Building geodatabase 2.

Geodatabase design steps

Ferenc Végső

Building geodatabase 2.: Geodatabase design steps

Ferenc Végső

Lector: Szabolcs Mihály

v 1.0

Publication date 2010

Abstract

In this module we are show the geodatabase design steps, including defining entities and its relations, intelligent individuals, and modeling the user's view.

Table of Contents

Chapter 2. Geodatabase design steps

1. 2.1 Introduction

The conceptual design of the GIS system is primarily an exercise in database design. It includes formal modeling (preparation of a data model) of the intended GIS database and the initial stages of the database planning activity. Database planning is the single most important activity in GIS development. It begins with the identification of the needed data and goes on to cover several other activities collectively termed the data life cycle - identification of data in the needs assessment, inclusion of the data in the data model, creation of the metadata, collection and entry of the data into the database, updating and maintenance, and, finally, retention according to the appropriate record retention schedule. A complete data plan facilitates all phases of data collection, maintenance and retention and as everything is considered in advance, data issues do not become major problems that must be addressed after the fact with considerable difficulty and aggravation. The product of the conceptual design activity is a data model which rigorously defines the GIS database and supports the detailed database planning activity.[1] The geodatabase model is essentially an object-oriented data model. The new data model is designed to define intelligent individuals and relationships with the other individuals. The geodatabase model brought closer the physical and logical data model, i.e., less abstraction was needed. As we saw earlier, the growth of abstraction reduces the similarity of data models to reality. The geodatabase model allows the definition of individual behavior without programming. Most of the property inheritance rules defined by or through pre-defined rules (such as the water separator assembly may be only on a conduit). Program should only be written if an object's behavior is very specific (e.g. transport network model).

The detailed database planning and design task includes the following activities: developing a logical or physical database design based on the data model prepared earlier, evaluating the potential data sources, estimating the quantities of geographic data, estimating the cost of building the GIS database and preparing the data conversion plan. Concurrent with the detailed planning for the database, pilot studies and/or benchmark testing that are desired can be executed. Information gained from these studies and tests will be needed to estimate the size of the equipment (disk space, main memory etc.) and to determine how much application development will be necessary. Subsequently, plans for staffing, staff training, equipment acquisition and installation, and user training must be completed. After the preparation of all these plans, the entire cost of the GIS will be known and the final feasibility assessment can be made.[2]

2. 2.2 Intelligent individuals

The previous paragraph mentioned the intelligent individuals. Before proceeding to the modeling of real-world, let's look at what makes individuals to smart. In the geodatabase vector data, raster data, surface data and object data can occur as well. Now, the vector data is presented to show the intelligence of individuals. The GIS database including a large amount of vector data.

Features have shape

The individuals - when they undergo abstraction - are distinct and can be described by their shape. The data of features is stored in the specialized field of attribute table that we know as 'geometry' field. The field is seen only through the user interface and has symbolic entries (point, line, polygon).

The type of shapes:

• Points and multipoints, which are a set of pints.

• Polylines, a set of line segments that connected (or may be not connected)

• Polygons, a set polylines that connected, closed and no intersected.

The line segments can be straight, circular, arcs, Bézier curves and elliptical arcs.

Features have a spatial reference

---

[1] Paul Becker et al.: GIS development guide, 2006.
[2] Paul Becker et al.: GIS development guide, 2006.

The shape of a feature described with Y, X values in Cartesian coordinate system. However, the shape of the earth is not flat, the best approach is the geoid. The projection specifies, how to the individual location mapped on the earth's surface.

Features have attributes

A feature maintains its attributes as fields in attribute table. The basic kinds of descriptive data can be numbers, text or other (date, object identification data or multimedia).

