DRAFT

EWEB Drinking Water Source Protection Program

ArcHydro GeoDatabase Migration Assessment

Prepared for Eugene Water and Electric Board

By Bob DenOuden
Lane Council of Governments

June 2004

TABLE OF CONTENTS

Introduction
EWEB's Drinking Water Source Protection Program (DWSPP) for the McKenzie River watershed includes an extensive data collection effort and analysis component. In 2002 LCOG, under contract to EWEB, designed and implemented a data base (SQL Server) to store water quality monitoring data generated be the DWSPP. In addition to data reporting and analysis of the tabular data housed in this database, EWEB's use of GIS for mapping and analysis of monitoring data has grown. To date, EWEB's use of GIS for interacting with the monitoring data has been to produce relatively simple maps showing various analyte concentrations in the basin. Future uses of GIS include pollution load analysis, network tracing to investigate pollution sources, and more advanced spatial analysis of the basin. In order to help facilitate these efforts EWEB contracted with LCOG to develop recommendations for integrating the monitoring database with EWEB's GIS tools. This report contains these recommendations.

Geodatabase Overview

What is a Geodatabase?
The geodatabase is an object-oriented data model introduced by ESRI that represents geographic features and attributes as objects and the relationships between objects, but is hosted inside a relational database management system. A geodatabase can store objects, such as feature classes, feature datasets, non-spatial tables, and relationship classes. Basically, 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. Geodatabases come in two varieties--personal and multi-user (or enterprise). Personal database support is built into ArcGIS implemented with Microsoft Jet and is suitable for project-level GIS. Multi-user databases are deployed using ArcSDE and require a RDBMS such as IBM DB2, Informix, Oracle, or Microsoft SQL Server.

Enterprise vs Personal Geodatabase
To the GIS user, the functionality of an enterprise geodatabase and a personal geodatabase are very similar. The primary difference to the user is how they connect to the data. The differences in how data are managed and stored are very different, however. An enterprise geodatabase, implemented through ArcSDE, offers the same benefits for spatial data that an RDBMS does for non-spatial data. GIS users employ the enterprise version of the geodatabase to provide efficient access to large databases by multiple concurrent users. The primary benefits offered by the enterprise geodatabase, implemented with ArcSDE and a RDBMS are:

- Multi-user editing
- Support for versioning
- Support for disconnected editing
- Virtually unlimited database size
- Robust RDBMS environment (security, backup and recovery, scalability)
- Support for raster data storage in the database, including compression

To the database administrator, of course, the differences are greater. SDE administration is a complex task requiring specialized knowledge analogous to the requirements imposed by SQL Server or Oracle when compared to an Access database.

Many users are confused by the role of ArcSDE in an enterprise geodatabase installation. ArcSDE is not a database, but a server software product used to access large multi-user geographic databases stored in relational database management systems (RDBMSs). It is an integrated part of ArcGIS and a core element of an enterprise GIS solution. Its primary role is to act as the GIS gateway to spatial data stored in the RDBMS.

