http://www.state.gov/cms_images/discussion.gif

GEOG 482/582: GIS Data Management

# Lesson 2: Data Models, Database Models, and Data Modeling

# Overview

Learning Objective Questions:

1. Why use three realms for representation modeling?
2. What is a data model?
3. What is a database model?
4. What is data (database) modeling?
5. What are some example application data models?

**Lesson Preview**

Learning objective questions act as the lesson outline.

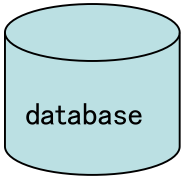
Questions beg answers.

# 1. Why use three realms of representation modeling in GIS?

Three realms of representation provide comprehensive understanding of data and information.

1. 

Worldview – people discuss issues and problems about the world

2. 


Database – people translate that conceptual understanding into means for data structuring and storage

3. 


Map analysis/display – people manipulate the data stored on disk and form a computer display or printed output

Image 1.
http://www.state.gov/cms_images/discussion.gif

Image 3.
http://www.wsdot.wa.gov/NR/rdonlyres/E7A54F0B-1CBB-4875-B93A-DCC4FA2475E9/0/Longview.pdf

**Key terms**
Worldview
Database
Map analysis

# On data, information, evidence and knowledge

Longley et al. 2010, p. 13 present what can be called a *knowledge ramp…*

**Data** – observations on a reality, whatever it might seem to be

**Information** – data placed in a meaningful context for use from which new insight about a topic can be gained.

**Evidence** – corroborated information that strengthens our trust in the information since others agree on the interpretation

**Knowledge** – evidence placed in a framework for use that allows us to take action in a repeated manner

The three realms help us use geospatial data to develop geospatial information that is used to think about and create evidence and knowledge.

**Key terms**
Data
Information
Evidence
Knowledge

# 2. What is a data model?

Everyone has a mental model of a problem, whatever sophistication it might be

- Data models help scaffold our mental models as we use information technology to solve complex problems.

**Key terms**
Data model

A complete data model consists of three components (Codd 1981):

1) geospatial constructs take the form of data structures,
2) operations that can be performed on those structures to derive information from the data, and
3) rules for maintaining the integrity of data.

The first component can be implemented as a data structure.
The second is implemented using software operations.
The third component specifies the constraints on the data structures and/or the software operations.

# 1st component – geospatial data constructs in geodatabase

generic feature classes are defined as:

- point – single point represented by ID
- multipoint – multiple points cluster represented by ID

network junction features:

- simple junction feature – like a node storing topology, but can have logical behavior e.g., valve connecting pipes
- complex junction feature – can contain internal parts, like a transformer or junction box, e.g., switch in electrical network

line:

- line segments – straight line from point to point
- circular arcs – parameterized by radius (pixel subpoints for shape)
- Bezier splines – multiple arcs to fit a series of points

network edge features:

- simple edge feature – lines play a topological role (no interior junctions) can have connectivity rules
- complex edge feature – support one or more junctions along edge

**Key terms**
Data constructs

# Making choices about data structuring

Data types are the foundation of **data model schemas**.

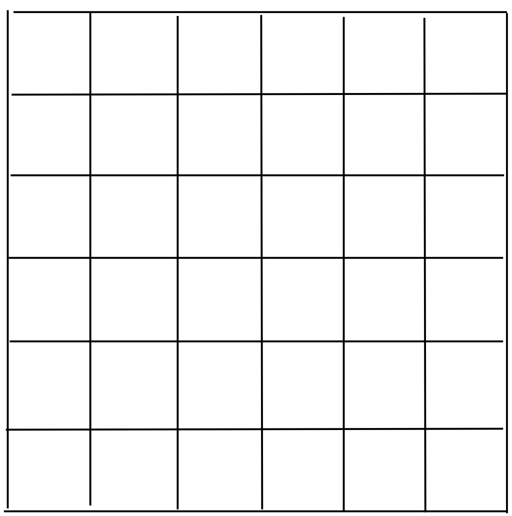
| Geospatial Data Constructs as Data Structure Primitives | Data Model | | | |
|---|---|---|---|---|
| | Raster | Shapefile | Coverage | Geodatabase |
| Image pixel | X | | | X |
| Point | | X | X | X |
| Multipoint | | X | | X |
| Polyline | | X | X | X |
| Chain / Arc | | | X | |
| Junction | | | | X |
| Edge | | | | X |
| Route | | | X | |
| Polygon | | X | X | X |
| Network | | | X | X |

