Using a Geodatabase to Improve Water Quality Assessment Data Management

By Jon Becker, US Environmental Protection Agency, Region 4
June 13, 2008

Abstract

The U.S. Environmental Protection Agency, Region 4, Water Management Division (WMD) has developed a geodatabase to manage the georeferenced water quality assessment data from the region's eight southeastern states for various years. The geodatabase's relationships between various assessment unit features and detailed tables provide WMD GIS staff with powerful and versatile analytical capability. By integrating disparate state data formats into a common structure, the geodatabase enables much easier multi-year, multi-state spatial analysis than before. Answers to complex management requests or environmental progress reports can be generated in a fraction of the time than when the georeferencing was contained in various unmatching shapefiles.

Overview

In accordance with the Clean Water Act, each state environmental agency tracks the water quality status of waterbodies in their states. Every even-numbered year an "assessment" is made of each waterbody based on whether the waterbody is meeting its designated uses. Each state agency tracks the status of its waterbodies in a local database and generates GIS datasets depicting the locations of these waters. Each lake, estuary or reach of a river is called an "assessment unit" (AU) and given a unique AU identifier. While EPA offers recommended formats for assessment data and the complimentary GIS data, states often deviate somewhat from these formats to meet their own needs.

States submit their assessment data and GIS files to EPA Region 4. These files are then forwarded on to EPA headquarters for eventual input into the national Assessment and TMDL Tracking And ImplementatioN System (ATTAINS). The GIS files are "reach-addressed" to the 1:100,000 scale National Hydrography Dataset (NHD), and incorporated into EPA's Watershed Assessment, Tracking & Environmental Results system which integrates various EPA water-related databases via reach addresses of the NHD.

Data Management Challenge

Over the years EPA Region 4 personnel have obtained numerous assessment databases and GIS datasets from its eight southeastern states. These datasets are in various incompatible formats. Thus, any sort of spatial multi-state or multi-year analysis of waterbody status has proven difficult and time-consuming, even for experienced GIS users, much less GIS novices. For example, in Figure 1 below, you will see the tabular structure of Tennessee's 2006 data (at top) is in a much different format than Georgia's 2004 data (bottom). In fact, Tennessee's assessment status for each AU is not even given in the GIS shapefile, but can only be determined through relating the GIS shapefile to a separate table in a one-to-many relationship. On the other hand, Georgia identifies violations for each AU in a field called "Violations," using abbreviated codes (e.g. "FC" for fecal coliforms).