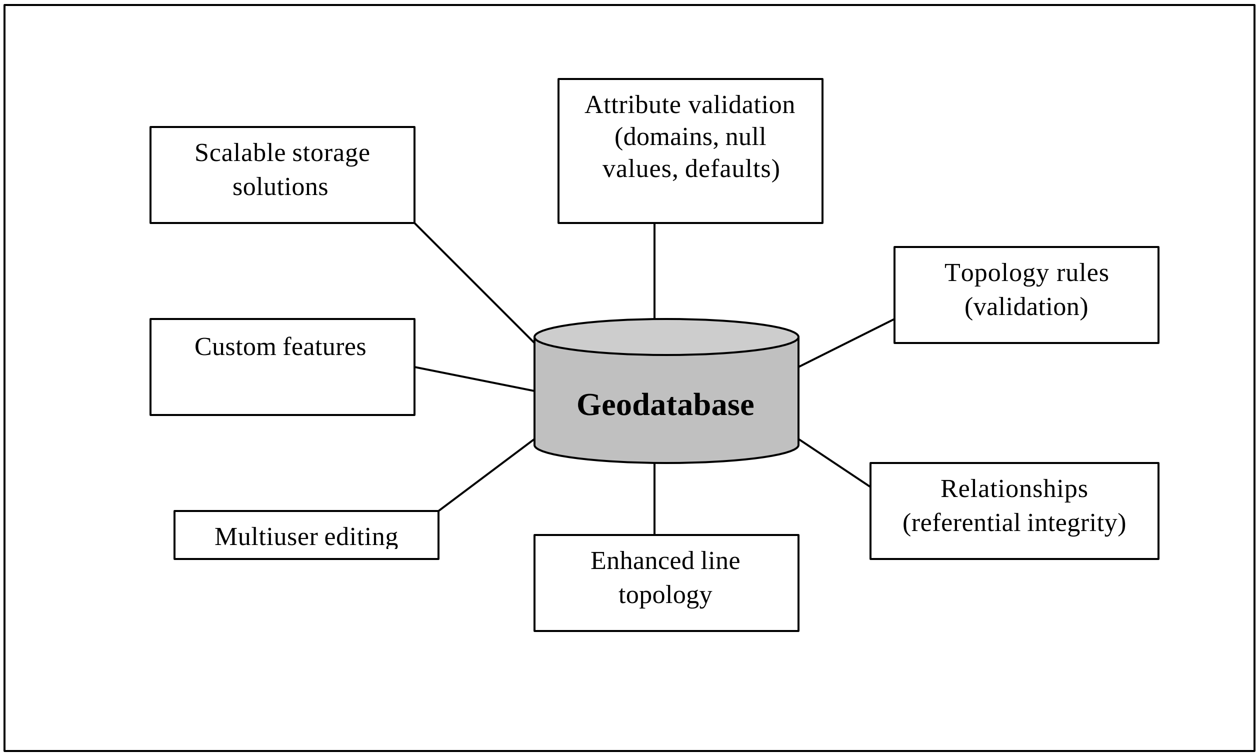
The ArcHydro data model can be implemented and deployed as either a personal or an enterprise geodatabase. The personal geodatabase implementation should provide a good place to begin with for the EWEB drinking water source protection program. If future needs dictate, it can be migrated to an enterprise geodatabase.

Benefits of the Geodatabase

Storing spatial and attribute data directly in a commercial database gives a geodatabase capabilities that are not available, or are more difficult to achieve, with other formats. Some of these benefits are listed below.

- A uniform repository for geographic data. All geographic data is centrally stored and managed in one database.
- Data entry and editing is more efficient. The use of subtypes, domains, and validation rules helps maintain database integrity and reduces database maintenance. For many users, this reason alone justifies adopting the geodatabase model.
- Sets of features are continuous. Geodatabases can accommodate very large sets of features without tiles or other spatial partitions.
- Multi-user editing. Enterprise geodatabases accessed through ArcSDE use a data management framework called versioning that lets multiple users access and edit features simultaneously and reconciles any conflicts.
- Topology rules can be defined to provide spatial data behavior that is closer to real world models.
- Feature-linked annotation. Geodatabase annotation can be linked to the feature that it describes. When the linked feature is moved or deleted, the related label is moved or deleted.
- Users work with more intuitive data objects. A properly designed geodatabase contains data objects that correspond to the user's model of data. Instead of generic points, lines, and areas, users work with objects of interest such as transformers, roads, and lakes.
- Much greater levels of customization are available in the geodatabase than in other data platforms.

A summary of the advantages of the Geodatabase is shown in the diagram below:



ESRI Geodatabase Data Models

Another benefit of the geodatabase is the ability to leverage ESRI's (and their partners') efforts in designing and implementing industry specific ArcGIS data models. The goal for the ArcGIS data models are to provide a practical templates for implementing GIS projects. The models provide a starting point for work on specific project data models. One of these data models is the ArcHydro data model, a specific design created by ESRI and the GIS Water Resources Consortium, headed by Dr. David R. Maidment, Director of the University of Texas' Center for Research in Water Resources. This data model has been in the works for three years and seems mature enough to build from.

ArcHydro Data Model and Tools

The ArcGIS Hydro Data Model (Arc Hydro) is an ArcGIS geodatabase model. It provides a standardized framework into which various types of water resources data can be loaded. In this manner the data forms an integrated water resources modeling and mapping database. Along with the data model, ArcHydro provides a set of tools designed to build the initial *hydrological information system* data sets. These tools provide functionality to process digital elevation models to produce synthetic stream networks in a geometric network that conforms to the ArcHydro data model specifications. There are also tools to delineate watersheds and catchments, load hydro point data, and perform some simple network tracing and other basic analysis.

