

Working with the Geodatabase: Scalable GIS Solutions for the Hydrographic Community

Authors: Neal Millett and Simon Evans
Environmental Systems Research Institute Inc.
Redlands, California, USA

Abstract:

This paper will provide information on how geodatabase technology can assist the hydrographic user community in maximizing the value added dimension to its digital data holdings.

The geodatabase (fundamental basis of ESRI ArcGIS software) is a storage mechanism for spatial and attribute data that contains specific storage structures for features, collections of features, attributes, relationships between attributes and relationships between features. The geodatabase has two major concepts: first, a geodatabase is a physical store of geographic information inside a database management system (DBMS); second, a geodatabase has a data model that supports transactional views of the database (versioning) that also support objects with attributes and behavior. Behavior describes how an object (feature) can be edited and displayed. Behavior includes, but is not limited to, relationships, validation rules, subtypes, and default values. With associated behaviors data entry can be regulated more efficiently, and data contamination issues can be avoided.

Some of the major benefits of the geodatabase are:

- Centralized management of a wide variety of geographic information in a DBMS
- Large data management in a continuous integrated environment (raster/vector)
- Full support for multi-user editing in an advanced versioning environment
- Support for advanced data types and geometry (i.e., the ability to store three dimensional coordinates, measures and true curves)
- Faster and more efficient data entry with data rules and relationships
- Create and edit feature-linked annotation
- Create and edit geometric networks
- Relationships with tabular data sources in any industry-standard DBMS
- Create and edit topologically integrated features
- Create subtypes and domains for maintaining database integrity
- Support for the development of industry-standard data models (e.g. S-57, VPF)

Overview

ArcGIS™ software is fundamentally based on a geodatabase (short for geographic database) data model. The geodatabase is a unique data format that is similar in structure to the coverage data model and also includes the functionality for multiuser editing. A geodatabase is a storage mechanism for spatial and attribute data that contains specific storage structures for features, collections of features, attributes, relationships between attributes, and relationships between features. The geodatabase has two major concepts: first, a geodatabase is a physical store of geographic information inside a database management system (DBMS); second, a geodatabase has a data model that supports transactional views of the database (versioning) that also supports objects with attributes and behavior. Behavior describes how an object (feature) can be edited and displayed. Behavior includes, but is not limited to, relationships, validation rules, subtypes, and default values. With associated behaviors data entry can be regulated more efficiently, and data contamination issues can be avoided.

Some of the major benefits acquired by migrating to the geodatabase are

- Centralized management of a wide variety of geographic information in a DBMS.
- Large data management in a continuous integrated environment (raster/vector).
- Full support for multiuser editing in an advanced versioning environment.
- Support for advanced data types and geometry (i.e., the ability to store three-dimensional coordinates, measures, and true curves).
- Faster and more efficient data entry with data rules and relationships.
- Create and edit feature-like annotation.
- Create and edit geometric networks.
- Relationship with tabular data sources stored in any industry-standard DBMS.
- Create and edit topologically integrated features.
- Create subtypes and domains for maintaining database integrity.
- Support for the development of industry-standard data models.

Why Use the Geodatabase?

The chief purpose of the geodatabase data model is to make it easier and more intuitive to use geographic information system (GIS) software across many applications. As such, the geodatabase is sometimes referred to as the next generation coverage. It is that and much more. The geodatabase data model centralizes data management and opens up the use of GIS to applications that were not feasible before. The geodatabase includes a variety of

sophisticated editing features that allow the GIS analysis to be more efficient and less time-consuming.

Personal and Multiuser Geodatabases

The geodatabase comes in two varieties—personal and multiuser. Personal geodatabase support is built into ArcGIS and implemented using standard desktop tools and the Microsoft Access file format for project-level GIS. Multiuser databases are deployed using ArcSDE™ software and require a DBMS such as IBM® DB2®, Informix®, Oracle®, or Microsoft® SQL Server™. ArcEditor™ 8.3 or ArcInfo™ 8.3 software can create and edit personal and multiuser geodatabases, coverages, shapefiles, CAD, and a variety of other data formats. ArcView® 8.3 software works with personal geodatabases and shapefiles.

The geodatabase, whether personal or multiuser, can store tables, feature classes, feature data sets, and functionality such as rules, relationships, and geometric networks. Using a geodatabase is a simple and straightforward process. Geodatabases can be created, accessed, and managed through the standard menus and tools in ArcCatalog™, ArcToolbox™, and ArcMap™ software applications. The geodatabase model supports intelligent features, rules and relationships that users can employ in GIS applications.

