IMPLEMENTING A GEODATABASE FOR WATER DISTRIBUTION AND WASTEWATER COLLECTION SYSTEMS: STRATEGIES AND BENEFITS

Bruce B. Taylor

AUTHOR: Engineering Services Supervisor, Clayton County Water Authority, 1600 Battle Creek Road, Morrow, GA 30260.
REFERENCE: Proceedings of the 2003 Georgia Water Resources Conference, held April 23-24, 2003, at the University of Georgia. Kathryn J. Hatcher, editor, Institute of Ecology, The University of Georgia, Athens, Georgia.

Abstract. Implementing a geodatabase involves a significant investment of resources in software, training, and time. This investment can yield substantial benefits when the reasons for adopting the geodatabase are clearly defined, and the design process is carefully planned. These benefits include improved data quality and integrity, enhanced feature editing, better performance, the ability to use "smart" features, and the use of industry-standard relational databases. This paper will discuss the trade-offs involved in implementing a geodatabase—a database that stores both geometry and attributes—for water distribution and wastewater collection systems, arguing that most systems should consider the geodatabase as an excellent option, since it provides managers with improved data storage, retrieval, and security. It will then discuss implementation strategies and give specific examples of how to make a successful transition from shapefiles, coverages and other data formats to the geodatabase.

INTRODUCTION

A Geographic Information System (GIS) combines a map with a database, relating a geographic feature with a record in a data table containing the feature's attributes. Water and wastewater utilities use GIS to map their pipes and other buried assets, to manage these assets by associating maintenance data with them, and to model the behavior of water and wastewater flow through the networks. While GIS technology was originally developed several decades ago, GIS data was stored in files or a combination of files and databases. Recently, however, several software vendors have developed a data storage method known as a geodatabase. As implied by its name, a geodatabase is a geographic database. A geodatabase uses database tables to store not just the GIS attributes, but also the geometry. The insights provided in this paper have been tempered by the author's experience implementing a water and wastewater geodatabase in one of Georgia's largest utilities. Clayton County Water Authority stores all of its corporate water and wastewater infrastructure data as well as its extensive land application system data in a geodatabase. Most water and wastewater would be wise to adopt the geodatabase as it provides many benefits in the storage, retrieval, and security of valuable corporate asset data.

GEODATABASE BENEFITS

Many benefits flow from this new way of organizing GIS data. One of the most important benefits is maintaining improved data quality and integrity while updating and editing data. A geodatabase allows the designer to impose constraints on both the geometry and the attributes, limiting the data that is added to the geodatabase to intended values. Geometric constraints allow the designer to control how features relate to each other geographically. A fire hydrant can only be connected to an 8-inch water line, for example. Attribute constraints allow the designer to control what type of data can be entered into a field. Through attribute domains, the designer can limit data to a range of acceptable values or to a list of predefined values. A sewer manhole can have a depth of no more than 100 feet, for example, or a valve can be manufactured by one of three companies. In addition to these constraints, most geodatabases require that the designer define the coordinate system and geographic precision before features are added. This requirement ensures that only geographically similar data is combined in the geodatabase, and eliminates potential errors that can arise from having to define the coordinate system and precision for each dataset.

Geodatabases allow GIS to take advantage of the broad range of excellent commercial databases that are currently on the market, since geodatabases are stored in industry standard databases such as ORACLE, SQL Server, and MS Access. These databases provide tools for data backup and recovery, rapid data storage and retrieval, indexing, and compression. By storing a geodatabase in one of these databases, GIS vendors can take advantage of the high performance of the commercial databases, and focus their efforts on data display and analysis. Many of the database vendors are currently developing native geographic data types, which will further enhance their performance. Additionally, most of these databases have been developed for large corporate datasets, and have ample capacity to store and query large geographic datasets, thus eliminating the need to divide GIS datasets into tiles as was once the case.

Storing geographic data in databases compares favorably to file-based storage. Files are vulnerable to inadvertent editing and corruption. Databases can be password protected and limit editing through record locking. Storing both geographic and attribute data in tables is inherently a more standardized way of storing data, and represents a move away from vendor-specific data formats. Some geodatabases also allow multi-user editing, which allows more than one user to edit a feature layer at the same time. Multi-user editing eliminates many of the maintenance and data-corruption issues that have traditionally plagued GIS editing. Additionally, since many vendors provide the ability to store data in a portable database such as MS Access, a geodatabase allows users to store their data in one database and then move or copy that database from one place to another without fear of losing or corrupting files.

Geodatabases allow for a more robust and intuitive graphical environment. File-based geographic data have in the past been limited to points, lines and polygons. Geodatabases have added circles, arcs and parametric curves as well as multi-point features. These additional geometries make the editing process more efficient, and allow for better representation of many real-world features. Geodatabases also allow for a more intelligent editing environment by allowing features to be represented as objects, not just as geometric shapes. Thus a layer representing valves can be subdivided into valves that perform specific functions, and the editor can easily differentiate them and use them only in appropriate situations.