Address	Year of built	Footage (nm)	No. of bedrooms
Kossuth u. 12.	1953	150	4

Features have subtypes

Features are forming a feature class. Feature classes are homogenous sets of features, but may include a significant subgroup of features. A feature class comprising buildings can be logically subdivided into subtypes (commercial, industrial, residential). Subtypes give you increased control on features. For example when a road belongs to a subgroup of unpaved roads, cannot continue like motorway.

Features have relationships

All features in the universe have relationship to other features. You can define spatial relationships between features (named topology), or you can define also non-spatial relationships between objects, such as relationship between a house and its owner.

Feature have domain

Each attribute of a feature can have an attribute domain. Domain can be a numeric range, or a list of valid rules or abbreviations. Each attribute can also have a default value automatic assigned, when the feature created.

Footage (nm)
150

The footage should be between 30 and 300

No. of bedrooms
4

The No. of bedrooms can be only 0, 1, 2, 3, 4 or 6

Features can be validated

Objects in the world follow rules when they are moved or changed. By these rules, you can check how to relate the part of the network (e.g. different diameter tubes can join only with restrictor), or be controlled to a number of property owner.

Features can have topology

The majority of features have clearly identifiable neighborhood relations, which represented by topology. The properties cannot be overlap and should be connected without any gap. This two-dimensional planar geometry called topology. The network of lines and fittings is to be connected without any gap. This network is called the one-dimensional geometry, network geometry.
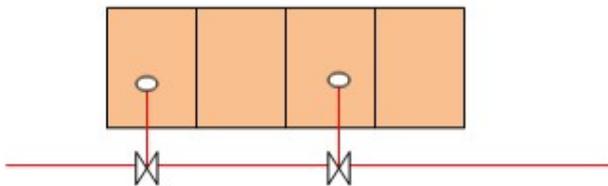
Features can have complex behavior

The simple features defined by geometry, topological properties, attributes, attribute domains and validation rules. More complex behaviors of features can be implemented to writing software code for a *custom feature*. Custom features permit complex behavior such as custom editing possibility, special analytical capabilities and so on.

2.1. 2.2.1 Benefits of the geodatabase model

Some of the benefits of the geodatabase model are:

• **Uniform collection of geographic data.** All geographic data can be stored and managed centrally in one database.

• **Users work with more intuitive data objects.** Instead of generic points, lines and polygons, users work with objects if interests, such as transformers, roads and lakes.

• **Data entry and editions are more accurate.** Most of data entry and editions error can be prevented by intelligent validation of behavior. For many users, this alone is enough reason to adopt geodatabase model.

• **Features have a rich context.** With smart features you not only define a feature's qualities, but its context with other features. This lets you specify, what should happen when a related feature moved, splitted, deleted or so.

• **Better maps can be made.** You have more control over how features are drawn, and you can add intelligent drawing behavior.

• **Features on a map display are dynamic.** The features can respond to changes neighboring features.

• **Sets of features are continuous.** The geodatabase can accommodate very large sets of features, rasters without spatial partitions.

• **Many users can edit geodatabase simultaneously.** This data model permits work flows where many people can edit features in a local area and reconcile any conflict that emerge.[3]

---

[3]Michael Zeller: Modeling our World. 2003.

3. 2.3 Geodatabase design

The geodatabase design is basically the same as any database design. The geodatabase is essentially a relational database with the addition of a database that includes geographic data. The geodatabase already contains the geographical and topological relations of spatial entities. This structure has a part of a special type of data, the topology, which is suitable to describe integrated systems, such as drainage or road network. The geodatabase data model is a bridge between the world's perception of the people and the objects in a relational data storage and information management technology.

Relational database design include two basic steps: drawing up of a logical data model and the physical implementation of database models. The logical data model shows the user's view of data and the database model implements the data model within the framework of relational database technology.

