How to Design a GeoDB

GEOG 419/519: Advanced GIS
Key is asking the right questions:

- How can GIS technology be implemented to streamline existing functions or change the way we achieve a goal?
- What data will benefit the organization the most?
- What data can be stored?
- Who is responsible for maintaining the database?
Truths about Design

• Time-consuming
• No end-use applications
• If not done:
  – Database may not meet requirements
  – Can end up with duplicate, missing, or unnecessary data
  – Lack of necessary management techniques
Objectives of Design

- Design should define goals, identify, analyze, and evaluate design alternatives, and create implementation plan
- Investment of time and money up front saves even more time and money later
- A functional, well-organized database:
  - Satisfies organizational objectives
  - Contains all necessary data
  - Accommodates different views of the data
  - Distinguishes maintenance apps from user apps
  - Organizes data so that different users access the same data
Design Guidelines

- Involve users
- Take it one step at a time
- Build a team
- Be creative
- Create deliverables
- Keep organizational goals and objectives in focus
- Do not add detail prematurely
- Document carefully
- Be flexible
- Plan from you model
- **Create from scratch**: create schema for features datasets, classes, and attribute tables.
- **Import Existing Data**: a schema is created, and existing data in imported in.
- **Use CASE tools**: computer aided software engineering principles and GUIs can be used to create a geodatabase.
- **THIS IS HOW TO BUILD A COMPUTER DATABASE, BUT DOESN’T MEAN ITS OF ANY USE** – for this you need good database design.
Steps in Building a GeoDB

- Model the user’s view of data
- Define objects and relationships
- Select geographic representation
- Match to geodatabase elements
- Organize geodatabase structure
Designing GIS Databases

- Model the users’ view
- Define entities and their relationships
- Identify representation of entities
- Match to GIS data model
- Organize into geographic data sets

The first three steps develop the conceptual model, classifying features based on an understanding of the data required to support the organization’s functions, and deciding their spatial representation.

The last two steps develop the logical model, matching the conceptual models to ArcGIS geographic data sets.
The Data Model

- **Data Model** is a formal definition of the data required in a GIS. Types include:
  - Structured List
  - Entity Relationship Diagram

- Purpose of the data model is to ensure that the data is identified and described in a *completely rigorous and unambiguous fashion*
### Example Structured List

<table>
<thead>
<tr>
<th>Feature</th>
<th>Layer</th>
<th>Type</th>
<th>Prim. Attr.</th>
<th>Owner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landuse</td>
<td>LU</td>
<td>Poly</td>
<td>LUI D</td>
<td>Planning</td>
</tr>
<tr>
<td>Soils</td>
<td>Soil</td>
<td>Poly</td>
<td>Soil ID</td>
<td>na</td>
</tr>
<tr>
<td>Elevation</td>
<td>DEM</td>
<td>Raster</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>Hydrography</td>
<td>Hydro</td>
<td>Line</td>
<td>ID</td>
<td>DEC</td>
</tr>
<tr>
<td>Roads</td>
<td>CL</td>
<td>Line</td>
<td>ID</td>
<td>Eng.</td>
</tr>
<tr>
<td>Buildings</td>
<td>Bldg</td>
<td>Poly</td>
<td>ID</td>
<td>Eng.</td>
</tr>
<tr>
<td>Parcels</td>
<td>Parcel</td>
<td>Poly</td>
<td>SBL</td>
<td>Assessor</td>
</tr>
</tbody>
</table>
Model the User’s View

• Identify the functions that support the organization’s goals and objectives
• Identify the data required to support the functions
• Organize the data into logical sets of features
• Define an initial implementation plan
• Identify organizational functions
Identify organizational functions

The geodatabase design will be influenced by the structure of your organization. Distinct departments may have responsibility for different segments of the geographic data.

At a basic level, you begin by identifying the providers and consumers of geographic information. The key data flows are modeled. This is the starting point for identifying logical groupings of data.

Determine data needed to support functions

<table>
<thead>
<tr>
<th>Land records</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Types of data</td>
<td>Data source</td>
</tr>
<tr>
<td>Parcel</td>
<td>Subdivision plans</td>
</tr>
<tr>
<td>Easement</td>
<td>Engineering records</td>
</tr>
<tr>
<td>Parcel description</td>
<td>Land title</td>
</tr>
<tr>
<td>Parcel photograph</td>
<td>Historic archive</td>
</tr>
<tr>
<td>Owner</td>
<td>Land assessment</td>
</tr>
<tr>
<td>Address</td>
<td>Phone database</td>
</tr>
</tbody>
</table>

For each function, identify all of the types of data that are necessary to fulfill this group's requirement to deliver information.

Organize data into logical groupings

Water utility
Land records
Streets
Terrain

From an inventory of all the types of geographic data that an organization maintains, identify a modest set of groupings that comprise all of your geographic data systems.