The Data model

The ArcGIS Hydro data model describes geospatial and temporal data on surface water resource features of the landscape. The data model addresses three issues:

- **Hydro Description –** What are the principal water resource features of the landscape?
- **Hydro Connectivity –** How does water move from feature to feature?
- **Hydro Modeling –** What are the time patterns of water flow and water quality associated with these features?

The data model is organized into feature datasets comprised of similar data elements. These include:

Hydrography: map hydrography, the "blue lines" on the map as well as various physical point data locations within the watershed.

Drainage: drainage areas derived from digital elevation models or manually digitized.

Channel: 3-D profile and cross-section representation of stream channels.

Network: a geometric network representation of the connectivity of the surface water features of the landscape, including hydrojunctions and hydroedges.

Time Series: Data collected within the watershed that fits a time series structure.

Associated with these five feature datasets are a set of object tables for additional information, such as events defined on the river network, and time series of monitoring data. The data model components themselves are fairly generic and most real world applications of GIS to hydrological systems should be able to take advantage of the basic building blocks defined by the ArcHydro data model.

ArcHydro Framework: The ArcHydro Framework is a subset of the full ArcHydro data model intended to implement the most common components of the full data model but with less complexity. The general recommendation for most agencies using the model is to start with the framework components and add the others as necessary. For EWEB it appears that the framework, plus time series, would be adequate for their needs. The exception to this may be the BASINS model requirements. Until more is known of what the ArcGIS version of BASINS will require, it makes sense to hold off on this decision. Alternately, one can start with the full model and eliminate the feature classes that are unnecessary.

The Tools

The ArcHydro toolset, which will be covered in more detail later in this document, includes the functionality for the following tasks:

- Terrain Processing
- Stream definition
- Catchment delineations
- Drainage point processing
- Slope calculations
- Watershed characterization
- Flow path calculation
- Hydrological network generation
- Calculating downstream length
- Find next feature
- Assigning HydroID

- Stream network tracing

By running through these tools in a defined sequence, the result will be a populated ArcHydro data model. The recommendation, from ESRI is to use these tools to construct and populate the initial ArcHydro geodatabase then modify it to meet local conditions and needs.

Benefits of the data model

The ArcHydro data model provides a standardized way to store hydrologic data. One of the primary benefits of the ArcHydro data model is that it comes with a suite of data processing tools designed to populate it, providing a starting point for further refinements. In addition, the data model provides a data standard that others have adopted, including the Environmental Protection Agency for it's BASINS (Better Assessment Integrating Point and Non-point Sources) watershed assessment model. The next version of BASINS, due in summer of 2004, will be built on the ArcHydro data model.

Drinking Water Quality Monitoring Database

EWEB's Drinking Water Source Protection Program for the McKenzie River watershed is supported by an extensive water quality monitoring data collection effort. In order to help facilitate these efforts, EWEB contracted with LCOG to design and implement a database solution to function as a repository for these data. The result of this project was an SQL Server database housing numerous disparate data sources in a common data structure that can be queried and analyzed through database front-end applications such as Microsoft Access or SQL Query Analyzer or, more importantly for this effort, through GIS tools.

Establishing a Linkage between the Geodatabase and the Monitoring Database

There are a number of ways in which tabular data (such as that contained within the drinking water monitoring database) can be linked to spatial data within a geodatabase. The basic options are:

- Relate or join spatial data to tables within ArcMap
- Establish a *relationship class* within the geodatabase to link feature class to object class (table)
- For point data, carry the X and Y coordinate values as attributes within the SQL Server database and add to ArcMap as point events
- Create static .dbf files and add them to an ArcMap document (then join or relate to feature class)
- Join or relate feature class to an Access table or query

The simple join or relate is the recommended (by ESRI) option for integrating tabular data sources with spatial data, primarily because relationship classes are "expensive" in the database and expose a great deal of additional functionality that might not be needed. One benefit of the relationship class is that it is persistent within the database, making it easy to establish linkages within ArcMap (though joins and relates can also be "saved" into a layer file to provide similar functionality. For monitoring site data, having the coordinate values stored in the data table in SQL Server makes it very convenient to work with the data in ArcMap, particularly for symbolizing data. A potential issue to examine is that of using the ArcHydro tools, which rely on a relationship to a Hydro Junction in order to take advantage of analysis tools. Spatial data that is brought in to the geodatabase without the use of the ArcHydro toolset will require additional processing to establish the correct HydroIDs and relationships to the data model.