Centralized GIS Data Management

A key feature of the geodatabase is that all data (vector, raster, metadata, measures, CAD, etc.) is stored in a commercial off-the-shelf DBMS. The geodatabase is the natural solution for handling the growing amount of digital spatial data and at the same time meets the need to leverage this data among the growing numbers of GIS users. This gives organizations the power to have an integrated data management policy covering all data, which can significantly simplify support and maintenance as well as reduce costs. The geodatabase supports multiple formats of spatial data including

- Simple features such as shape files
- Custom features with business logic and editing rules
- Attribute data
- Metadata
- Images
- Raster/Grid data
- CAD data

Multiuser Editing

The geodatabase can be used for multiuser access to databases storing continuous (nontiled) spatial data through the use of versioning and long transactions. Versioning allows simultaneous editing by multiple users (subject to DBMS permissions). This framework lets the analyst create versions of a geodatabase for the stages of a project, reconcile differences between versions, and update the master version of a geodatabase with the design as-built. Versioning is not supported by the personal geodatabase.

Support for Intelligent Features, Rules, and Relationships

The geodatabase data model supports, as standard, a rich collection of objects (rows in a database table) and features (objects with geometry). The geodatabase supports advanced feature types such as geometric and logical networks, true curves, complex polylines, and user-defined features. Vector features can now have two, three, or four dimensions (x,y, z, and m). Users can define topological and associative relationships and rules that define how features

classes interact. All this can be done without programming using the menu-driven graphical user interface.

Feature-Linked Annotation

Once you have labeled your spatial features, you can then convert the labels into an annotation feature class. An option in this process is to make the annotation linked to the spatial features so that when a spatial feature is moved, the label will move with it.

Edit Topologically Integrated Features

Topologically associated feature classes in a geodatabase should be stored in an integrated feature data set. A feature data set is a collection of feature classes. The user can use the topological editing tools (such as the Shared-Edge edit tool) in ArcMap to maintain the topological associations between features in an integrated feature data set. Since it is common for vector data sets to have features that share common boundaries or corners, the ability to directly edit all of these features within one spatially common feature data set is a huge advantage for the GIS community.

For example, DNC must maintain coincidence between line and area features. The shoreline must be coincident with obstruction areas, limit areas, cliffs (occasionally), built up areas, and hydro areas to name a few. When reshaping the shoreline you want to make sure you reshape all coincident features. By creating a map topology in ArcMap, it allows you to view what's coincident in your edit area. You can then choose which feature class or features within that feature class can participate in your edit.

Integrity Constraints for More Efficient Data Entry

By managing data in a geodatabase, the user can define rules about how the data can be edited. Using the geodatabase, any user can set up subtypes and domains to ensure the attribute integrity of the database.

Subtypes

With subtypes the analyst can define categories for his/her features, which allow the analyst to distinguish objects base on different behaviors, attributes, and access privileges. For example, in DNC aids to navigation, a fixed light may have a accuracy of 1 (Accurate), 2 (Approximate) and 3 (Doubtful). The accuracy subtypes can be set up for each light so that during data entry, users can easily associate the proper attribute. Users can also set defaults for the attribute value that is most likely to be associated with lights. For example, if 90 percent of lights have an accuracy code of 1 (Accurate), having this attribute set as a default value can save time. ArcCatalog lets the user browse through easily understood descriptions of any subtypes, so the user does not have to remember cryptic coding schemes.

Domains

Domains control the allowable set of values that may be input for any field in a geodatabase table. This limits the possibility of data entry error because the choices for each data value are limited within defined parameters. For example, in DNC a sounding would have a hydrographic depth range between 0.1 and 12000.0 meters; therefore, if a value outside this range were entered it would be invalid. There are two types of attribute domains.

- Coded Values-Specify listings of acceptable codes. For example, in DNC the sounding accuracy category (ACC) could be coded as 1 (Accurate), 2 (Approximate), or 3 (Doubtful).

- Range- Specify acceptable starting and ending numeric values. For example, in DNC the hydrographic depth (HDP) could be between 0.1 and 12000.0 meters.

