EEOS 381 - Spatial Databases and GIS Applications

Lecture 5
Geodatabases
What is a Geodatabase?

- “Geographic Database” – ESRI-coined term
- A standard RDBMS that stores and manages geographic data
- A modern object-relational GIS data model or format of GIS data based on relational table principles
- A manager of geographic data
- A concept of an organization’s GIS database
What can the Geodatabase do?

- Provide (controlled) access (read/write) to large volumes of geographic data stored in both files and databases.
- Handle rich and diverse data types as well as other objects.
What can the Geodatabase do?

- Deploy business logic and GIS functionality to desktops, servers (including the Web), and mobile devices
- Apply sophisticated rules and relationships in an “intelligent” GIS—all stored in the database. The rules closely model reality.

The geodatabase architecture is based on simple relational storage and combined with advanced application logic.

Geodatabases are used to manage and store diverse collections of geographic information types.
What can the Geodatabase do?

From a conceptual standpoint, the geodatabase implements advanced GIS logic in an application tier on top of numerous RDBMSs, although the logic is stored in the geodatabase.

The same application logic animates GIS behavior on file-based data sets, but must be managed outside the database.
The geodatabase is a common framework shared by all ArcGIS products and applications.

Data stored in an ArcSDE geodatabase may be accessed in non-ArcGIS and non-ESRI applications as well.
Types of Geodatabases

- **Personal** - Microsoft Jet Engine (Access database - .mdb file) (Debuted at 8.0)
- **File** - Folder on disk named with .gdb extension
- **Desktop, Workgroup SDE** - use SQL Server Express 2005/2008
- **Enterprise SDE** - in RDBMS (Oracle, SQL Server, IBM DB2, IBM Informix, PostgreSQL/PostGIS)
# Comparison of GDB Types

<table>
<thead>
<tr>
<th></th>
<th>Personal Geodatabase for Microsoft Access</th>
<th>File Geodatabase</th>
<th>Desktop/Workgroup/Enterprise Geodatabase</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Storage Technology</strong></td>
<td>MS Access</td>
<td>Folder of binary files</td>
<td>RDBMS</td>
</tr>
<tr>
<td><strong>Storage capacity</strong></td>
<td>2 GB</td>
<td>No Limits (256 TB per table)</td>
<td>Depends on server</td>
</tr>
<tr>
<td><strong>Licensing</strong></td>
<td>Basic, Standard, Advanced</td>
<td>Basic, Standard, Advanced</td>
<td>ArcGIS Desktop/ArcGIS Server</td>
</tr>
<tr>
<td><strong>Number of users</strong></td>
<td>Single editor, Multiple readers</td>
<td>One editor per table, Multiple readers</td>
<td>Multiple editors &amp; readers</td>
</tr>
<tr>
<td><strong>Versioning support</strong></td>
<td>None (Supports check-in / checkout replication only)</td>
<td>None (Supports check-in / checkout replication only)</td>
<td>Versioning, replication, archiving</td>
</tr>
</tbody>
</table>

* Before 10.1 known as “ArcView”, “ArcEditor” and “ArcInfo”

>> **ESRI:** “File GDBs have better performance than Access PGDB”
Benefits of the Geodatabase

- Non-tiled (seamless) layers
- Rapid spatial searching (indexing)
- Storage and integration of vector, raster, tabular data, metadata and other objects
- Topology and other custom-based behavior
Storage

- Stores each feature (coordinates and attributes) as a row in a table, with special “spatial” columns – e.g. SHAPE, area, length.
  - Feature classes actually comprise several related tables.
- All objects are stored in tables
- GDB metadata are stored in yet another set of tables
Overview of GDB Elements

Feature Dataset
Contains spatially-related feature classes together with the topology and network objects that bind them. Feature classes in a feature dataset have spatial reference.

Feature class
A table with a shape field containing point, line, or polygon geometries for geographic features. Each row is a feature.

Table
A collection of rows, each containing the same fields. Feature classes are tables with shape fields.

Domain
Defines a set or range of valid values for a field.

Relationship class
Associates objects from a feature class or table to objects in another feature class or table. Relationship classes can optionally have user-defined fields.

Topology
Integrity rules that define the behavior of geographically-integrated features.

Schematic Dataset
A dataset used for graphically representing network connectivity and sets of relationships.

Terrain
Triangulated irregular network (TIN)-based dataset

Geometric Network
Rules for managing connectivity among features in a set of feature classes.

Parcel Fabric
A seamless parcel boundary network created and maintained using data from survey records.

Raster Dataset / Raster Catalog / Mosaic Dataset
Contains rasters which represent continuous geographic phenomena.

Metadata Document
An XML document that can be associated with every dataset, commonly used in ArcIMS and other server applications.