**Key terms**
A **logical data model** both enables and constrains choices simultaneously.

# Raster and Vector Data Structure Primitives

Raster data structure – data constructs
– a matter of size for data array

- **pixel** primitive, e.g., 30 meters by 30 meters, commonly used to partition image space
- **cell** primitive, e.g., 1 kilometer by 1 kilometer, commonly used to partition map space

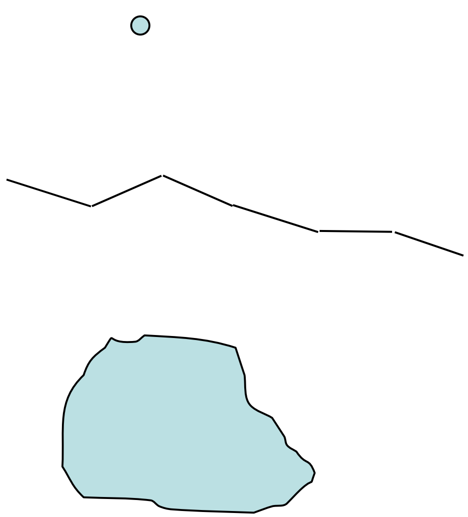


Vector data structure – data constructs
- a matter of dimension (D)

0-D **Point**

1-D **Polyline**

2-D **Polygon**



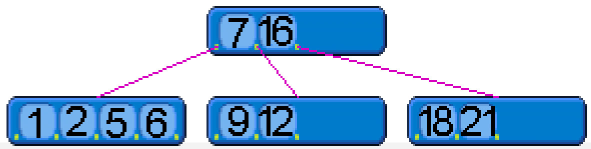
# Specifying data formats for data structure fields

... a choice of data type formatting for spatial coordinate storage,
e.g., single precision or double precision

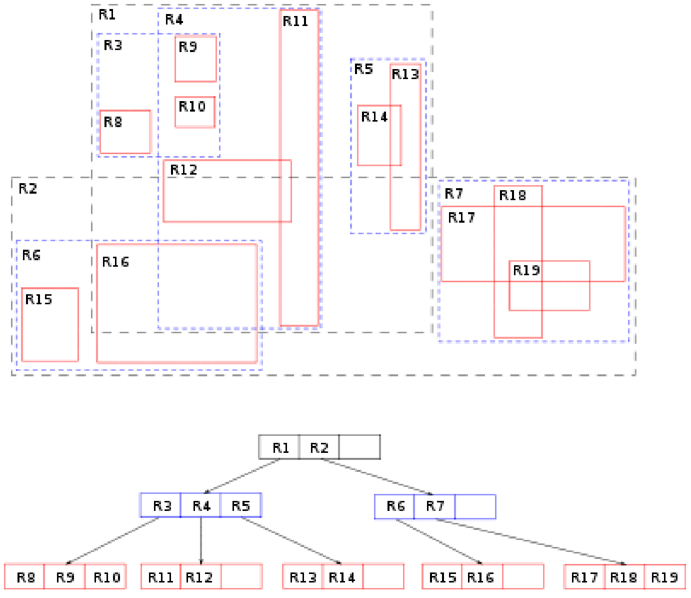
... a choice of data type formatting for time data storage,
e.g., date and time, just date

... a choice of data type formatting for attribute data fields,
e.g., string, character, alphanumeric, numeric

... a choice of indexes for performance retrieval from disk,
e.g., B-tree (binary tree), R-tree (region tree)




A B–tree of order 2 (Bayer & McCreight 1972 or order 5 (Knuth 1997) from Wikipedia




Simple example of an R–tree for 2D rectangles from Wikipedia

# $2^{nd}$ component - data model operations

CRUD operations

Query operations

Geometry operations

Topology operations

Spatial analysis operations

Spatial SQL

Transactions

**Key terms**
operations

# Database Operations of Logical Data Models

Y&H Chapter 2 Database operations

Create, Read, Update, Delete (CRUD) operations

Create – add row and/or data value

Read – retrieve data value

Update – change the data value

Delete – eliminate the data value

**Key terms**
CRUD

# Query Processing Operations

Database queries are formed using operators

SELECT – query rows of a table, all data values or those meeting criteria

PROJECT – query columns of a table, removing duplicate data values

JOIN – concatenates data from rows in one table with rows in another table on same value

