Lecture 5 Measurement frameworks as a Core of Data Structures – 1

Learning Objectives
5.1 What are we representing in coastal databases?
5.2 What are the two popular meanings of the term ‘data model’?
5.3 What is a measurement framework in relationship to a data model?
5.4 What is the difference between a raster data structure and a vector data structure in terms of fixed, controlled, and measured treatments of data dimensions for space, time, attribute?

5.1 What are we representing in coastal databases?

A database can be called a “representation model” of some (past, present, future) world. See Beatley: Ch. 2 Understanding the Coastal Environment, See Fig 2.1 shore zone on p.14, Fig 2.2 coastal zone, and Table 2.1 processes, plus PSNERP process unit shoreform graphic for comparison of characteristics of a world. We represent objects in vector data models and fields in raster data models to compose geospatial databases. Geospatial data constructs, e.g., points, lines, polygons, cells and pixels are the building blocks for data models, hence databases. We organize the measurements for constructs using measurement frameworks.

Measurement framework is a foundation for a data structure abstraction: Dimensions of space, time & attribute with units of measurement each at a level of measurement within a reference system, wherein dimensions are bundled for characterizing geospatial phenomena.

5.2 What are the two popular meanings of the term ‘data model’?

Measurement frameworks are the core of data structures within data models! A data model (as a way to abstract the character of phenomena) specifies:
- **data structure** (constructs) used for representing a world at some level of meaning (e.g. everyday meaning, organizational meaning, software meaning, database meaning)
- the operations that are possible on those data constructs for deriving information from the data structure; that is, establishing new relationships as insights among data constructs
- a set of integrity constraints (rules) for keeping database content and structure robust – aka clean.

1) **Most popular use of term data model** includes the data construct component only, or what we call basic data structure; data content plus relationships forms the data structure (e.g. feature classes and relationships among feature classes). This interpretation is the foundation of data model, but a partial understanding; not wrong, just incomplete. Common marine data types from Wright 2007 Ch 2 Common Marine Data Types.

2) The more complete understanding of data model considers all three components – constructs, operations, and constraints according to Edgar Codd who elucidate the idea in the early 1980’s and awarded fellow status at IBM for that work.

Measurement frameworks are used to create feature classes or rasters ( grids) as the foundation of data content and structures used in data processing. However, the data content and structure offers the foundation of information only. Operations are critical for deriving information from data structures by deriving relationships. We address the operations in more detail in lecture 7. Let’s take a deeper look at the concept of measurement framework (data structures in lectures 5 and 6) that synthesize information addressed in lectures 1 – 4.
5.3 What is a measurement framework in relation to a data model?

Basic Measurement Framework: The Geographical Matrix – Brian Berry

Geography has a set of "entities" (places) and each "has" certain attributes. Thus a simple matrix (cases and variables) serves the model...

<table>
<thead>
<tr>
<th>City Name</th>
<th>Population 1990</th>
<th>% Office Vacancy</th>
<th>Debt/ Person</th>
<th>Rainy Days / Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>7072000</td>
<td>18.1</td>
<td>$4778</td>
<td>111</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>3485000</td>
<td>14.3</td>
<td>2296</td>
<td>35</td>
</tr>
<tr>
<td>Chicago</td>
<td>2784000</td>
<td>22.1</td>
<td>2160</td>
<td>114</td>
</tr>
<tr>
<td>Houston</td>
<td>1631000</td>
<td>19.1</td>
<td>2430</td>
<td>90</td>
</tr>
</tbody>
</table>

A lingering question about measurement in this matrix: What is a "City"? City ‘has’ attributes that can be measured, e.g. ‘human population’ measured as “count”, e.g. population in 1990. This is the model of social science data: "cases" as places, and it doesn't allow spatial relationships in any “deep” way – location might be considered point, or assumed. GIS emphasizes “location” as grounded in a spatial reference system. Coordinates can be represented as variables, but they do not provide relationships in an explicit manner.

A city has a boundary in space as specified in a ‘city charter’ granted to it by a governing jurisdiction; in the US that jurisdiction would be a state or territory. How we treat the relationship among “space, attribute, and time” in an integrated manner is understood through use of a “measurement framework”. Therefore, a **measurement framework specifies a relationship among a collection of dimensions (from space, time, attribute meta-dimensions) using “treatment rules” to organize the dimensions of data.** Remember that dimensions are organized within reference systems having units of measurement at some level of measurement for each dimension. The levels of measurement help guide/constrain operations, which is why Chrisman added five levels to Stevens four levels (scales) of measurement.