Figure 1
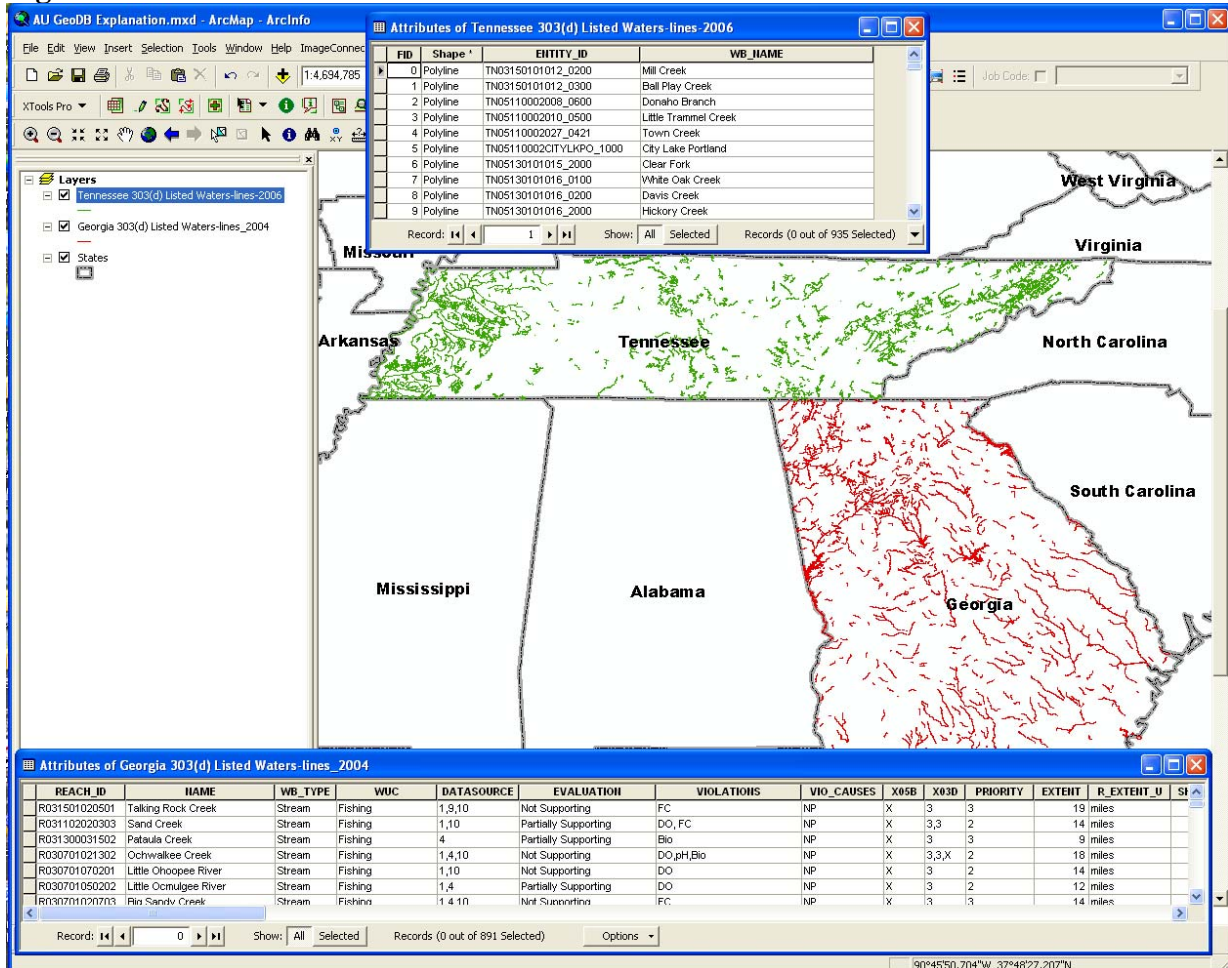
Geodatabase Approach

In 2007, EPA Region 4 staff explored the possibility of using a geodatabase to better manage assessment data and the GIS features rather than assorted shapefiles, tables and Access databases. However, the obvious roadblock was the varying formats of the datasets. As mentioned earlier, state assessment and GIS data are eventually reformatted by EPA headquarters and its contractors into consistent formats in the national ATTAINS and WATERS systems respectively. Region 4 contracted Research Triangle Institute to develop an initial prototype of the geodatabase, populating it with reach-addressed version of the states' assessment datasets. Region 4 staff then generated tables detailing the status of each AU from ATTAINS (then called the National Total Maximum Daily Load Tracking System). Relationship classes were created to relate the point, line and polygon AU features to the assessment status tables using the AU identifier field. This was a one-to-many relationship because the AU may sometimes be listed multiple times in the assessment status table if it has been tracked for multiple cycles and/or multiple water quality standard impairments (e.g. failure to meet swimming designated use because of high levels of pathogens).

Although this geodatabase format functioned fine, Region 4 staff soon became aware that the reach-addressed versions of the state AU GIS features might be insufficient in some cases. Sometimes the reach-addressed versions of the GIS features did not match up adequately with the original state GIS shapefiles, at least for our purposes. In some cases the state's approach to identifying an AU didn't match well with the NHD framework. For example, because Florida has mostly flat topography and complex and highly modified hydrography not depicted well at the NHD scale, Florida Department of Environmental Protection prefers to use its own Waterbody Identifiers (WBIDs) which depict a small contributing drainage areas. Because Region 4 staff often deal directly with state agency staff in discussing these AUs, it was important for our purposes to populate the geodatabase with the state's GIS features of the AUs rather than the reach-addressed versions used in EPA WATERS. That said, Region 4 recognizes the value of having all of the state assessment GIS data in a consistent, NHD-based format, and continues to encourage its states to adopt the 1:100,000 NHD as the framework for their AUs. Most of our states are now using or plan to use NHD or a modified version of it to represent their AUs so hopefully this will not be much of an issue soon.

Because of the Region's desire to populate the geodatabase with GIS features from the states, the tables from each state GIS dataset had to be reformatted to a consistent format. The essential component for each AU record was an AU identifier, since that was used to relate back to the ATTAINS tables. Features were also dissolved by the AU identifier so that there was only one feature per AU. Waterbody name and type fields about the AU were populated if the information was readily available in the state GIS dataset, and a hyperlink field provided a link to the EPA webpage about that waterbody. Length in miles was calculated for linear features and area in acres for polygon features.