PRODUCT – concatenate every row in one table with every row in another table

UNION – generate new table by appending rows from one table onto another

INTERSECT – generate new table consisting of all rows appearing in both tables

DIFFERENCE – generates new table consisting of all rows in one table not appearing in another

DIVIDE – takes a two column table and a one column tables and cerates new table consisting of all vales of one column of binary table that match values in the single-column table

**Key terms**

# Geometry Operators

Spatial reference – returns reference system

Envelope – returns the minimum bounding rectangle of geometry

Export – converts geometry into different representation

IsEmpty – tests if geometry is empty set or not

IsSimple – returns TRUE if geometry is simple

Boundary – returns of the boundary of the geometry

**Key terms**

# Topological Operators

Equal – tests if geometries are spatially equal
Disjoint – tests if geometries are disjoint
Intersect – tests if geometries intersect
Touch - tests if geometries touch each other
Cross - tests if geometries cross each other
Within - tests if geometries is within another geometry
Contain – tests if a given geometry overlaps another given geometry

Returns TRUE
if spatial relationship specified by the 9-matrix

Relationship between point, line, area (3) and
another point, line, area (3) , so 3 x 3 = 9-matrix

**Key terms**

# Spatial Analysis Operators

Distance – Returns shortest distance between any two points

Buffer – Returns a geometry that represents all points at a pre-defined distance

ConvexHull – Returns convex hull of a given geometry

Intersection – Returns the geometry of two geometries

Union – Returns the combination of two geometries

Difference – Returns the separate geometry of the two geometries

SymDifference – returns the symmetric difference of two geometries; 0 if the same; 1 if different

**Key terms**

# Spatial SQL

Check out Y&H Figure 4-15 p. 124 for examples of spatial SQL

SELECT, PROJECT, JOIN etc. originally attribute only

In 1980's, SQL became coordinate-enabled

**Key terms**

# Database Transactions – following principles

Transaction – input what is to be done, process the data, and provide the output – as a package.

Atomicity – all or nothing of transaction is executed

Consistency preservation – data remains consistent

Isolation – results of simultaneous transactions are independent

Durability – after transaction, the result can be traced from beginning

**Key terms**

# Four transaction control mechanisms

Concurrent control – locks data items involved in transaction

Logging transactions – keep track of all changes made to data, enables redo

Transaction commitment – prevents change to database unless transaction is ready to complete

Rollback – allows database to undo an incomplete transaction

Y&H Figure 2-4 p. 36 provides the steps in a database transaction

**Key terms**

# 3rd component – data model rules

Rules are specified to preserve the integrity, veracity, and validity of the data.

We set the "bounds" of the data instances that can be stored.

Data Model Rules include

- Integrity Rules for Domains Y&H pp. 30-31, 116-117
- Topology Rules
- Subtypes Rules
- Relationship Rules

**Key terms**

# 3. What is a database model?

A database model makes use of a data model to express the design of a database using feature classes, relationships, and other constructs for a particular problem or set of problems.

There is a close relation between data model and database model.

A data model expresses the design for DBMS software.

A database model expresses the design, that is, the application for a database. A schema articulates that design in the form of data structure primitives, e.g. points or polygons etc.

The data model is a 'design framework', whereas the database model is an application of the design framework for creating a particular database design.

**Key terms**

# 4. What is data (database) modeling?

Data (Database) modeling – development work performed across three levels of data abstraction

- Conceptual model – important entities of the world provide meaning for organizing data(database)
- Logical model – data classes designed to be stored in DBMS, based on a particular DBMS software design
- Physical model – data elements formatted within a particular data model for storage on disk for retrieval

Each of the levels contains a schema, that is, the mechanism for specifying the database description at that level.

- If development is about a data model then the work task is about software at conceptual, logical and physical levels.
- If development is about a database model then the work task is about database design at conceptual, logical, and physical levels.

