Geog 461 Learning Objective Outline

LOO 08 Geospatial Database Design
08.1 What are the general steps database design?
Nyerges and Jankowski *GISDS* Chapter 3 Section 3.3 Database Design Process for Green County Functional Planning

A database model (using a particular data model schema) is an expression of a collection of object classes (entities), attributes and relationships for a particular domain context. A database model and the associated database are the foundation of the "representation model" described by Steinitz (2004) in his six-phase GIS workflow process.

A database model can be expressed at each of the three levels of abstraction, conceptual, logical and physical as described previously. These are called levels of database abstraction because we choose to select (abstract) certain salient aspects of a database design.

Below are steps outlining a geodatabase design process adapted from Arctur and Zeiler (2004) *Designing Geodatabases*, ESRI Press. The process includes conceptual, logical, and physical design phases to help clarify the overall process of database development (requirements scoping, design and implementation). Each of those phases ends in the creation of a product called a database model at the respective level of conceptual, logical, and physical.

Table 3.5 Geodatabase Database Design Process as Data Modeling

Conceptual Design of a Database Model

- Identify the information products or the research question to be addressed
- Identify the key thematic layers and feature classes.
- Detail all feature class(es)
- Group representations into datasets

Logical Design of a Database Model

- Define attribute database structure and behavior for feature classes
- Define spatial properties of datasets

Physical Design of a Database Model

- Data field specification
- Implementation of schema
- Populate the database – an implementation step

NOTE: This is an iterative process, NOT just a linear pass through these steps. Something you discover in a lower step helps re-orient what you had done previously. Thus, you "upgrade and refine" your earlier ideas and create a better design.

08.2 What are the steps for conceptual design of a database model?
Nyerges and Jankowski *GISDS* Chapter 3 Section 3.3.1 Conceptual Design of a Database Model

1. **Identify the information products or the research question to be addressed.** To the best information available, identify the information products that will be produced with the application(s). For example a product might be a water resource, transportation, and/or land use plan as an array of project improvements conceptualized to improve a community over the next twenty years (give or take).

A GIS data designer/analyst would converse with situation stakeholders about the information outcomes to appear in the product, rather than guess. If you are the stakeholder, then mull it over a bit to make sure you have an idea. Some guidance should be available in terms of a project statement. In the Green County project, the County Council provided the purpose and the objectives in siting a wastewater treatment facility. In another context, perhaps the purpose is a research statement, in which one or more research questions have been posed. Sometimes such questions are called "need to know questions". For example, what do the stakeholders "need to know" about the geographical decision situation under investigation? What are the gaps in information, evidence, and/or knowledge? What information is not available that should be in order to accomplish tasks related to decision situations? What changes (processes) in the world are important to the decision situation? What are the decision tasks? Those questions should help the reader articulate "information needs" as a basis of data requirements.

From a landscape modeling perspective as described in chapter 2, we can develop value structures that underpin the information needs of decision models. What we store in databases are "data values" to be able to derive information from the representation models through to decision models. From where does this "value" arise? What fosters the development of certain (data) values in our databases? Looking back to the conceptual data modeling process, there is undoubtedly some reason why certain data categories are chosen and others not. The answer lies in what is "valued" to be represented.

> The single most important factor determining the future of our environment is people's sense of values. … The problems of the environment are not, fundamentally, scientific or technical – they are social … Values are the hardest things to discuss, but society's values are the driving force which determines what it does and does not do. Only when we know who we want to be and why, can we start to question whether our current actions are true to that ideal. (IUCN 1997 pp. 16-18)

2. **Identify the key thematic layers and feature classes.** A thematic layer is a superclass of information, commonly consisting of a dataset(s) and perhaps several feature classes (hence feature layers), convenient for human conversation about geographic data. For each thematic layer, specify the feature classes that compose that thematic layer. For each feature class specify the data sources potentially available, spatial representation of the class, accuracy, symbolization and annotation to satisfy the modeling, query and or map product applications.