Geoprocessing Tools
A collection of dataflow and workflow processes for performing data management, analysis, and modeling.

Network Dataset
A collection of topologically connected network elements (e.g., edges, junctions, and turns) that are derived from network sources, typically used to represent an undirected-flow system network such as a road or subway system.

Address Locator

Annotation

Dimension
GIS Data Types in a GDB
Feature Classes

Simple:
- Point/Multipoint, Line or Polygon (single or multipart), Multipatch
- Annotation
  - Feature-linked
  - Simple
  - Dimension

Topological (geometric network)

Names should be descriptive, must begin with a letter and can contain only letters, numbers and underscores, and must be unique within a geodatabase
Feature Datasets

Collection of related feature classes that share a common coordinate system/spatial reference.
Use for:
- Topology, based on rules (see slides 17-19)
- Terrain dataset
- Network dataset
- Geometric network
- Parcel fabric

Geodatabase Topology

- Refers to spatial relationships between geographic features (adjacency, connectivity, coincidence, etc)
- Helps ensure data quality and integrity
- Users create, define, validate, and maintain topology in a geodatabase, based on many spatial constraints ("topology rules") – see PDF poster on S: drive and online.
  - The rules are stored in the geodatabase.
  - Requires Standard (ArcEditor) or Advanced (ArcInfo) license to create
Upon creation, users define which layers participate in a topology.

- One or multiple polygon, point, and line layers can participate in the same topology.

Users specify which rules are appropriate for their data layer(s).

Topology may be validated to find errors, and errors may be fixed or flagged as exceptions in ArcMap.
## Geodatabase Topology

### Types of Topology

<table>
<thead>
<tr>
<th>Type</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>arc-node topology</strong></td>
<td>Line features can share endpoints</td>
</tr>
<tr>
<td><strong>region topology</strong></td>
<td>Area features can overlap with other area features</td>
</tr>
<tr>
<td><strong>polygon topology</strong></td>
<td>Area features can share boundaries</td>
</tr>
<tr>
<td><strong>node topology</strong></td>
<td>Line features can share endpoint vertices with point features</td>
</tr>
<tr>
<td><strong>route topology</strong></td>
<td>Line features can share segments with other line features</td>
</tr>
<tr>
<td><strong>point events</strong></td>
<td>Point features can share vertices with line features</td>
</tr>
</tbody>
</table>

Geodatabase Topology

Examples of rules – one feature class:

### Must not overlap

Polygons must not overlap within a feature class or subtype. Polygons can be disconnected or touch at a point or touch along an edge.

- Use this rule to make sure that no polygon overlaps another polygon in the same feature class or subtype.

- A voting district map cannot have any overlaps in its coverage.

### Must not have gaps

Polygons must not have a void between them within a feature class or subtype.

- Use this rule when all of your polygons should form a continuous surface with no voids or gaps.

- Soil polygons cannot include gaps or form voids—they must form a continuous fabric.
Geodatabase Topology

Examples of rules – multiple feature classes:

**Line**
- Must be covered by boundary of
- Lines in one feature class or subtype must be covered by the boundaries of polygons in another feature class or subtype.
- Use this rule when you want to model lines that are coincident with the boundaries of polygons.

**Point**
- Must be properly inside polygons
- Points in one feature class or subtype must be inside polygons of another feature class or subtype.
- Use this rule when you want points to be completely within the boundaries of polygons.
**Other Common GDB Objects**

- **Tables**
  - not features on a map
  - can be associated with features via Joins and Relates

- **Raster Datasets**
  - Single or multi-band images
  - Mosaics
  - In PGDB stored in separate Imagine format image files

- **Raster Catalogs, Mosaic Datasets**

- **Terrains (TIN datasets)**
**Relationship Classes**

- Collection of relationships between feature classes and/or tables
- Model dependencies between objects, built into the GDB rather than by the user on-the-fly in ArcMap
- With relationships, you can control what happens to an object when its related object is removed or changed (ensures referential integrity).
Attribute Domains

Constraints on attributes. Two types:

- **Range** domains
  - For numeric fields
  - Constrains the value to **between** prescribed **minimum** and **maximum** values (like between 2,000 and 14,000).

- **Coded value** domains
  - For all attribute types (except object IDs and BLOBs)
  - More readable values instead of codes
  - Defined set of valid values, to ensure that the attribute has one of the expected values (like “residential”, “commercial”, etc.)

* Domains are set at the **top-level** of a geodatabase

* When editing in ArcMap, upon validating your work, invalid attribute values are highlighted for editing.
Create and manage domains inside of ArcCatalog and/or with the tools in the ‘Domains’ tools in ArcToolbox.