**Key Terms**
data modeling
abstraction
conceptual
logical
physical

# In data (database) modeling there are three components and three levels of abstraction…

| Three Levels of Abstraction | Three Components | | |
|---|---|---|---|
| | **Schema Constructs** for describing | **Operations** for manipulating | **Constraints** for validating |
| Conceptual | …worldly features | …worldly processes | …features and processes |
| Logical | …data primitives of the database | …data primitives of the database | …data and operations on the database |
| Physical | …of disk storage | …of computer bytes and bits | …reads and writes to disk |

**Key Terms**
constructs
operations
constraints

# Geodatabase Design Process – ten steps

**Conceptual Design of a Database Model**

1. Identify information products and/or "need to know" questions
2. Identify the key thematic layers and feature classes.
3. Detail all feature class(es)
4. Group feature classes into datasets (logical collections)

**Logical Design of a Database Model**

5. Define attribute database structure and behavior for descriptive attributes, e.g. using rules
6. Define spatial properties of datasets, e.g. using rules
7. Propose a database design

**Physical Design of a Database Model**

8. Implement, prototype, review and refine schema design
9. Design workflows for building and maintaining each layer
10. Document design using appropriate methods

# Specifying conceptual data model schema

**1. Identify the information products or the research question discussed as part of the problem definition.**

- Over the past few years Professor Nyerges has been compiling resources about coastal data models
  - Existing data models, e.g. ArcHydro, ArcMarine
  - Current literature, e.g. Puget Sound Partnership strategic materials, Puget Sound Nearshore Ecosystem Restoration Program reports
  - Textbook materials
  - Experts in the field
- In particular, materials about non-point source pollution have been important because of the social, economic, and ecological implications of such processes in coastal watersheds

**Key Terms**

Knowledge sources

# Specifying conceptual data model schema

## 2. Identify the key thematic layers and feature classes

| |
|---|
| 2 Barrier Islands |
| 2 Estuaries |
| 2 Coastal Marshes |
| 24 Coral Reefs |
| 2 Rocky Shores |
| 2 Bluffs |
| 2 Wetlands |
| 12 Habitats |
| 24 Soil Composition |
| 24 Land Cover |
| 124 Pollution and toxic containments |
| 24 Land Use and Zoning |
| 2 Building Code |
| 2 Ports |
| 4 Precipitation |
| 134 Water quality |

| |
|---|
| 2 Ferry Systems/Water Taxi |
| 2 Present Buildings/Structures |
| 2 Roads Network |
| 3 Sewage Utility Piping Network |
| 2 Sea Walls |
| 12 Tides |
| 12 Currents |
| 2 Winds Patterns/Flow |
| 24 Erosion and Accretion |
| 2 Migratory Animals (e.g. birds) |
| 12 Catch Basins/catchments |

| |
|---|
| 124 Watershed Elevation |
| 12 Streams/Rivers Flow |
| 12 Continental Shelf/Slope |
| 12 Water Depth/Slope |
| 12 Canal Shape/Depth/Slope |
| 3 Critical Area Ordinances (CAO) Spaces |
| 3 Freshwater sources / treatment plants for freshwater |
| 3 Waste Treatment locations |
| 4 Agriculture |
| 1234 Shoreline/shorezone |

**Key Terms**

Feature class

**Identified from:**
1 – ArcHydro and ArcMarine Data Models
2 – Beatley, Brower, Schwab 2002
3 – Puget Sound Nearshore Task Force, 2006
4 – N-SPECT model materials, 2005

# Specifying conceptual data model schema

**3. Detail all feature classes**

Each of the features can be detailed in terms of attributes that further characterize the feature class.

Estuary
    Size, Form, Water Quality

Watershed
    Size

Land cover

Land use and zoning

Key to database design is knowing when to identify a characteristic of the problem as a feature class or an attribute. Almost any feature class can be transformed into an attribute, and vice versa. For example, "land use" above can be transformed from a feature class into an attribute, since parcels are what really "have uses" and each parcel is "zoned" for some use

Parcel
    Land use, Zoning

**Key Terms**
Attribute
Feature class
Transformation

# Conceptual schema design

**4. Group representations into datasets – schema design pattern 1.**

1st of 9 schema design patterns as a Foundation of Geodatabase Design

Feature Datasets (see A&Z 16-17)

Use feature datasets to group feature classes for which you want to design topologies or networks or those you wish to edit simultaneously because you can develop dependencies between/among features.

A list or class diagram could be used. Class diagrams as a dialect within the Unified Modeling Language can be used to express the design. MS Visio creates class diagrams.