For each data type, identify the likely source of data. A part of the project plan must include an estimate for cost of data capture, processing, and validation.
Define Entities and Relationships

• Identify and describe entities
• Identify and describe the relationships among these entities
• Document the entities and relationship with UML diagrams
• Create statements about how the system works then evaluate those statements in terms of entities and relationships
Define objects and relationships

Identify and describe objects

- Form sentences that state the entities and their behavior. The nouns are entities and the verbs are relationships.
- This step can be done by writing a progressive series of statements starting with "a" water system is composed of devices and water lines." Each statement should be simple and accurate.

```
A valve controls the flow of water.
A water device connects to one or more water mains.
A water system is composed of devices and water lines.
A water main is a type of water line.
```

Specify relationships between objects

- Many entities have close relationships with other entities. Relationships guide your geodatabase design.

```
The land title lists me as owner.
```

Document model in diagram

- Once you have collected your list of entities and relationships, it is a good practice to create a data model diagram.

```
Using business graphics software, start by making boxes for entities and lines with arrows for relationships. This diagram will facilitate discussion with domain experts and advance the refinement of the model.
```
The Nature of Geographic Data

- Geographic data has been described as:
  - **Object** - a thing that can be seen or touched.
  - **Entity** - objects or things to be included in a database
  - **Feature** - the make, shape, form or appearance of a person or thing. Term that derives itself from cartography (*features on a map*)
  - **Attribute** - characteristics of the entities
Identify the representation of entities

- Is the feature represented on a map?
- Is the shape of a feature important?
- Is the feature best accessed through its relationship with another feature?
- Will the feature have different representations at different scales?
Select geographic representation

Represent discrete features with points, lines, areas

You can model the richest expression of features with the vector types. These entities are well defined on a map and are permanent.

- **Point**: an entity too small to map with a line or area
- **Line**: a long entity too narrow to map with an area
- **Area**: an entity with length and width at the map scale
- **Annotation**: a descriptive label on an entity
- **Object**: a nongeographic entity, such as an owner

Characterize continuous phenomena with images

Images have versatile application in a GIS. You would specify images for aerial or satellite photographs, photographs of facilities, and any scanned documents.

- **Image**: a file that contains a continuous valued map, aerial photograph, copy of a plat, or picture of a building

Model terrain with surfaces

When you model a continuous phenomenon that has a z value, specify surface. (Later, you will decide whether TIN or raster is better for the surface.)

- **Surface**: a system of points or locations with elevation values that form a mesh for a mathematical approximation of the shape of the earth
Classical Entities and Spatial Component

Design your geodatabase
- What data?
- Location and projection?
- Feature classes and subtypes?
- Geometric networks?
- Rules?
- Relationships?

Entity

Entity Attributes
- Street (name, number)
- Soil Zone (name, perm.)
- Well (type, date tested)

Spatial Component
- line (coordinates, topology)
- polygon (coordinates, topology)
- point (coordinates)
E-R Diagram

- Platform
- Side Number
- Part of
- Station
- Name Zone

© Arthur J. Lembo, Jr.
Salisbury University
# Spatial Relationships

<table>
<thead>
<tr>
<th>Spatial Relationship</th>
<th>Verbs</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connectivity</td>
<td>Connect, link</td>
<td>![Symbol for Connectivity]</td>
</tr>
<tr>
<td>Contiguity</td>
<td>Adjacent</td>
<td>![Symbol for Contiguity]</td>
</tr>
<tr>
<td>Containment</td>
<td>Contained</td>
<td>![Symbol for Containment]</td>
</tr>
<tr>
<td>Proximity</td>
<td>Nearest</td>
<td>![Symbol for Proximity]</td>
</tr>
<tr>
<td>Coincidence</td>
<td>Coincident</td>
<td>![Symbol for Coincidence]</td>
</tr>
</tbody>
</table>
Considerations in Modeling Geography in an E-R Diagram

- Correct Identification and Definition of Entities
- Defining a Corresponding “Spatial” Entity for Each “Traditional” Entity
- Recognition of Multiple Instances of Geographic Entities
  - time and scale
  - Represented by: entity simple, entity spatial, entity time
Representation of Spatial Objects

Object (entity)

Spatial Object | G | T

Associated Spatial Object Type

Object Name

Topology: point, line, polygon

XY Coordinate system
Modeling Spatial Relationships

ENTITY

RELATIONSHIP

ATTRIBUTE
Developing a Spatial E-R Diagram

- Is Derived From Needs Assessment
- Relationships Determined from Application Descriptions

Soil Sample
- Point
- Contained In
- /Contains

Farm Plot
- Polygon

© Arthur J. Lembo, Jr.
Salisbury University
Match to a GeoDB model

• Determine appropriate geodb representation for entities
  – Spatial type = point
    • Unconnected – point feature
    • Connected – simple junction
    • Connected with internal topology – complex junction
  – Spatial type = line
    • Stand-alone line = line feature
    • Line in a system = simple edge
    • Line with connected sections = complex edge
  – Areas = polygon feature (with potential planar topology)
  – Objects = objects
## Match to geodatabase elements