Geometric Networks

The movement of people, the transportation and distribution of goods and services, the delivery of resources and energy, drainage, and the communications of information all occur through definable network systems. Line and point feature classes in a single feature data set can be organized into a geometric network. The network combines the line and point feature classes into a single database entity to model linear networks-for example, undersea pipeline networks-and maintain topological relationships between the participating feature classes. The geometric network automatically maintains the explicit topological relationships between all the features that participate in a geometric network. With ArcCatalog, ArcInfo, and ArcEditor users can easily create a geometric network using the Geometric Network wizard. An entirely new and empty geometric network can be devised that a user can then populate. Alternatively, geometric networks can be assembled from existing features in a feature data set.

Sources and Sinks

Networks are often used to model real-world systems in which the direction of movement through the network is well-defined. For example, the flow of natural gas in a pipeline network is from the well location to a series of pumping stations/platforms. Flow direction in a network is calculated from a set of sources and sinks. Flow is away from sources, such as the well location, and toward sinks such as the pumping station/platform.

Dimensions

Dimensions are a special kind of map annotation that show specific lengths or distances on a map. A dimension may indicate the length of a wharf, measured distance line, or it may indicate the distance between two features such as traffic separation zones.

The geodatabase currently supports two types of dimensions: aligned and linear. Aligned dimensions run parallel to the baseline and represent the true distance between the beginning and ending dimension points. Linear dimensions do not represent the true distance between the beginning and ending dimension points, but aid in analysis and assessment of corresponding features.

Relationships

The geodatabase gives users the power to create relationships among multiple tables in the DBMS. A relationship is an association between two objects in a feature data set. These objects can be nonspatial (objects) or spatial (features). Besides identifying the associated objects, relationships can have additional properties.

Relationships are organized into relationship classes. Each relationship in a relationship class has the same origin class and destination class. Any object class may participate in many relationship classes. With relationships, the geodatabase ensures referential integrity for example a mooring trot (destination class) can consist of many mooring buoys (origin class).

Essentially, tabular relationships in a geodatabase give the users the ability to simultaneously interact with related information for their specific analysis.

Validation Rules

The geodatabase supports four broad types of validation rules for maintaining database integrity. These include attribute validation, network connectivity rules, relationship rules, and common clusters.

Validation rules can ensure that the data entry operator does not enter invalid entries for the data in question, for example a starboard buoy has to have an attribute value of GREEN. When the data entry changes values for an attribute, users might want to make sure that only legal or valid values are inserted into the attributes.

Attribution Rules

Attribute validation rules include subtypes for assigning data categorically as well as attribute domains for constraining the values allowed in any particular attribute for a table, feature class, or subtype. Essentially, attribute domains constrain the values that can be applied to the data so that the data entry operator will be less likely to enter incorrect values. Subtypes make it easier for the operator to assign the data into easily understandable categories for example there may be several subtypes of beacons defined within the beacon class.

Connectivity Rules

Network connectivity rules constrain the type of network features that can be connected to one another and the number of features of any particular type that can be connected to features of another type.

There are two types of connectivity rules: edge-junction rules and edge-edge rules. An edge-junction rule is a connectivity rule that establishes that an edge of type A may connect to a junction type B. An edge-edge rule is a connectivity rule that establishes that an edge of type A may connect to an edge of type B through a set of junction types.

By establishing these rules, along with other rules such as attribute domains, the analyst can maintain the integrity of the network data in the database.

Relationship Rules

Relationship classes can have an associated set of relationship rules. Relationship rules control which object subtypes from the origin class can be related to which object subtypes in the destination class. They can also be used to specify a valid cardinality range for all permissible subtype pairs. Cardinality describes how many objects of type A are related to an object of type B. For example, the subtype wood pole may be able to support from zero to three transformers, whereas the subtype steel pole may support zero to five transformers. In the first case, the cardinality range would be 0-3; in the second case, it would be 0-5. In ENC you might want to build a relationship between buoys and topmark features. If you delete the buoy the topmark must be removed as well.

There are dependency or condition rules as well. In DNC a rock can be awash, submerged or covers and uncovers. In the submerged rock case, the hydrographic drying height would be null and hydrographic depth would be 0 (Unknown) or the actual value to the nearest 0.1 meter (valid range 0.1 to 12000.0). In the covers and uncovers case, the hydrographic drying height could be 0 (Unknown) or the actual value to the nearest 0.1 meter (valid range 0.1 to 1000.0) and the hydrographic depth would be null.