Drinking Water Quality Monitoring Mapping Needs Assessment

LCOG has been producing a series of maps showing various summarizations of water quality monitoring data for baseline conditions and storm events. Based on these maps and an iterative process of defining the mapping needs for water quality monitoring reporting, we now have a better idea of these requirements. In addition to specific mapping requirements, EWEB also needs to have the ability to interface with the monitoring data through their GIS tools in a more ad hoc manner.

Specific revealed mapping needs thus far include:

- Average baseline values for selected analytes at monitoring points
- Differences between storm event monitored value and average baseline measure in magnitude as well as percentage.
- Storm event measured value for selected analytes at monitoring points
- Benchmark value exceedance for selected analytes for selected storm events as well as baseline values.

Once a full list of requirements can be developed, additions to the SQL Server data model can be implemented, in the form of additional tables and/or views, in order to support these needs. The combination of ArcHydro's personal geodatabase data model, the ArcHydro tool set, and links to the SQL Server tables that comprise the water quality monitoring database shows great promise for EWEB in implementing a "hydrological information system" for the McKenzie watershed in support of drinking water quality protection.

Migration to ArcHydro

As previously mentioned, ArcHydro consists of a data model as well as a set of tools designed to populate the data model and access the resulting data sets. These tools will make it much easier to implement the data model. The migration process will entail using ArcHydro's tool set to build the initial geodatabase, then incorporating elements of the monitoring database into this geodatabase as either linked views or related tables. This process is described in more detail below:

Building a Geodatabase

The ArcHydro data model consists of five primary groupings of data (organized as feature datasets in the ArcHydro geodatabase). These are:

- Drainage: Drainage areas and stream lines (the "blue lines" on the map)
- Hydrography: The base data from topographic maps and tabular data inventories (including monitoring points and other user defined points).
- Network: Connected sets of points and lines showing pathways of water flow. This will be created from elevation data (DEMs) using the ArcHydro tools.
- Channel: A 3-D representation of the shape of stream channels.
- Time Series: Tabular attribute data describing time varying water properties for any hydro feature

Drainage feature sets: Catchments, watershed, drainage points, drainage lines, and drainage areas are all created as a output from the ArcHydro tools. In order to create the complex integrations between feature sets (linked by HydroID) it is recommended that the tools are used to populate the initial geodatabase. Preliminary testing using these tools indicates that this will be feasible. A single DEM spanning the McKenzie watershed is necessary. This exists for the 30 meter data, but it is not yet clear whether this is the case for the 10 meter DEMs (which might be desirable for stream definition). If 10 meter data is desired more work will be required to assemble the data at a watershed sale and mosaic it into a single grid.

Hydrography: Loading EWEB's monitoring site point data into the MonitoringPoint data structure is straight-forward. It is recommended to use the item EWEB_SITE_ID as the HydroCode item in the ArcHydro data model. The model was designed to contain an internal integer ID (HydroID) as well as an "external" text ID (HydroCode) which nicely matches our current design. Other Hydrography features will include the 1:24,000 scale hydrography data (lines and polygons) that already exist – though the Linn County data should be appended to the Lane County data and additional attribution applied. Additional work in preparing a hydrography feature dataset includes the merging of editing done on the regional 1:24,000 hydro layer (Countywide) with the edits done by EWEB on the McKenzie watershed 1:24,000 hydro data in order to create a seamless countywide hydrography layer that EWEB and other could use.

Network: The network feature sets form a geometric network that allows modeling, tracing, and other hydrologic/hydraulic analysis. The process of creating a geometric network within ArcCatalog is fairly straight-forward but the specific structure of that network required by the ArcHydro tools makes it a more demanding task. These feature sets are best created through the ArcHydro tool set. As noted above, 10 meter DEMs might be needed for defining the synthetic stream network to a sufficient level of detail for EWEB's modeling needs. More research and experimentation is needed. The tools were used to create a network from the 30 meter DEM data and this worked well.
There are potential issues with scale and accuracy of the synthetic stream network. The DEMs will not lend themselves to creating a structure that includes the engineered channels that EWEB maintains (diversions, canals, etc…) nor will it likely produce an accurate representation within the more urbanized areas of east Springfield. This will have to be further investigated.