### Apply feature geometry and topology

<table>
<thead>
<tr>
<th>Entity</th>
<th>Related to</th>
<th>Spatial Type</th>
<th>ArcInfo Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water utility</td>
<td>-</td>
<td>point</td>
<td>object</td>
</tr>
<tr>
<td>Pump</td>
<td>-</td>
<td>point</td>
<td>point feature</td>
</tr>
<tr>
<td>Motor</td>
<td>-</td>
<td>point</td>
<td>point feature</td>
</tr>
<tr>
<td>Meter box</td>
<td>Motor</td>
<td>point</td>
<td>point feature</td>
</tr>
<tr>
<td>Valve</td>
<td>-</td>
<td>point</td>
<td>simple junction</td>
</tr>
<tr>
<td>Water main</td>
<td>-</td>
<td>line</td>
<td>complex edge</td>
</tr>
<tr>
<td>Treatment plant</td>
<td>-</td>
<td>point</td>
<td>complex junction</td>
</tr>
<tr>
<td>Land records</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parcel</td>
<td>-</td>
<td>area</td>
<td>polygon feature</td>
</tr>
<tr>
<td>Easement</td>
<td>-</td>
<td>line</td>
<td>line feature</td>
</tr>
<tr>
<td>Parcel description</td>
<td>Parcel</td>
<td>text</td>
<td>annotation feature</td>
</tr>
<tr>
<td>Parcel photograph</td>
<td>-</td>
<td>image</td>
<td>raster</td>
</tr>
<tr>
<td>Owner</td>
<td>Parcel</td>
<td>object</td>
<td>object</td>
</tr>
<tr>
<td>Address</td>
<td>-</td>
<td>location</td>
<td>address</td>
</tr>
<tr>
<td>Streets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Street</td>
<td>-</td>
<td>line</td>
<td>line feature</td>
</tr>
<tr>
<td>Bridge</td>
<td>-</td>
<td>point</td>
<td>point feature</td>
</tr>
<tr>
<td>Name</td>
<td>Street</td>
<td>text</td>
<td>annotation feature</td>
</tr>
<tr>
<td>Traffic light</td>
<td>-</td>
<td>point</td>
<td>point feature</td>
</tr>
<tr>
<td>Bus route</td>
<td>-</td>
<td>line</td>
<td>line feature</td>
</tr>
<tr>
<td>Bus stop</td>
<td>-</td>
<td>point</td>
<td>point feature</td>
</tr>
<tr>
<td>Environment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Historic monument</td>
<td>-</td>
<td>point</td>
<td>point feature</td>
</tr>
<tr>
<td>Fence</td>
<td>-</td>
<td>line</td>
<td>line feature</td>
</tr>
<tr>
<td>Vegetation cover</td>
<td>-</td>
<td>area</td>
<td>polygon feature</td>
</tr>
<tr>
<td>Place names</td>
<td>-</td>
<td>text</td>
<td>annotation feature</td>
</tr>
<tr>
<td>River valley</td>
<td>-</td>
<td>surface</td>
<td>TIN</td>
</tr>
<tr>
<td>Satellite image</td>
<td>-</td>
<td>image</td>
<td>raster</td>
</tr>
</tbody>
</table>

### Determine feature and geometry type

- For non-geographic objects, select row.
- For simple geographic objects, select feature.
- For features in a network, select simple or complex edge or junction feature.

### Specify topological graphs

- For linear systems, such as transportation or utility, select geometric network.
  A geometric network has custom behavior built in to make the editing of networks easy.
- For systems of land or jurisdictions, a planar topology manages the shared geometry of a set of features.
  A planar topology enforced that no feature can cross another without an intersection.

### Implement attribute types for objects

Each entity can have many attributes. These are the attribute types.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multipolygon</td>
<td>polygon</td>
</tr>
<tr>
<td>Long integer</td>
<td>long</td>
</tr>
<tr>
<td>Real</td>
<td>real</td>
</tr>
<tr>
<td>Date</td>
<td>date</td>
</tr>
<tr>
<td>Integer</td>
<td>int</td>
</tr>
<tr>
<td>String</td>
<td>string</td>
</tr>
<tr>
<td>Text</td>
<td>text</td>
</tr>
<tr>
<td>Geometry</td>
<td>geometry</td>
</tr>
<tr>
<td>Boolean</td>
<td>bool</td>
</tr>
<tr>
<td>SpatialReference</td>
<td>spatial</td>
</tr>
</tbody>
</table>

---

**Salisbury University**

Chapter 12 • Geodatabase design guide • 193
Physical Design

• Logical Design Performed Independent of Physical Design
• Must Now Move Logical Design to Physical Design
• Physical Design Example (ARC/INFO, geodatabase):
Conceptual vs. Physical Design

Temp Logger
- Point: G T

Stream Segment
- Line: G T

Physical Design

Stream Reach Order
- Stream INFO
- PAT BND AAT TIC NAT

History
- ID Date Temp

TempLog
- ID

© Arthur J. Lembo, Jr.
Salisbury University
Organize into Geographic Datasets

• Assign entities to feature classes and subtypes
  – Feature classes vs. subtypes

• Group related sets of features into geometric networks or planar topologies
  – Simple edges & junctions, complex edges & junctions – geometric network
  – Need space-filling and no overlapping – planar topology

• Organize feature classes and datasets into geodb’s