The assessment data feature classes and tables and relationship classes have been loaded into the Region's ArcSDE server. This has provided a stable, fast-drawing platform. By having all the data for different states and cycles in one feature dataset, that allows us to build various layer files pointing to these feature classes which are just different definition queries. A layer depicting locations of approved TMDLs was also developed. By using the Identify button in ArcGIS or ArcReader, a user can click on an impaired water and see it's impairments for all cycles and what TMDLs have been established for it. (Figure 2). The built-in relationships give users the ability to query 303(d) listings or approved TMDLs by certain parameters, such as all pathogen impaired waters, and then activate the relationship classes to display those features in the GIS.

Figure 2
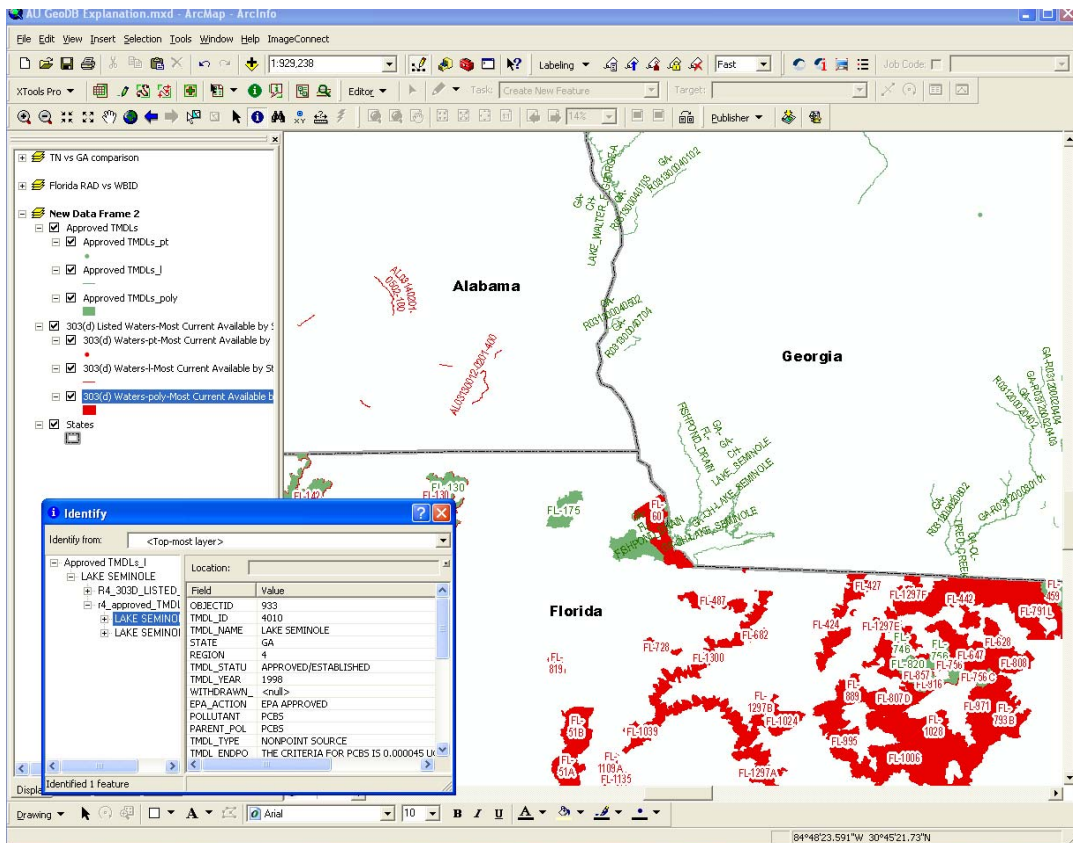
Now when new GIS data files arrive from the states, they are modified to fit the table format and loaded into the appropriate feature class. The tables from ATTAINS with the waterbody assessment status are periodically updated. Serving the data via ArcSDE also gives the Region the ability to include these layers in ArcReader projects or other applications.

Reformatting the historical GIS files into this new consistent tabular framework involved a lot of work, but the resulting product has proven well worth the effort. GIS users can perform complex queries in the related tables and then activate the related features, or simply use the Identify tool to click on an AU feature and see the related tabular information there. Because of the power of the relationship classes linking the ATTAINS tables of impaired waters and approved total maximum daily loads (TMDLs), we now have a much more powerful and versatile analytical dataset that can generate useful information in a fraction of the time than before when we had numerous separate files of varying formats.

If you wish to receive a copy of the Powerpoint presentation given at the conference (which will include more graphics and detail than this paper version), please contact Jon Becker of US EPA Region 4, Watershed Management Office at becker.jon@epa.gov, 404-562-9927.