Channel: Initially, it is assumed that these feature classes would not be necessary for EWEB's ArcHydro geodatabase.

Time Series: The time series dataset consists of two object classes (database tables) that correspond fairly closely to the MEASUREMENTS table in the monitoring database. Initial tests have shown that it is fairly easy to populate these tables from the existing table structure, though some research will be needed in order to actually utilize the time series viewer in ArcMap. According to Dean Djokic, ArcMap 9.1 will add a significant update in time series data management functionality. The time series functionality provided by ArcHydro Tools is limited so it makes sense to investigate the new ArcMap functionality that is forthcoming. Also, according to Dean Djokic, the company, DHI, has been developing a time series manager that will be included in the Arc Hydro when it is done (the "lite" version for free, the full version for a fee). This promises to provide better functionality for displaying and analyzing the time series data stored in the water quality monitoring database.

Loading data into the ArcHydro Geodatabase

The ArcHydro toolbar lays out a straight forward set of steps designed to produce drainage area and flow network features, all to the ArcHydro data specification. An initial experimentation with the process proved successful once the database and ArcHydro tools were set up. It is necessary to turn on the Spatial Analyst extension before running the terrain processing commands. It is also necessary to install MSXML 4.0 before installing the tools. ArcHydro itself comes with a SETUP.EXE that makes installation simple. Once installed, the commands are organized on an ArcGIS toolbar into functional areas; terrain processing, watershed processing, attribute tools, and network tools.

Terrain Processing

The terrain processing commands include the following (in order of operation)
DEM reconditioning: This process matches a vector streams representation onto a DEM in order to improve the accuracy of the synthetic streams created later. The streams need to be single line (flow line) representations (as opposed to shorelines and double bank streams).
Fill Sinks: "Raises" elevation of cells that are lower than all of their neighbors until a flow direction from that cell can be determined.
Flow Direction: Sets direction of flow for each cell using the D-8 methodology (using the steepest slope between each cell and it's eight neighboring cells.
Flow Accumulation: Calculates total flow accumulated by counting all upstream cells that drain through each cell.
Stream Definition: Creates a stream course grid.
Stream Segmentation: Attributes the stream grid into sections.
Catchment Grid Delineation: Assigns catchment attribute to each cell in the grid.
Catchment Polygon Processing: Converts catchment grid to polygons.
Drainage Line Processing: Converts stream grid into a vector drainage line feature class.
Adjoint Catchment Processing: Generates aggregate upstream catchments for each catchment polygon.
Drainage Point Processing: Generates drainage point (outlet) for each catchment.
Longest Flow Path Calculation for Catchments: Calculates longest flow path for each catchment.
Longest Flow Path Calculation for Adjoint Catchments: Calculates longest flow path for each adjoint catchment.
Slope: Creates a slope grid.
Slope > 30: Identifies cells with slope greater than 30%.
Slope > 30 and Facing North: Identifies cells with slope greater than 30% that face north.
Weighted Flow Accumulation: Allows a weight grid to be applied in the calculation of flow accumulation.

Watershed Processing

Batch Watershed Delineation: Creates a watershed (polygon feature class) for a set of batch points supplied by the user. Watershed features are overlapping if input points are on the same stream.
Batch Sub-watershed Delineation: Creates sub-watershed for a set of batch points supplied by the user (calculates partial watersheds between points).
Drainage Area Centroid: Generates the centroid of drainage areas as centers of gravity (used to calculate the average elevation of a catchment).
Longest Flow Path: Identifies and computes the length of the longest flow path in a selected set of drainage areas (used to calculate the slope of a catchment).
Longest Flow Path USGS Method: Identifies and computes the length of the longest flow path in a selected set of drainage areas using the USGS method
Longest Flow Path for Watersheds: Identifies and computes the length of the longest flow path for selected watersheds.
Longest Flow Path for Sub-watersheds: Identifies and computes the length of the longest flow path for selected sub-watersheds.
Flow Path Parameters: Computes the following parameters for the longest flow path:

- LengthMi: Length of the flow path in miles
- SlpFtMi: Slope along the flow path in feet per mile
- Slp1085FtMi: 10-85 slope along the flow path in feet per mile