3.1. 2.3.1 Designing a logical data model

The key task is to precisely define the set objects of interests and to identify the relationships between them. The objects usually consider as everyday things, such as roads, land, buildings, owners, etc.. The relationship is expressed in natural language, such as the "next", "owner", "part". Designing the data model is generally not a straithforward process. The initial model may be filled with data, can be tested, may be compared to the needs of the user and the user's organization (company, agency, authority), a business practice or policy. Particularly important to involve to the design the representatives of users group. This is the key to reaching the needs of users by a satisfactory data model. As already mentioned, the creation of logical data model is an iterative process, and are typically based on experiences. Does not exist somewhere in a "real" model, but there are better and less good models. It is difficult to determine exactly when the model is a good and complete, but there is some indication:

• Is the logical data model contains all the information possible, without repetition? • Does the logical data model supports the organization's business (legislative) policies? • Does the logical data model contains the different views of different groups of users for the data? (for the engineer the pipeline a material with diameter, for the accountant the cost element).

3.2. 2.3.2 Representing logical data model

The physical database is prepared on the basis of the logical database model. In most cases, a skilled professionals for relational database construction will receive the logical data model from the data modeler. By the commands of the database management software, the database manager creates a database structure, and defines the parts of the database. Then you can read or typing in the data. The physical database is very similar to the logical database, but for technical reasons it differs much from it. For display, in the tables the objects can be merged or splitted. The rules and relationships can be expressed in different ways. The main advantage of the geodatabase model, so that the data is included, that they can best recall to the logical data model. In other words, the physical transformation is not completely hiding the database logic from the user. The former (e.g., file-oriented) databases were only available through a programmable interface, the user can not see anything about the internal structure of the database. The main disadvantage was that the user always had to rely to the developer.

3.3. 2.3.3 Elements of the logical and physical database models

The following table shows the basic elements of the logical data model and their corresponding database elements:

Logical elements	Database elements
feature	row
attribute	field or column
class	table

The logical data model is an abstraction of the feature, that we encounter in the specific application. This abstraction is converted to the database elements. The feature represents a real object, such as buildings, lakes or the consumer. The feature is stored as row in the table header. Objects have a set of attribute data. The attribute data shows the quality parameters of the object, such as name, size, quality or identifier as the key to another object. The attributes are stored in columns (or fields) in the database. The class is a collection of similar features. In a class all instances of the same data set descriptor. The class is stored in the database as a table. The rows and columns in a table forms a two –dimensional matrix.

3.4. 2.3.4 Problem of the complex data

The relational database management implements a simple, elegant, easy to understand and transparent structure. The simplicity is also the disadvantage - a relational database is easy to define, but difficult to model complex data. However, the geographic database contain complex data types. The line or a series of closed polygon are structured sets of coordinates, and can not be described as an elementary data type like integer, floating number or string. In addition, the features have collected in a database, should include topological information and spatial relations and overall relationship as well. A relational database is the foundation for the geodatabase. The main purpose of the geodatabase is to handle complex data in unified data model, independent of the relational database management underneath.

3.5. 2.3.5 Steps for geodatabase design

The structure of geodatabase lets you design geographic datasets that are close to their logical data model.

These are the basic steps in designing a geodatabase:

• **Modeling the user's view.** Perform talk (interview) with users, understand organization, and analyze the current and possibly future needs of users. • **The definition of objects and their relationships.** The objects should be incorporated into the logical data model to take account of how relations are between them. • **Select the geographic representation.** We need to decide whether a vector, raster, or the surface is best to represent the data of interest. • **Match to elements of the geodatabase.** Fit the objects in the logical data model into the elements of a geodatabase. • **Organize geodatabase structure.** Build the structure of a geodatabase. Consider thematic groupings, topological associations and department responsibility of data.

4. 2.4 Model the user's view

The objective of this step is to ensure a common understand between the design team and those who have interested to use a GIS.

4.1. 2.4.1 Needs Assessment

The GIS needs assessment is designed to produce two critical pieces of information:

• The list of GIS functions that will be needed

• A master list of geographic data.