TRADE-OFFS

As with any new technology, there are trade-offs associated with adopting a geodatabase. To take advantage of the geometry and attribute constraints, more time must be invested up front to design and implement them. This additional design and implementation time is not limited to constraints, particularly with a geodatabase stored in a large-scale database such as ORACLE, SQL Server or DB2. These larger implementations require considerable up front design time as well as knowledge of relational database management, plus time and effort to set up the relational database.

In most cases software must be upgraded to interface with a geodatabase. Both front-end GIS packages such as GeoMedia or ArcGIS and back end packages such as SDE must be purchased. (The back end packages are only necessary with some large-scale implementations). These upgrades, in turn, require training, process changes, and rewriting of scripts and interfaces to other corporate software. In addition, there may be compatibility issues with users of existing software. Though many GIS packages provide the capability to export data from a geodatabase, few export routines can be completely automated, and many require extra time and user intervention to complete the export process.

In a few cases, a geodatabase may actually provide lower performance than file-based geographic data sets. Although uncommon, this limitation highlights the fact that the geodatabase is an emerging technology. While geodatabases are evolving rapidly, the technology has not completely stabilized. And while third-party software vendors are rapidly adopting the geodatabase model, not all have made the conversion, and some software tools and utilities may not yet be available.

IMPLEMENTATION

All software and data setups are different, and there can be great differences depending on the size of the organization and the nature of the data. Despite these differences, all geodatabase implementations should contain the following steps: pre-planning, planning and the actual implementation. For small implementations, for example, the planning stage may simply involve an hour-long meeting with two or three individuals. For large ones, a week-long geodatabase modeling meeting may be required.

In pre-planning, the implementers and the end-users should answer several questions. Why are we adapting the geodatabase? Can we take advantage of the benefits discussed above? How do any of the trade-offs apply to us? Do we just want to stay “cutting edge” (technology for technology’s sake)? Is the timing right relative to other software upgrades? Is the current system somehow deficient? What is the size of the organization and staff? What is the size of the datasets to be converted (overall number of features, number of feature types, number of layers)? Answering these questions should give a good indication of when and how to proceed with the implementation.

In planning the implementation, implementers and end-users should plan for and schedule software and hardware upgrades, schedule implementation team and end-user training, and perform the data design. This last step is the most important: once a geodatabase design is implemented it is difficult to redesign, modify or undo. A geodatabase design workshop should be held, preferably in a location without chance for interruption. During the workshop all data to be incorporated in the geodatabase must be written out (use paper at this point in the process rather than the computer), including data types, sub-types, relationships, and domains. Existing geodatabase design templates should be evaluated, as well as existing data structures. While the functionality of the geodatabase model can be alluring, designers should always keep in mind the time and effort required to implement and maintain this functionality. Often less is more. Extreme care should be exercised when developing ID schemes. Designers should consider how the IDs will be used to link tables, how they will be updated, and how they may be affected by importing and exporting data.

Implementing the geodatabase can be quite simple for small (few feature types) datasets, and quite involved for large or complex datasets. Implementers can use database design software such as MS VISIO to develop the final schema and export it to the geodatabase, or they can use native GIS tools (e.g., ArcCatalog) to develop the schema. Once the schema is finalized, the geodatabase tables are created and the data is loaded. Loading data can be challenging, depending on the quality of the original data, and how closely it "maps" to the tables and fields in the geodatabase. Often the data will need to be modified, and sometimes individual data cells must be updated by keying in new values. Backups should be made at each stage, when the source data is modified. If at all possible the source data should be put on editing hold during the implementation.

CONCLUSIONS

The Clayton County Water Authority has successfully implemented the geodatabase model and is enjoying the many benefits listed above. Hiring a consultant to lead the geodatabase design workshop ensured a well-designed and easily maintained geodatabase, while doing the implementation in-house allowed us to learn the details of the system.

For most, adopting the geodatabase will be less a question of "if" and more a question of "when." With its many benefits and its adoption by the major GIS vendors, the geodatabase will become the standard for GIS data storage. By considering the trade-offs and doing other pre-planning activities, the transition can be smooth and the results rewarding. Organizations that currently store a significant amount of data in other GIS formats, may reasonably decide to delay geodatabase implementation until the trade-offs can be minimized. For organizations without much invested in GIS data or software, the benefits of adopting the geodatabase should far outweigh these trade-offs. The water utility industry is one where template geodatabase schema are quite evolved, and the major software vendors, as well as third-party vendors, have made significant progress with front-end GIS tools and utilities. Since many utilities in Georgia have not yet developed mature GIS systems, they would be particularly able to capitalize on geodatabase technology.