Network Tools

Hydro Network Generation: This function builds the drainage network from the features created through the previous toolsets. It converts drainage features into network features as a geometric network. The function performs the following operations:

- Convert the Drainage Line feature class into the Hydro Edge feature class.
- Create the Hydro Junction feature class as the From Nodes/To Nodes without duplication (a junction is defined by one and one point only)
- Create an Arc Hydro Geometric Network based on the Hydro Edge and Hydro Junction feature classes, and set the flow direction in Hydro Edge in the digitized direction.
- Establish attribute connectivity between the drainage and network feature classes.
- Attribute each junction with HydroID and NextDownID field.
- Create a relationship between the Catchments and Hydro Junctions: HydroJunctionHasCatchment

Node/Link Schema Generation: Generates a node-link schema. The nodes are defined by the centers of the polygons representing basins and by points that represent locations of interest in the model. The points include basin outlets, river junctions, water intakes and other facilities.
Store Flow Direction: Reads the flow direction for a set of edges from the network and writes the value of the flow direction to a FlowDir field in the Edge Feature Class.
Set Flow Direction: Sets the flow direction for selected edges in a network edge feature class. If no features are selected, the tool sets the flow direction for all the edges in the feature class. Assigning flow direction based on an attribute allows the user to delete the network (for maintenance or distribution reasons) and to still retain the proper flow direction values. Once the network is re-established, flow direction can be assigned by attribute without having to create sinks in the network. Flow directions for indeterminate cases (such as edges in loops) can be

assigned manually with this tool. Situations in which flow directions may change (such as in canals in flat areas) can also be modeled.

ArcHydro Database Creation Plan

The ArcHydro Tools provide an efficient means of building an initial geodatabase for McKenzie watershed spatial data analysis and potential modeling support. By running the tools, described above, a geometric network (synthetic streams) for the basin will be created along with related watershed components (catchments) and network junction points. This structure forms the basis for modeling (with BASINS or potentially other models) and other analysis. The monitoring database will be linked to this structure through the creation of monitoring points linked by the HydroCode (EWEB_SITE_ID) and the Hydro_ID generated through ArcTools. Additionally, time series data (contained in the MEASUREMENT table of the monitoring database) can be brought into the Geodatabase though initial time series functionality is minimal. Other feature classes that should be brought in include the Hydrography features and 1:24,000 streams and water bodies.
Steps for preparing data to build the ArcHydro Geodatabase for the McKenzie watershed:

- ✍ Prepare 1:24,000 streams and water body layers that extend into Linn County and is fully attributed for the McKenzie watershed.
- ✍ Investigate the used of 10 meter DEMs for the watershed vs. the 30 meter DEM data already available. Initial testing with 30 meter data created a fairly sparse stream network even with low threshold values. The smaller resolution elevation data would have to be integrated with Linn County in order to get the whole watershed. The 30 meter data is available on the regional server for the entire McKenzie watershed (the grid covers the entire state of Oregon).
- ✍ Created the initial ArcHydro database from the ArcHydro data repository provided (this is done in ArcCatalog)
- ✍ Run the ArcHydro Tools as described above.
- ✍ Add the MONITORING_SITES in the monitoring database as monitoring points in the geodatabase. Associate them with network junctions (through Hydro_ID)
- ✍ Create ArcHydro time series structure on the SQL Server database housing the monitoring data (two tables, TSType and TimeSeries, must be created and populated from the MEASUREMENT table. Some additional information regarding the data's time interval, whether is a regularly time-spaced data set, and the nature of the measurement values will be needed in addition to the data already available)

Summary

If EWEB has as a goal the development of hydrologic modeling capability for the McKenzie watershed, then ArcHydro provides significant benefits. Initially there may not a great deal of benefit derived from the effort required to build and populate the ArcHydro data set, especially if most of the usage of the data is limited to viewing and mapping. It is when modeling and network-based analysis come into play that the model will provide the greatest benefit. That said, the familiarity gained by working with the ArcHydro data model and geodatabase will be a plus. Though no panacea, the geodatabase also will provide many more tools and options for integrating tabular data from the water quality monitoring program with spatial data in GIS. The primary factor driving the schedule for implementing the ArcHydro data model will likely be the release date for EPA's BASINS model as an ArcGIS 9.X extension. The latest information on this describes a "Summer 2004" release.