These two information sets are extracted from a set of GIS application descriptions, a list of important data, and a description of management processes. Standard forms are used to document the results of user interviews. The information gained in the needs assessment activity goes directly into the Conceptual GIS Design activity.

4.2. 2.4.2 Survey of Available Data

A survey of available data can commence once needed data have been identified in the needs assessment. This task will inventory and document mapped, tabular and digital data within the local authorities as well as data available from other sources, such as federal, state, or other local authorities and private sector organizations. The entries in this inventory may include other GIS systems within the local area from which some of the needed data may be obtained. If there exists an organized data sharing cooperative or other mechanism for government data sharing, it should be investigated at this time. There also exists the possibility that one or more of the commercial GIS database developers may be able to supply some of the needed data and should therefore be investigated. The documentation prepared at this point will be sufficient to evaluate each potential data

source for use in the GIS. Information collected at this point will also form part of the metadata for the resulting GIS database. [4]

The following table shows an example to determine data needed to support functions:

Land records	
Type of data	Data source
Parcel	Cadastral map
Easement	Land title
Parcel description	Land title
Parcel image	Field work
Name of owner	Land title
Address of owner	Land title or population registry

4.3. 2.4.3 Organize data

To interact with your GIS, you should make a top-level grouping of all the data. These groupings represent main groups such as "water utility", "land records", "streets" or "terrain".

Of the functional side these groups receive and transmit information. For example, the combination the surface model and the rainfall data gave information for hydrological drainage model. All groups should be are strictly in the same coordinate system and the topological types (network, planar or without topology).

The table below illustrates this kind of grouping:

water utility
land title
road network
terrain surface

4.4. 2.4.4 Define entities and its relationships

In this step you examine the data classification more closely. You must select an object - we call them feature - which are have common characteristics. The main steps of this process:

•identification and description of the entities • the establishment and the description of relations between these entities •documentation of the entities and relationships in the form of UML diagram

It is recommended that you document this design using commercially available graphics UML software (i.e. SparxSystems). On this diagram, you would have boxes for entities and lines for describing its relationships.

This step is important because it gives the user a detailed overview of the data and the data relations. The most important thing is that the user may be involved in this work in order to verify and validation the results. In this

---

[4]Paul Becker et al.: GIS development guide, 2006.

step, you may have to manage huge amounts of data. To divide the task into manageable units, focus on one function at a time. It may take several approximations to clarify the definition of entities and its relationships.

## 4.5. 2.4.5 Articulating entities and relationships

The individuals and their relationships can be expressed in the statements. The entities were regarded as nouns, their relationships in verbs.

Some examples:

• The shut-off valve regulates the flow of water. This statement describes an entity. • The connector assembly connects two or more of the pipeline. This statement describes the structural relationship between individuals. • The water network consists of pipes and fittings. This statement describes the aggregation of entities to make a new more complex entity. • A water main is a type of a water line. This statement describes the hierarchical grouping of individuals.

