Lecture 8 DEEP Representation – relating the building blocks of data models

Learning Objectives
8.1 What is data modeling abstraction in regards to the concept of representation?
8.2 What is the emphasis of each of three data modeling levels of abstraction - conceptual data model, logical data model, and physical data model?
8.3 How is a measurement framework related to a GIS-based logical data model? Why must measurement frameworks and operations work together?
8.4 What raster and/or vector data structuring might we use for coastal GIS?

8.1 What is “data modeling abstraction” in regards to the concept of representation?

SURFACE Representation involves the symbolization strategies chosen to create “map displays”. A large portion of material in Geography 360 is about surface representation – we call it map design. Geography 462 is about DEEPer representation used for analysis.

DEEP Representation involves the combination of dimension, measurement unit, reference system, measurement framework, data structure, and data models. But first let us consider the idea of abstraction of data representation.

Levels of Abstraction of GIS Data Representation
GIS data modeling involves several levels of abstraction, each level relying upon a type of representation. Let us consider two contexts of human-computer interaction…
- Some representations, (e.g., natural language), are understandable largely by humans.
  “Reasonably concrete” expressions for humans, e.g., words like nouns and verbs in natural language together with their meaning, convey ideas about everyday life, but are “reasonably abstract”, i.e., generally lacking in detail as representations for computers.

- Some representations, (e.g., bits and bytes), are understandable largely by computers.
  “Reasonably concrete” expressions for computers, (e.g., representations like 0’s and 1’s as bits to form bytes in computer words), are “reasonably abstract” for humans; few of us speak to each other in 0’s and 1’s like a computer communicating with another computer.

Abstraction is a process of identifying the salient (that is, most significant) characteristics and working with them to develop information.

8.2 What is the emphasis of each of three data modeling levels of abstraction - conceptual data model, logical data model, and physical data model?

Data models are composed of representations for 1) geospatial data construct types, 2) operations, and 3) integrity constraints. Data models are the foundation of GIS software and databases. There are three levels of abstraction, each level having a different representation emphasis. What is articulated at the conceptual level is carried forward to the logical level. What is articulated at the logical level (with conceptual level embedded within) is carried forward to the physical level. Let’s consider the data structure part of the data model (representation):
- Conceptual Data Model – data construct content representations, i.e., what portion of (some) world does data model represent? e.g., a coast or perhaps a shore
• Logical Data Model – data structure processing representations, i.e., how of are data to be organized to perform data operations of points, lines, polygons and/or cells etc.?
• Physical Data Model – data format representations, i.e., how is the data to be stored on disk and what is the primitive result of the operation? e.g., numeric or text?

What is the difference between a data structure and a data model? A complete specification of data model adds “operations” and “rules” to “structure” for keeping database content “robust”.

**Foundation of logical data models**

Geospatial data construct types together with measurement frameworks provide the underpinning for GIS logical data models (see table below). Numbers (x.x.x) for table entries are from table 2.6 in lecture 6 (same as table 2-6 in *Exploring GIS*), and provided as p. 4 of this handout.

**Note:** not an exhaustive categorization of all data construct types or measurement frameworks.

<table>
<thead>
<tr>
<th>Geospatial Data Construct Type</th>
<th>Raster Data Models</th>
<th>Vector Data Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Image</td>
<td>Grid</td>
</tr>
<tr>
<td>Image cell</td>
<td>2.1.1, 2.1.2</td>
<td>2.2.1 - 2.2.6</td>
</tr>
<tr>
<td>Grid cell</td>
<td>2.1.1, 2.1.2</td>
<td></td>
</tr>
<tr>
<td>Point</td>
<td></td>
<td>1.1.1</td>
</tr>
<tr>
<td>Multipoint</td>
<td></td>
<td>1.1.2</td>
</tr>
<tr>
<td>Node</td>
<td>1.2.1, 1.2.2</td>
<td>1.2.1, 1.2.2</td>
</tr>
<tr>
<td>Polyline (path/line/segment)</td>
<td>1.1.1, 1.1.2</td>
<td>1.1.1, 1.1.2</td>
</tr>
<tr>
<td>Link</td>
<td>1.2.1, 1.2.2</td>
<td></td>
</tr>
<tr>
<td>Chain (Arc)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Face</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>Ring</td>
<td>1.1.1</td>
<td>1.1.1</td>
</tr>
<tr>
<td>Polygon</td>
<td>1.2.2</td>
<td>1.2.2, 4.1</td>
</tr>
<tr>
<td>Tic</td>
<td>X - special</td>
<td></td>
</tr>
<tr>
<td>Annotation</td>
<td>X - special</td>
<td>X - special</td>
</tr>
<tr>
<td>Simple junction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>simple edge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complex junction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complex edge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Section</td>
<td>4.x other</td>
<td>4.x other</td>
</tr>
<tr>
<td>Route</td>
<td>4.x other</td>
<td>4.x other</td>
</tr>
<tr>
<td>Region</td>
<td>4.x other</td>
<td>4.x other</td>
</tr>
</tbody>
</table>

**Special Note:** This categorization provides an example for understanding the association between construct types (primitives) and measurement frameworks to form “logical data models”. Missing from data model descriptions above are the list of operations and data integrity
constraints associated with each of the data models. Learning to use GIS data operations on raster and vector data structures (measurement frameworks) is much of the effort in learning about GIS, and how to use it. No one has ever created an exhaustive enumeration of all operations for all data models – it has been too big a job. A couple of dissertations have come close for basic raster and vector structures.

