8. Geographic Data Modeling

Geographic Information Systems and Science
SECOND EDITION
Paul A. Longley, Michael F. Goodchild, David J. Maguire, David W. Rhind
© 2005 John Wiley and Sons, Ltd
Outline

- Definitions
- Data models / modeling
- GIS data models
  - Topology
- Example
  - Water facilities
Definitions

• **Data model**
  - set of constructs for representing objects and processes in the digital environment

• **Representation**
  - Focus on conceptual and scientific issues
Role of a Data Model
Levels of Data Model Abstraction

- Reality
- Conceptual Model
- Logical Model
- Physical Model

Increasing Abstraction

Human-oriented

Computer-oriented
Two representations of San Diego, California: (A) panchromatic SPOT raster satellite image collected in 1990 at 10 m resolution; (B) vector objects digitized from the image.
GIS Data Models & Applications

- CAD
- Graphical
- Image
- Raster/Grid
- Network
- Geo-relational
- TIN
- Object

- Engineering design
- Simple mapping
- Image processing and analysis
- Spatial analysis / modeling
- Network analysis
- Geoprocessing geometric features
- Surface /terrain analysis / modeling
- Features with behavior
Raster and Vector Models

- **Raster** – implementation of field conceptual model
  - Array of cells used to represent objects
  - Useful as background maps and for spatial analysis

- **Vector** – implementation of discrete object conceptual model
  - Point, line and polygon representations
  - Widely used in cartography, and network analysis
Raster – Satellite Imagery
Vector Data Model

Points

<table>
<thead>
<tr>
<th>Point number</th>
<th>(x,y) coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(2,4)</td>
</tr>
<tr>
<td>2</td>
<td>(3,2)</td>
</tr>
<tr>
<td>3</td>
<td>(5,3)</td>
</tr>
<tr>
<td>4</td>
<td>(6,2)</td>
</tr>
</tbody>
</table>

Polylines

<table>
<thead>
<tr>
<th>Polyl ine number</th>
<th>(x,y) coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(1,5) (3,6) (6,5) (7,6)</td>
</tr>
<tr>
<td>2</td>
<td>(1,1) (3,3) (6,2) (7,3)</td>
</tr>
</tbody>
</table>

Polygons

<table>
<thead>
<tr>
<th>Polygon number</th>
<th>(x,y) coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(2,4) (2,5) (3,6) (4,5) (3,4) (2,4)</td>
</tr>
<tr>
<td>2</td>
<td>(3,2) (3,3) (4,3) (5,4) (6,2) (5,1) (4,1) (4,2) (3,2)</td>
</tr>
</tbody>
</table>
Topology

- Science and mathematics of geometric relationships
  - Simple features + topological rules
  - Connectivity
  - Adjacency
  - Shared nodes / edges

- Topology uses
  - Data validation
  - Spatial analysis (e.g. network tracing, polygon adjacency)
Topological Polygon Data Layer
Contiguity of Topological Polygons
Geo-relational Polygon Dataset

![Diagram showing soils layer, nodes, and polygons with attribute data](image)

<table>
<thead>
<tr>
<th>ID</th>
<th>Soil</th>
<th>Class</th>
<th>Suitability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A3</td>
<td>113</td>
<td>HIGH</td>
</tr>
<tr>
<td>2</td>
<td>C6</td>
<td>95</td>
<td>LOW</td>
</tr>
<tr>
<td>3</td>
<td>B7</td>
<td>212</td>
<td>MODERATE</td>
</tr>
<tr>
<td>4</td>
<td>B13</td>
<td>201</td>
<td>MODERATE</td>
</tr>
<tr>
<td>5</td>
<td>Z22</td>
<td>86</td>
<td>LOW</td>
</tr>
<tr>
<td>6</td>
<td>A6</td>
<td>77</td>
<td>HIGH</td>
</tr>
<tr>
<td>7</td>
<td>A1</td>
<td>117</td>
<td>LOW</td>
</tr>
</tbody>
</table>
TIN Surface of Death Valley, California
TIN Surface of Death Valley, California
TIN Surface of Death Valley, California
A TIN is a topologic data structure that manages information about the nodes that comprise each triangle and the neighbours to each triangle.

<table>
<thead>
<tr>
<th>Triangle</th>
<th>Node list</th>
<th>Neighbours</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1, 2, 3</td>
<td>- , B, D</td>
</tr>
<tr>
<td>B</td>
<td>2, 4, 3</td>
<td>- , C, A</td>
</tr>
<tr>
<td>C</td>
<td>4, 8, 3</td>
<td>- , G, B</td>
</tr>
<tr>
<td>D</td>
<td>1, 3, 5</td>
<td>A, F, E</td>
</tr>
<tr>
<td>E</td>
<td>1, 5, 6</td>
<td>D, H, -</td>
</tr>
<tr>
<td>F</td>
<td>3, 7, 5</td>
<td>G, H, D</td>
</tr>
<tr>
<td>G</td>
<td>3, 8, 7</td>
<td>C, - , F</td>
</tr>
<tr>
<td>H</td>
<td>5, 7, 6</td>
<td>F, - , E</td>
</tr>
</tbody>
</table>

Triangles always have three nodes and usually have three neighbouring triangles. Triangles on the periphery of the TIN can have one or two neighbours.
Three Dimension Landscape of First Turn on Yangtse River in CHINA

Human Settlements Research Center, Tsinghua University
Example of split and merge rules for parcel objects: (A) split; (B) merge
Example Water Facilities Data Model

- Start with objects and relationships
- Model as object types and relationships
  - Topological network
  - Hierarchical ‘type of’
  - Collection ‘composed of’
- Add related attribute tables
Water Distribution system
Water Distribution System
Visio CASE Tool (UML Representation)
Common Mistakes

- Design in abstract without reference to GIS software core data model
- Don’t budget right amount of time
  - Too much, too little
- Try to be too wide ranging and generic instead of specific and practical
- Design for elegance instead of performance
Conclusions

- Data modeling is an **art** and a **science**
- Can’t really understand it without practical experience
- Mature tools available to help
  - CASE, UML
- Never forget its **GIS** data modeling