What are the basics of the treatment rules? David Sinton in 1978 articulated how we ‘treat’ dimensions of space, attribute and time (really any measured dimension) within map analysis as a way to form a data category, i.e., feature class or raster in ArcGIS. Remember, that a feature class and raster is a basic data structure.

Three **treatment rules** offered by Sinton.
- Fix – a data value for a dimension when controlling and measuring other dimensions
- Control – in a systematic way a second data dimension by varying the data value(s) within some known range when measuring another data dimension
- Measure – observations with a unit of measurement for a third data dimension using some reliable sampling mechanism

Example of human population data categories (i.e., feature classes as simple data structures) for dot map and choropleth map:
- population data category for a dot map: time is fixed, attribute is controlled (selected for observation), space is measured (where the population ‘dots’ occur)
- population data category for a choropleth map: time is fixed, space is controlled, attribute is measured

Time is commonly fixed for maps; although time is often seen as a second control within space-time data analysis.

In the book *Validity and the Research Process* (Sage 1985), David Brinberg and Joe McGrath enumerated six modes of treatment (independent from Sinton). In addition to three rules above offered by Sinton, Brinberg and McGrath also included:

- Systematic match – establish measurement similarity for additional comparisons
- Random – let the variable take on any data value because we know it exists, but not sure of actual measurements.
- Ignore – recognize, but do not specifically treat the data values; thus we make assumptions that the data are not important enough to treat.

**Putting a Measurement Framework to Work as Data Structures**

5.4 What is the difference between a raster data structure and a vector data structure in terms of fixed, controlled, and measured treatments of data dimensions for space, time, attribute?

**Remember:** A (geo)spatial data structure is composed of the spatial data constructs and the relationships among these constructs; and a measurement framework is the foundation of a data structure. Two main families of data structures are raster and vector, which are foundational for understanding most GIS data structures.

<table>
<thead>
<tr>
<th>FIX</th>
<th>CONTROL</th>
<th>MEASURE</th>
<th>Sinton's interpretation (1978) of data structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Space</td>
<td>Attribute</td>
<td>RASTER (Location controlled by grid as point or cell area)</td>
</tr>
<tr>
<td>Time</td>
<td>Attribute</td>
<td>Space</td>
<td>VECTOR (a generic type as point, line, polygon)</td>
</tr>
</tbody>
</table>

Both raster and vector approaches provide a framework using a single value at all points in a region. From a geographic perspective, a continuous sampling is called a 'surface', and from a mathematical perspective it is called a 'field', even if the data value level of measurement is nominal. Objects (vector) and fields (raster) are terms that used more recently to differentiate elements of vector and raster data structures.
Example data structures interpreted using measurement frameworks:

**Space controlled example** - grid of world population
- time is fixed in terms of calendar year(s)
- space as grid location is controlled (for representing total count within area)
- attribute human population is measured (estimated) from census sources

**Temporal controlled example** - Coastal erosion at Washaway Beach WA
- attribute is fixed as “shoreline” interface between land and water
- time is controlled – observations taken at certain “year” controlled observations
- spatial (feature) location is measured – location of the (shore)line

**Space and Time controlled example** - World Climate stations index map
Average precipitation and temperature by month at climate stations. Thus, space and time controlled, temperature and precipitation measured.

**Generalizing across Measurement Framework Families**

*based on relationships between location and attribute
Chrisman's revision/reinterpretation of Sinton

**Space-Controlled (Raster / Grid / Field)**

*basic rules: spatial location serves as control, attributes measured to suit the topic*
- Point-based: center point
- Area-based: (many rules addressed in next lecture)

**Attribute-Controlled (Vector / Feature / Object Frameworks)**

*basic rules: attribute serves as control, locations measured to suit the topic*
- Isolated Objects: each category taken as Yes/No
  - 'Spatial Object' (or cartographic feature view): each entity (point,line,area) surrounded by the void as in the Brian Berry geographical matrix shown earlier.
  - Isoline: continuous attribute sliced, disjoint (contour lines do not intersect)
- Connected Objects: multinomial (two or more) categories
  - Network: object connectivity relationships (e.g. highway network; stream network)
  - Categorical Coverage: exhaustive classification divides space with contiguous boundaries. Example: legislative districts within a state for a given year

**Relationship Control**
- Control by Pairs (matrix as origin-destination (O-D) of flow; migration matrix as to-from flow)
- Triangular Irregular Network (TIN) – nodes, links, triangles in relation to one another

**Composite Frameworks**
- Choropleth (control by categories [names of zones] then by space [irregular collection zone])