# Conceptual schema design

## 4. Group representations into datasets

| Human Infrastructure/Impact |
|---|
| * Pollution and toxic containments |
| * Land Use and Zoning |
| Building Code |
| Ports |
| Ferry Systems/Water Taxi |
| Present Buildings/Structures |
| Roads Network |
| Sewage Utility Piping Network |
| Sea Walls |

| Water and Water Bodies |
|---|
| * Catch Basins/catchments |
| * Watershed elevation |
| * Streams / Rivers Flow |
| * Precipitation |
| * Water Quality |

| Dynamic Natural Phenomena |
|---|
| Tides |
| Currents |
| Winds Patterns/Flow |
| * Erosion and Accretion |
| Migratory Animals (e.g. birds) |

| Underwater Topography |
|---|
| Continental Shelf/Slope |
| Water Depth/Slope |
| Canal Shape/Depth/Slope |
| Critical Area Ordinances (CAO) Spaces |
| Freshwater sources / treatment plants for freshwater |
| Waste Treatment locations |

| Physical/Natural |
|---|
| Barrier Islands |
| Estuaries |
| Coastal Marshes |
| Coral Reefs |
| Rocky Shores |
| Bluffs |
| Wetlands |
| Habitats |
| * Soil Composition |
| * Land Cover |
| * Agriculture |
| * Shoreline / shorezone |

**Key Terms**

Critical feature class

*** Critical class for N-SPECT**

# Logical Design

**5. Define attribute database structure and behavior for descriptive attributes**

5a. Define that attribute database structure and behavior for features using domains – schema design pattern 2

Specify valid values and ranges for attribute fields, apply subtypes to control behavior, create relationships with rules for association, and classifications for complex code domains. Specify primary keys and types of indexes.

Classification and Domains (A&Z 32-34)

- Attribute domains define valid set of values for a field
- range of valid values
- lists of values, and
- standard classifications

**Key terms**

# Logical Design

5b. Define that attribute database structure and behavior for features using subtypes – schema design pattern 3.

Subtypes (A&Z p18-19) define

- Attribute domains
- Default attribute values
- Split-merge policies
- Connectivity rules
- Relationship rules
- Network rules
- Topology rules

**Key terms**

# Logical Design

5c. Define that attribute database structure and behavior for features using relationships – schema design pattern 4

*Relationships* – If the spatial and topological relationships are not quite suitable, a general association relationship might be useful to relate features.

Relationships (A&Z pp 20-21)

Implement general associations for features and objectives, depending on display and query requirements. Associates objects from a feature or table to objects in another feature class or table.

**Key terms**
Relationships

# Logical Design

6. **Define spatial properties of datasets** to establish the spatial reference system for the dataset.

6a. Specify the survey datasets that provide control for coordinates using design pattern 5
6b. Specify rules to compose topology for connected systems of features as design pattern 6
6c. Specify the raster datasets as appropriate as design pattern 7
6d. Networks
6e. Raster data – schema pattern
6f. Labeling and Annotations

**Key terms**
Rules

# Logical Design

*6a. Survey data* – survey data allow an analyst to integrate survey control (computational) network with feature types to maintain the rigor in the survey control network as design patter 5.

**Key terms**

Survey data (A&Z 26-27)

- Comprehensive survey measurements used to manage computation networks.
- Linked to and used to update features using
    - survey points – survey locations can be associated with feature geometry
    - feature – vertices in survey-aware feature classes
    - measurements – store dependencies among computations and refines positional accuracy

# Logical Design

*6b. Geocoding data in the form of*

- coordinates,
- street addresses,
- river miles,
- road reference points etc

need to be identified to provide fundamental locational orientation.

**Key terms**
Geocoding

# Logical Design

6c. *Topology* (A&Z 22-23)

Topologies define how feature share geometry and control of integrity through rules and editing behavior (e.g., census blocks cannot overlap one another and share geometry with street centerlines.) as design pattern 6.

Two types of topology

- Geodatabase topology
    - grouping of feature classes
    - ranked feature classes
    - topology rules

- Map topology
    - simple topology impose relationships among the point, line, polygon features
    - shared-edge editing tools
    - edit feature classes across feature datasets and shapefiles

**Key terms**
Topology

# Logical Design

6d. *Networks*

Geometric networks offer quick tracing in network models. These are rules that establish connections among feature types on a geometric level and are different than the topological connectivity. Such rules establish how many edge connections at a junction are valid as design pattern 7.