Handy tools to export and import domain values to/from tables.
Default Values

- The most commonly expected value for a new feature.
- Set in the *field properties of a layer or table* in ArcCatalog
- Applied during data entry in ArcMap
- Ex: applying “residential” as the default landuse classification of a new or split land parcel.
- Can be applied to text, short-integer, long-integer, float, double, and date attribute types.
Split and Merge Policies

- When **splitting features**
  - default value applied to both, or
  - duplicate value to both, or
  - geometry ratio (proportional values)

- When **merging features**
  - default applied to merged feature, or
  - sum value - add values, or
  - weighted average

>> Both are set for each attribute **Domain**
# Split and Merge Policies

## Splitting features

<table>
<thead>
<tr>
<th>owner</th>
<th>land zoning</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>G. Gould</td>
<td>R-4</td>
<td>25000</td>
</tr>
</tbody>
</table>

**Split policy**
- Attribute of original feature is duplicated in split features.
- Default value is applied to split features.
- Numeric attribute is subdivided by ratio of split area or length.

## Merging features

<table>
<thead>
<tr>
<th>district</th>
<th>yield</th>
<th>% harvested</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lakeview</td>
<td>24000</td>
<td>35</td>
</tr>
<tr>
<td>Riverside</td>
<td>45000</td>
<td>47</td>
</tr>
</tbody>
</table>

**Merge policy**
- Default value for attribute is applied to the merged feature.
- Numeric attribute is summed.
- Numeric value is weighted average of attribute from the two features.

---

**Database Properties**

<table>
<thead>
<tr>
<th>Field Type</th>
<th>Split policy</th>
<th>Merge policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short Integer</td>
<td>Duplicate</td>
<td>Default Value</td>
</tr>
<tr>
<td>Coded Values</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Set policies here
Domains with Split and Merge Policies

Steps to setting attribute domains

1. select an attribute
2. set a domain type
   - domain type: range
     - minimum value
     - maximum value
   - domain type: coded value
3. set split and merge policies
   - split policy
     - duplicate
     - default value
     - geometric ratio
   - merge policy
     - default value
     - sum values
     - weighted average

Text can only have a coded value domain. All other attribute types can have a coded value or range domain.
Subtypes

Optional way to group related features in one feature class (e.g. road classes), based on the most significant classification of features (or objects in an object class).

Assign distinct simple behavior for different classifications of your objects or features. All subtypes of a class share the same set of attributes.

Can define distinct attribute domains and default values and relationship classes and apply validation rules.

Help preserve integrity of your data.

Set in Properties of a feature class.
# GDB Behavior

## Features sorted by subtype

### Roads with subtypes

<table>
<thead>
<tr>
<th>fid</th>
<th>geom</th>
<th>subtype</th>
<th>width</th>
<th>ln</th>
<th>name</th>
</tr>
</thead>
<tbody>
<tr>
<td>102</td>
<td>concrete</td>
<td>65</td>
<td>4</td>
<td>US Highway 285</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>concrete</td>
<td>75</td>
<td>4</td>
<td>NM Highway 14</td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>concrete</td>
<td>75</td>
<td>4</td>
<td>US Interstate 25</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>asphalt</td>
<td>45</td>
<td>2</td>
<td>Grant Paige Ave</td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>asphalt</td>
<td>35</td>
<td>2</td>
<td>Shakedown Street</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>asphalt</td>
<td>40</td>
<td>2</td>
<td>Acequia Wier</td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>asphalt</td>
<td>45</td>
<td>2</td>
<td>Hart Alley</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>gravel</td>
<td>25</td>
<td>2</td>
<td>Garcia Road</td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>gravel</td>
<td>15</td>
<td>1</td>
<td>Lesh Ranch Road</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>gravel</td>
<td>20</td>
<td>1</td>
<td>McKernan Lane</td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>gravel</td>
<td>15</td>
<td>1</td>
<td>Kreutzman Road</td>
<td></td>
</tr>
</tbody>
</table>

### Realizing simple behaviors

- **Default values**: A new concrete road is given a default value of four lanes. A new asphalt road is given a default width of 35 feet. A new gravel road is given a default width of 15 feet.
- **Attribute domains**: Valid widths are 30, 35, 40, and 45. Valid lane counts are 1, 2, and 4.
- **Split/Merge policy**: A split highway retains all highway designations. Placed “Highway” and “Interstate.”
- **Connectivity rules**: A split gravel road retains its width. A gravel road cannot directly connect to a freeway.
- **Relationship rules**: A merged asphalt road takes a default value for lanes. An asphalt road can be related with bridges or tunnel crossovers.

### Validation rules

- A concrete road can connect to an asphalt road but not to a gravel road. Two concrete roadways can be associated with a highway route.