Data Tomlin (U Pennsylvania) for raster – the easier one, and the basis of Spatial Analyst Jochen Albrecht (City University of New York) for vector – the hard one; everything else

8.3 How is a measurement framework related to a GIS-based logical data model? Why must measurement frameworks and operations work together?

Building Blocks of a Data Model:
1) **Dimensions and Measurement Units** - needed to describe particular phenomena about our world to understand the differences and similarities among phenomena. We need to agree about the basis of measurements to come to know the world in an agreed upon way. That is where “standards of measure” are used for knowing how to work with units of measurements.
2) **Measurement levels** - provide a generalization about units of measurement so we can work with the measurement units in systematic ways. That is, it does not make sense to add a nominal measurement to an interval measurement; what would be the result?
3) **Reference system** organizes measurements, i.e., establishes relationships among measurements in regards to dimensionality, e.g. like a coordinate system, or a soils sampling scheme, or in a calendar of dates. A reference system relies upon you understanding dimensions, measurement units, and measurement levels.
4) **Measurement framework** - organizes reference systems and measurements into data constructs so we can develop information about the world and build relationships between and among our measurements in terms of fixed, controlled, measured, randomized or ignored as the basis of operations on a data structure.
5) **Data model** - three components (data constructs, operations, and rules) that organize abstraction levels 1-4 above for software design (for data management, analysis, and/or display) so we can undertake GIS data processing. A **data structure** “represents” data constructs and relationships among data constructs used within measurement framework in a way that software can implement. **Operations** sequence data processing of the data structures deriving information from data. Integrity constraints provide rules for data structures and operations within contexts for meaningful data processing, keeping data understandable. Important note: PLEASE DO NOT make the error of thinking three levels of abstraction are the same as the three components in a data model; we have three components at each level.

8.4 What raster and/or vector data structuring might we use for spatial-temporal coastal GIS?

Consider the data model building blocks mentioned above in the context of 4 scale levels of change influencing **PSNERP process unit shoreforms**. A **substantial analysis** was performed by the PSNERP team for scale levels 1 and 2, but not 3 and 4. This course highlights the importance of scale levels 3 and 4 in the context of important developments for GIS involving **geographic dynamics**. Spatial-temporal data model designs are needed to fully incorporate time into analysis; **some progress is be made**. Nonetheless, spatial-temporal change analysis can be performed with current logical data models as exemplified by the PSNERP activity.
### Table 2.6 Summary of Measurement Frameworks

#### 1. Control by Attribute

**1.1 Isolated Objects**
- **Spatial Object** - Single category distinguishes from void, e.g. islands in the ocean.
- **Isoline** - Regular slices of continuous variable, e.g. elevation is systematically controlled to sample contour lines

**1.2 Connected Objects**
- **Network** - Spatial objects connect to each other, form topology, e.g., a street network or utility network
- **Categorical Coverage** - Network formed by exhaustive classification, e.g., soils across a landscape

#### 2. Control by Space

**2.1 Point-based Control**
- **Center point** - Systematic sampling in regular grid, e.g. for a digital elevation matrix
- **Systematic unaligned** - Random point chosen within cell, e.g., location of a houses within land use grid cells

**2.2 Area-based Control**
- **Extreme value** - Maximum (or minimum) of values in cell, e.g., the highest point or lowest point in a grid cell
- **Total** - Sum of quantities in cell, e.g. the reflected light (as type of land cover) of a cell
- **Predominant type** - Most common category in cell, e.g., predominant land use or land cover in a cell
- **Presence / absence** - Binary result for single category; e.g., it is or is not a type of land cover
- **Percent cover** - Amount of cell covered by single category; e.g. 65% impervious surface of a cell
- **Precedence of types** - Highest ranking category present in cell, e.g., as in land use mostly commercial, some residential, little industrial

#### 3. Control by Relationships

**3.1 Measurement by pair** - Control by pairs of objects, e.g. origins and destinations of immigrants

**3.2 Triangular Irregular Network (TIN)** - Control by uniform slope, e.g., three points define a surface and thus together are used to compute slope.

#### 4. Composite Control

**4.1 Choropleth** Control by categories (name of zones) then by space; e.g. US counties have names, and each named county has a legal boundary for which population is counted.