Linear networks (A&Z 24-25)

- Geometric network
- Logical network

Rules for managing connection among features in a set of feature classes.

**Key terms**
Networks

# Logical Design

*6e. Raster data* – schema pattern

Analysts can introduce high performance raster processing through raster design patterns. Raster design patterns allow for aggregating rasters into one overall file, or maintain them separately as design pattern 8.

- Raster (A&Z 28-29),
    - Raster dataset – stores single raster
    - Raster catalog – container of raster datasets

- Four classes of raster datasets:
    - Orthorectified imagery
    - Scanned maps
    - Raster elevation data
    - Raster time series

**Key terms**
Raster

# Logical Design

*6f. Labeling and Annotations* – At least two major types of text are important, labeling that describes particular features and annotations that describe other phenomena relevant to a map reader as design pattern 9.

**Key terms**

Labeling & Annotation (A&Z 30-31)

- Descriptive text specified as either automatic labeling, based on feature attribute values, and use of annotation features, each edited and positioned manually.

Dynamically displayed labels

- Feature-linked annotation – annotation class joined in composite relationship
- Simple annotation – standalone annotation class
- Labels – text properties, such as font and size, and defined labeling field

# Logical Design

**7. Propose a geodatabase Design**

- Assemble the descriptions into specific database design using all information compiled.

- Document using database design language, e.g., as in

**ArcGIS Diagrammer 10.2**

- Specify the indexes use to build directories

**Key terms**
ArcGIS Diagrammer

# Physical Design

**8. Implement, prototype, review, and refine design**

Five approaches to implementation of geodatabase schema

See diagram A&Z *Designing Geodatabases* p. 381

- Create a geodatabase schema with ArcCatalog from scratch
- Load existing data into empty geodatabase to create basic schema
- Specify geodatabase schema with Unified Modeling Language (e.g. using MS Visio)
- Use a geodatabase schema template (download for ESRI site)
- Use scripting e,g. Python, to develop specification

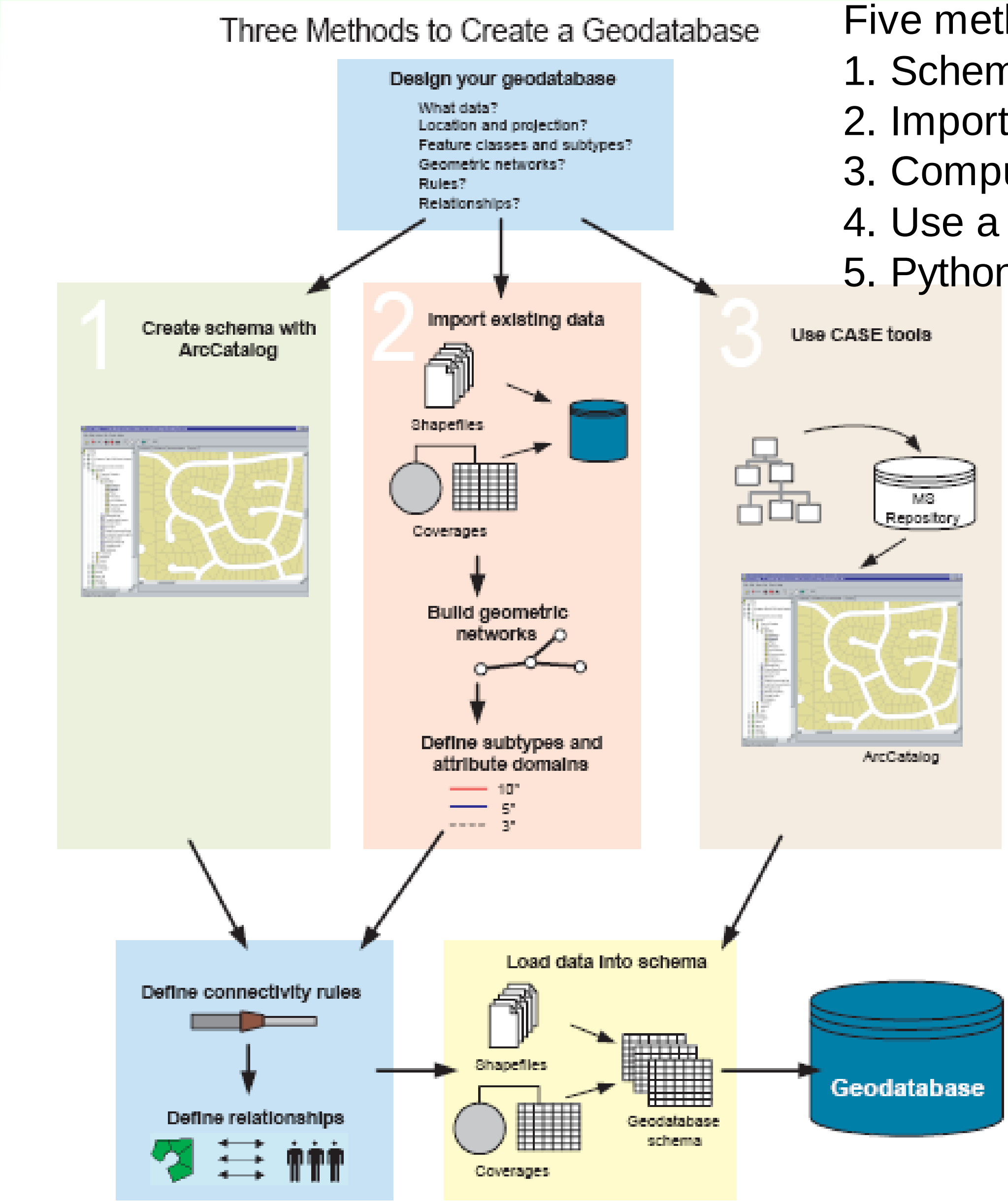
Modify the design as appropriate A&Z p. 382

