantages:
  - Easy to understand
  - Mathematical operations
  - Representing continuous phenomena
**Disadvantages of Raster Model**
- Disadvantages:
  - Representing discrete phenomena
  - Data redundancy
  - No topology (for discrete data)
  - Spatial accuracy

**Raster Data Structures**
- Method by which raster model is implemented in a particular GIS software application
- Typically, user is unaware of how data structure works

Cell by Cell

```
Row 1: 0 0 0 0 1 1 0 0  
Row 2: 0 0 0 1 1 1 0 0  
Row 3: 0 0 1 1 1 1 1 0  
Row 4: 0 0 1 1 1 1 1 0  
Row 5: 0 0 1 1 1 1 1 0  
Row 6: 0 1 1 1 1 1 0 0  
Row 7: 0 1 1 1 1 1 1 0  
Row 8: 0 0 0 0 0 0 0 0  
```

Figure 5.6 The cell-by-cell data structure records each cell value by row and column.
Run Length Encoding
8,8,1
0,4,1,2,0,2
0,3,1,3,0,2
0,2,1,5,0,1
0,2,1,5,0,1
0,1,1,6,0,1
0,1,1,6,0,1
0,8

Figure 5.9
The run length encoding method records the cell values in runs. Row 1, for example, has two adjacent cells in columns 5 and 6 that are gray or have the value of 1. Row 1 is therefore encoded with one run, beginning in column 5 and ending in column 6. The same method is used to record other rows.

Block Codes
1,1,2,4,6,7,2
4,1,1,5
25,1,3,3

Chain Codes
0=N, 1=E, 2=S, 3=W
1,5,1
1.1.2.2.1.1.2.4.3.5.0,1,1,1,0.3
1.1.0,1,1,1,0,1
Quad Tree

Figure 5.10
The regional quad tree method divides a raster into a hierarchy of quadrants. The division stops when a quadrant is made of cells of the same value (gray or white). A quadrant that cannot be subdivided is called a leaf node. In the diagram, the quadrants are indexed spatially; 0 for NW, 1 for SW, 2 for SE, and 3 for NE. Using the spatial indexing method and the hierarchical quad tree structure, the gray cells can be coded as 02, 03, and so on. See text for more explanation.

Mr SID

Figure 5.11
The Haar wavelet and the wavelet transform: (a) Three Haar wavelets at three scales (resolutions); (b) A simple example of the wavelet transform.

Vector Data Model

- Study area is referenced to a coordinate system that is used to identify the location of points, lines, and polygons using x, y, (z) coordinates.
**Vector Representation of PLPs**

Vector Data Sources

- Common vector data sources include:
  - NTDB data (National Topographic Database)
  - MLI data (1:20k roads, hydrology, etc.)
  - Tiger Data (U.S. Census)
  - GPS or Total Station survey data
  - Digitized data

**Vector Data Model**

- Advantages:
  - Representation of discrete objects
  - No data redundancy
  - Spatial accuracy
  - Topology
Vector Data Model

- Disadvantages:
  - Mathematical analysis
  - Not suitable for continuously distributed data

Vector Data Models

- Non-topological

- Topological

Today there are two basic vector data models

- Georelational Vector Data Model
  - ArcInfo coverages & ArcView shapefiles

- Object-based Vector Data Model
  - ArcGIS geodatabases
Figure 3.2
Based on the georelational data model, an ArcInfo coverage has two components: graphic files for spatial data and INFO files for attribute data. The label connects the two components.

Figure 3.7
The data structure of a point coverage.

Figure 3.8
The data structure of a line coverage.
Georelational Data Model

- ArcView shapefiles
  - Non-topological vector data model
- Topology on the fly
- Advantages
Object-based Model
- Based on object oriented programming theory
  - Objects:
    - represent features
    - have defined properties and methods
    - can be organized into classes

Geodatabase Model
- Geospatial features represented using objects
- Features have properties and methods
- Features and attributes stored in single file/database
Table 4.1 Topology rules in the geodatabase data model
ArcGIS Geodatabase Structure

- Advantages
  - Validation rules
    - Topology Rules
    - Attribute Domains
    - Relationship Rules
    - Connectivity Rules
    - Custom Rules

Raster – Vector Conversion

- Rasterization
- Vectorization

Figure 5.12
On the left is an example of conversion from vector to raster data, or rasterization. On the right is an example of conversion from raster to vector data, or vectorization.
**Why Convert?**

- **Rasterize**
  - Statistical or other mathematical operations

- **Vectorize**
  - Extract thematic information from classified RS imagery (roads, rivers, land cover classes)
  - Scanned maps (ArcScan)

---

**Why Integrate?**

- As an image background
- Image hotlinks
- To subset and classify imagery
- Still seldom used in analysis together

---

**Rasterization**

- Specify output grid cell size
- Attribute to be used as cell value
- Grid cell size v. important
  - Min. dimension of smallest feature
- Inevitable distortion
Line Rasterization
- Straight line rasterization
- Bresenham's Line Algorithm

Polygon Rasterisation
- How are boundary pixels handled?
  - Central Point
  - Dominant Unit
  - Ranked List
Scan Line Coherence

- Common polygon rasterisation algorithm
  - Fill grid cells b/w boundary cells

Vectorisation

- Vital for feature extraction
  - Scanned map products
  - RS imagery second largest source of geospatial data

Figure 4.11
Large format drum scanners. (Courtesy of GTCO Calcomp, Inc.)
**Scanning Maps**

- Considerations:
  - Colour depth
  - Resolution

**Vectorization**

- For classified RS images
  - Reclassification
- For scanned maps
  - Image preprocessing
  - Thresholding
  - Bilevel images

*Figure 6.12*
A binary scanned file: the lines are soil lines, and the black areas are the background.
**Vectorization**

- Editing/pre-processing boolean image
  - Remove gaps
  - Noise
  - Line thinning

**Figure 6.13**
A raster line in a scanned file has a width of several pixels.

**Vectorisation**

- Vectorization of boolean image
  - Node detection
  - Cleaning
  - Build topology