The conceptual design of the Green County geodatabase **depicted in Figure 3.6** contains several feature datasets that are the thematic layers in the database design. Although the original Green County GIS project did not need feature datasets because the siting problem was cast as a rather simple problem, the expanded GIS project can make use of them. Each dataset is created using a *package* in UML class diagrams (**Figure 3.7**)

3. **Detail all feature class(es).** For each feature class, describe the spatial, attribute, temporal data field names for the class. For each feature class specify scale range for spatial representation, and hence the associated spatial data object types? This will determine if multiple resolution datasets for layers are needed. Revisit step 2 as needed to complete the specification. Identify the relationships among the feature classes.

4. **Group representations into datasets.** A feature dataset is a group of feature classes that are *organized based on relationships identified among the feature classes* that help in generating the information needed by problem stakeholders. The dataset creates the instance of "thematic layer" or a portion of the thematic layer in which the relationships among feature classes are important for deriving information. Analysts name feature classes and feature datasets in a manner convenient to promote shared understanding among analysts and stakeholders. We use feature datasets to group feature classes for which you want to design topologies or networks or those you wish to edit simultaneously.

See **Figure 3.8** Land feature dataset in the Green County conceptual database design.

08.3 What are the steps for logical design of a database model?
Nyerges and Jankowski *GISDS* Chapter 3 Section 3.3.2 Logical Design of a Database Model

5. **Define that attribute database structure and behavior for features.** Apply subtypes to control behavior, create relationships with rules for association, and classifications for complex code domains.

Subtypes – Subtypes of feature classes and tables preserve coarse-grained classes in a data model, improve display performance, geoprocessing and data management, while allowing a rich set of behaviors for features and objects. Subtypes let an analyst apply a classification system within a feature class and apply behavior through rules. Subtypes help reduce the number of feature classes and improve performance of the database.

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

6. **Define spatial properties of datasets.** Specify rules to compose topology that enforces spatial integrity and shared geometry and specify rules to compose networks for connected systems of features. Specify the spatial reference system for the dataset. Specify the survey datasets that provide control for coordinates if needed. Specify the raster datasets as appropriate.

Spatial Reference
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.
GIS DB Primer Chapter 2

Geocoding data – geocoding in the form of coordinates, street addresses, river miles, road reference points etc need to be identified to provide fundamental locational orientation.
GIS DB Primer Chapter 2

Topology – Topology rules are part of the geodatabase schema and work with a set of topological editing tools that enforce the rules. A feature class can participate in no more than one topology or network. Geodatabase topologies provide a rich set of configurable topology rules. Map topology makes it easy to edit the shared edges of feature geometries.
GIS DB Primer Chapter 3 Section 3.3.1 Figure 3.15

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.

See Figure 3.9 Wastewater network class diagram (additional part of Green County database design)

Raster data – 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.

08.4 What are the steps for physical design of a database model?
Nyerges and Jankowski *GISDS* Chapter 3 Section 3.3.3 Physical Design of a Database Model

7. **Data field specification**. For data fields, specify valid values and ranges for all domains, including feature code domains. Specify primary keys and types of indexes.

Classifications and domains – Simple classification systems can be implemented with coded value domains. However, an analyst can address complex (hierarchical) coding systems using valid value tables for further data integrity and editing support.

At this time that primary and secondary keys for the data fields are specified, based on the valid domains of each of the fields. A data key reduces the need to perform a "global search" on data elements in a data file. A primary (data) key is used to provide access within the collection of features that can be distinguished by a unique identifier.

Table 3.4 Data Types
See Figure 3.10 Datatype specifications
See Figure 3.11 ObjectID for primary key

8. **Implementation**. Construct data schema to reside in a database management system. Test the computability of the data schema.

See **Figure 3.12** for schema semantics check – error report
Report is generated when using the MS Office Visio UML schema generator

9. **Populate the database –** not really part of design, but major part of DB development.

This involves loading the data into the schema.
Data loading can come from a variety of sources, including the geography server.