4.6. 2.4.6 Documenting entities and its relationships

A concise and understandable form for documentation of this stage the design is to preparation simple UML diagram. The following illustration shows a UML diagram of an example:
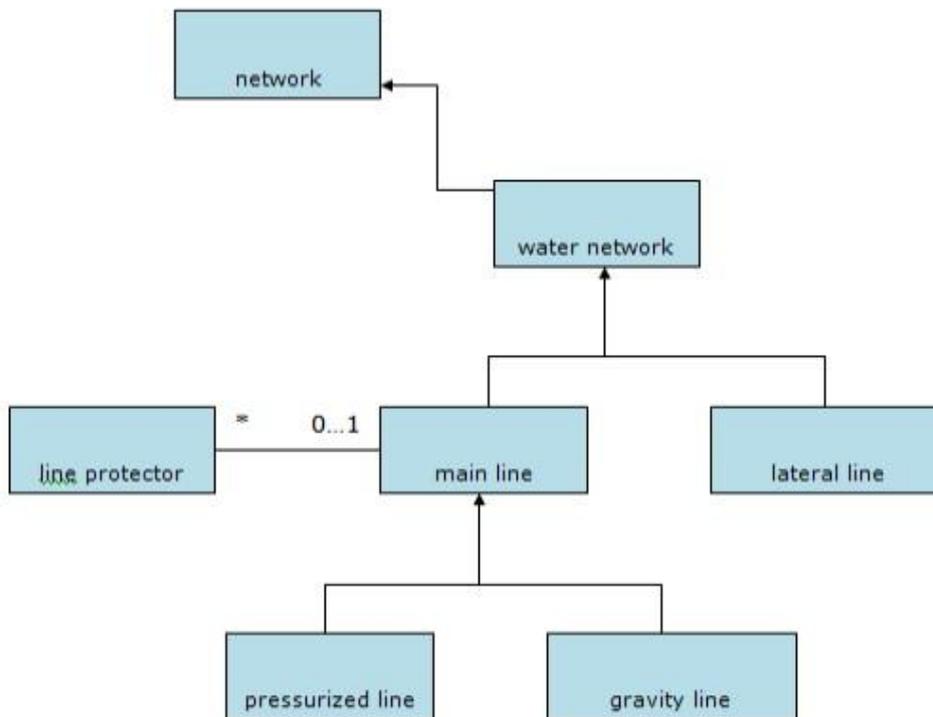
The above diagram states the following:

• A water line is a type of network line

• A main line and a lateral line make up a type of water line

• A main line can be associated with zero to many line protectors

• A pressurized and gravity main are types of main lines

The steps created so far can be summarized in the following table:

entity	relation
Water utility	
pump	-
water meter	-
meter box	water meter
valve	-
main line	--
treatment plant	-
Land title	
parcel	-
easement	-
parcel record	parcel
parcel image	-
owner	parcel
address of owner	-
Street network	
street	-
bridge	-
street name	street
traffic light	-
bus route	-
bus stop	-
Environment	
monument	-
fence	-
vegetation	-
place names	-

river valley	-
satellite image	-

4.7. 2.4.7 Identify representation of entities

In this step you classify entities by the type of its representation. Some entities will show geometric representation along with its attribute data. Other individuals will only representing by descriptive data, while others by raster, vector drawings or photos.

The following considerations should be taken:

• the feature has to appear on the map or not

• the shape of the feature might be significant in performing geographic analysis

• the feature is data that can be accessed and visualized through its relationship with another feature (for example ownership information can be accessed by selecting a parcel)

• the feature will have different representations at different map scales

• text attributes will be displayed on the screen or map products

The following rules will help in the choice of representation. The information developed in this step should be summarized in a data dictionary. This dictionary is intended to document the appearance of individuals in the geodatabase.

• Point – for a feature whose is too small to be defined as area in a database of a given resolution

• Line – for a feature whose is too narrow to be defined as area in a database of a given resolution

• Area – illustrates the location and polygonal shape of a feature on a map of a given scale

• Surface – illustrates the shape as in an area, but shows the shape resulting from changes in elevation

• Raster – represent features as an area of rectangular cells

• Image, photo, drawing – each represent a digital picture and cannot be used for spatial analysis

• Object (or Binary Large Object) – identifies a feature which are not like above, and not have geometric representation (video, voice, office document, web page)

If features could be represented in two forms (i.e. the can be a point or an area in the same geodatabase), can be identify in both forms in data dictionary. You can use this possibility to represent features in a more complex way and making more sophisticated spatial analysis.