Load the data A&Z p. 384-388

Test the design, A&Z p. 388-389

**Key terms**

# Geodatabase Schema Development




Five methods for creating a schema

1. Schema with ArcCatalog
2. Import existing data
3. Computer Aided Software Engineering
4. Use a template from Esri
5. Python program – write a script

Lab assignments will guide us through steps that combine these approaches

# Physical Design

**9. Design workflows for building and maintaining each layer**

- Activity modeling
- What is the layer for?
- Why is the layer maintained?
- Who is responsible for maintaining the layer?
- When and how often is the layer maintained?

**Key terms**

# Physical Design

**10. Document schema design using methods available**

**Three tools can be used to examine schemas:**

**Geodatabase Diagrammer** – A&Z p. 12-13

ArcCatalog plug-in command used for creating graphic presentations as posters.

**ArcGIS Diagrammer 10.1**

A stand-alone tool for designing and documenting geodatabases

Created after Geodatabase Diagrammer, and considered easier to use. No longer supported by Esri developers.

**X-Ray for ArcCatalog**

Used for inspecting/editing particular fields within a schema

**Key terms**
Geodatabase Diagrammer
ArcGIS Diagrammer
X-Ray for ArcCatalog

# 5. What are some example application data models?

ESRI provides resources for industry-specific data models to get GIS projects started. A list is available at…

http://support.esri.com/technical-article/000011644

For example,

Address

Agriculture

Atmospheric

Basemap

Biodiversity

BroadbandStat

Building Interior Space

Carbon Footprint

Census - Administrative Boundaries

Defense - Intel

Energy Utilities (includes ArcGIS MultiSpeak)

Environmental Regulated Facilities

These are only a sample, but the list shows a diverse array of templates are available.

# Databased models continued…

- Fire Service
- Forest Service
- Forestry
- Geology
- GIS for the Nation
- Groundwater
- Health
- Historic Preservation and Archaeology
- Homeland Security
- Hydro
- International Hydrographic Organization (IHO) S-57 for ENC
- Irrigation
- Land Parcels
- Local Government
- Marine
- National Cadastre
- Petroleum
- Pipeline
- Raster
- Seabed Survey
- Telecommunications
- Transportation
- Water Utilities

# Summary

In this lesson, you learned about…

1. Three realms for representation modeling

2. The character of data models

3. Character of database models

4. The process of data (datsbase) modeling

5. Examples of application data models

Contact me at nyerges@uw.edu if you have questions or comments about this lesson.

GEOG 482/582: GIS Data Management

# END Lesson 2: Data Models, Database Models, and Data Modeling