Each subtype in an object class or feature class has a default value for new attributes, domains of valid attributes, rules to validate how features connect or relate, and the type of relationship possible for a new object.
**Validation Rules**

Control feature and attribute integrity. Three types:

- **Attribute Rule** - attribute domain applied to a subtype of a class
  - Ex: the field named DIAMETER can represent only pipes that are 10, 15, 25, or 50 centimeters in diameter

- **Connectivity Rule** - specifies the valid pairs of attribute values for subtypes for connected network features.
  - Ex: an electric line with phase ABC may be connected to a downstream line with phase AC.

- **Relationship Rule** - constrains the cardinality of a relationship between an origin class and destination class.
  - Ex: a parcel of land can have no, one, or two owners
Aliases

Alternative, more user-friendly descriptive name for a field or feature class in a GDB

- Unlike true names, aliases do not have to adhere to the limitations of the database, so they can contain special characters such as spaces.
- For example, you might specify an alias of "Date of Deed" for a field stored in the geodatabase as DEED_DATE.
Aliases

This name will appear in ArcMap table of contents when the feature class is added to the map.

Feature class

Field

This name will appear in the table and in Identify Results

Click any field to see its properties.

Field Properties

<table>
<thead>
<tr>
<th>Name</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE_30_39</td>
<td>Long Integer</td>
</tr>
<tr>
<td>AGE_40_49</td>
<td>Long Integer</td>
</tr>
<tr>
<td>AGE_50_64</td>
<td>Long Integer</td>
</tr>
<tr>
<td>AGE_65_UP</td>
<td>Long Integer</td>
</tr>
<tr>
<td>MED_AGE</td>
<td>Double</td>
</tr>
<tr>
<td>MED_AGE_M</td>
<td>Double</td>
</tr>
<tr>
<td>MED_AGE_F</td>
<td>Double</td>
</tr>
<tr>
<td>HOUSEHOLDS</td>
<td>Long Integer</td>
</tr>
<tr>
<td>AVE_HH_SZ</td>
<td>Double</td>
</tr>
<tr>
<td>HSEHLD_1_M</td>
<td>Long Integer</td>
</tr>
<tr>
<td>HSEHLD_1_F</td>
<td>Long Integer</td>
</tr>
<tr>
<td>MARRH_CHD</td>
<td>Long Integer</td>
</tr>
<tr>
<td>MARRH_NO_C</td>
<td>Long Integer</td>
</tr>
</tbody>
</table>

Alias: Average Household Size

Allow NULL values: Yes

Default Value: None
Creating Geodatabases

To create a Personal or File GDB:

- Right-click in a folder connection and choose **New** and then select the desired GDB type

Also use the Toolbox tools; With these you can create GDBs compatible with previous versions of ArcGIS.
Creating Objects in a GDB

Right-click a GDB and choose **New**, then select the desired object.
Importing Data into a GDB

Right-click a GDB and choose Import

Then select the desired format

Enter required and optional information into dialog

You can also Export data in other locations and choose your GDB as the Output Location.
Loading Data into Existing Layers

Right-click layer in ArcCatalog and choose Load > Load Data...

- Select source(s)
- Map fields
- Define query

Another way is to load XML Recordset
Exporting Data from a GDB

Right-click GDB and choose Export

Then select the desired format

Enter required and optional information into dialog

This image shows export of all objects in a file GDB
Exporting Data from a GDB

- Right-click feature class and choose Export
- Then select the desired format
- Enter required and optional information into dialog

This image shows export of a specific feature class in a file GDB
Geodatabase XML

- **Extensible Markup Language**
- An open interchange format for exchanging information between geodatabases and other external systems; ASCII text format
- Exchange complete data sets losslessly
- Exchange data *changes*--*Delta record sets*--using XML streams to pass updates and changes among geodatabases and other external data structures
- Exchange and share full or partial GDB *schemas* between ArcGIS users
Geodatabase XML

- **XML Recordset Document** – Features (records) in XML that may be exchanged among existing objects with the same schema

- **XML Workspace Document** – Contains schema, and may contain features as well
Distributing Geodatabases

- **Personal** – just transfer the single .mdb file
  - Compact database first
  - May use WinZip or other utility to compress

- **File**
  - Zip .gdb folder

- **Enterprise SDE**
  - SDE "export file" with command-line tool to produce .sdx files for individual layers

- **XML files**
  - For objects or entire GDB (personal/file or SDE)
    - Zip for more file compression
GDB Resources

ArcGIS Help; ESRI Press Books

http://resources.arcgis.com/en/help/main/10.2/#/What_is_a_geodatabase/003n000000100000/

On S: drive as PDF

ArcGIS online documentation