Example of data dictionary:

entity	relation	spatial representation
Water utility		
pump	-	point
water meter	-	point
meter box	water meter	point

valve	-	point
main line	--	line
treatment plant	-	point
Land title		
parcel	-	area
easement	-	line
parcel record	parcel	text
parcel image	-	image (raster)
owner	parcel	object
address of owner	-	location
Street network		
street	-	line
bridge	-	point (line)
street name	street	text
traffic light	-	point
bus route	-	line
bus stop	-	point
Environment		
monument	-	point
fence	-	line
vegetation	-	area
place names	-	text
river valley	-	surface
satellite image	-	image (raster)

The objective of next step is to determine how data is represented in the GIS software. The focus now turns to developing an efficient and effective database. The developing team should have members who understand the geodatabase data model and analysis capabilities.

4.8. 2.4.8 Decide on geodatabase representation

We should consider the following.

If a spatial type is point:

- for an stand alone point (like historical monument) enter a point feature

- for a connected point (like water valve) enter a simple junction feature

- for a connected point with internal structure (like water treatment plant) enter a complex junction feature

If the spatial type is line:

- for an unconnected line (like fence) enter a simple line feature

- for a line that participates in a system (like road network) enter a simple edge feature

- for a line with connected sections (like water network) enter a complex feature

If the spatial representation is area:

- for a stand alone area (like the parcel) enter a polygon feature

- for a space filling area like vegetation enter a polygon feature with planar topology

If the spatial type is image (satellite image, scanned map, photograph, etc.) enter a raster type.

If the spatial feature is a surface:

- for a terrain surface enter a TIN surface representing type

- for a surfaces changing continuously (thermal surface, voice surface, spread of pollution) enter a surface

If the spatial type is an object and do not have geometric representation (like owner of properties) enter an object type.

Example to mach to geodatabase elements:

entity	relation	spatial representation	geodatabase representation
Water utility			
pump	-	point	object
water meter	-	point	point feature
meter box	water meter	point	point feature
valve	-	point	simple junction
main line	--	line	complex edge
treatment plant	-	point	complex junction
Land title			
parcel	-	area	polygon feature

easement	-	line	line feature
parcel record	parcel	text	annotation feature
parcel image	-	image (raster)	raster
owner	parcel	object	object
address of owner	-	location	address
Street network			
street	-	line	line feature
bridge	-	point (line)	point feature
street name	street	text	annotation feature
traffic light	-	point	point feature
bus route	-	line	line feature
bus stop	-	point	point feature
Environment			
monument	-	point	point feature
fence	-	line	line feature
vegetation	-	area	polygon feature
place names	-	text	annotation feature
river valley	-	surface	TIN
satellite image	-	image (raster)	raster

The last step is grouping entities to *feature classes* and subtypes, group related sets of features into *network or topology*, and organize feature classes and datasets into *geodatabase*. It is very important, that they should share a common *spatial reference*.

Literature

Everest G.C,: *Managing Data in Organizations IN Database Management Objectives,* System Functions and Administration. London, McGraw-Hill, 1986

Goodchild M.F. - Gopal S., (Eds.) : *Preface IN Accuracy of Spatial Databases,* Taylor and Francis, London, 1989

HMSO,: Report of the Committee of Enquiry, chaired by Lord Chorley, 1987

Mather P.M., : *Data Representation section 1.3 in Chapter 1 and Introduction to Computers IN Mather P.M., (1991) Computer Applications in Geography, Chichester.* 1991

Mather P.M.,: *Locational Data section 2.3 in Chapter 2 Computers and Geographical Data IN Mather P.M., (1991) Computer Applications in Geography, Chichester* 1991

Peterson, J. - Platt, J.,: *What Good are Objects to GIS Users Who Are not Programmers?* 1993

Raper, J.F. - Kelk, B., : 1991

Robinson A.H., Sale R., Morrison J.L., Muehrcke P.C. : *Elements of Cartography. (5th edition).* John Wiley & Sons, 1984

Samet H., (1989), : *Applications of Spatial Data Structures: Computer Graphics, Imaging, Processing and Other Areas. Ontario: Addison-Wesley* 1989

Michael Zeiler: *Modeling our world.* ESRI press, 2003