Title

Managing Fish and Aquatics Data Using the ArcHydro Data Model

Author

Christian Weik

Abstract

The Foothills Model Forest (FtMF) is a not-for-profit partnership conducting research on sustainable forest management. The Fish and Watershed program and its partner agencies have traditionally stored field-collected data in ad-hoc, non-spatial spreadsheet and database systems. To better consolidate all field-collected data and to enforce correct spatial placement of field sampling locations a database was designed using traditional entity-relationship techniques and the ArcHydro data model to tie together spatial and non-spatial features in a Personal Geodatabase. The resulting database functionality includes data entry in Microsoft Access and fine-scale mapping and analysis capabilities to illustrate and predict species richness.

Introduction

The FtMF is a not-for profit corporation that conducts research on sustainable forest management. The Foothills Model Forest (FtMF) is one of eleven Model Forests across Canada. The FtMF covers an area of approximately 2.75 million hectares and lies within the boreal, montane, sub-alpine and alpine forest regions of Canada.

The Foothills Model Forest (FtMF) conducts research into both the ecological and economic aspects of forest management. Some of the on-site projects at the FtMF include:

- Grizzly bear research project
- Natural disturbance research project
- Fish and watershed research project
- Foothills Growth and Yield Association

Background

The FtMF has been collecting fish inventory data for the past seven years and currently manages data for over 1300 surveys on the FtMF landbase with a staff of between two and five people. These data include survey site locations, fish species, fish lengths and weights as well as aquatic and terrestrial data such as water depth and riparian vegetation that pertain to fish habitat. These data were stored in a Microsoft Access database and in accompanying Microsoft Excel files. The Access data entry and reporting components of this system worked well initially, but the database design did not enable expanding collection needs, did not constrain field values to ensure data were logical and did not effectively capture the spatial dimension of the data. Over time this tool resulted in several chronic and typical problems in managing these data. They included:

- Significant data errors created during data entry due to little or no database constraints.
- Significant data errors created in placement of survey locations due to data entry errors and the inability to place the survey points using GIS.

- Survey locations not falling on hydrographic features due to data entry errors or a differing accuracy in method of capture (i.e. GPS).
- Significant time spent in entering spatial reference information such as management units, watersheds, stream name, stream tributary etc.
- Poor database design resulting in the inability of end users to create some queries.
- Inability or difficulty in effectively mapping the in-stream single point or linear sections representing surveyed areas.
- Inability or difficulty in creating fish species presence/absence maps.
- Significant time spent revisiting data in an attempt to control effort data quality.

These problems related to spatial reference were particularly apparent given the complex nature of the stream network system. Users found it impossible to reference, in a tabular database, position in a stream relative to downstream and upstream features.

Approach

With the emergence of systems to integrate spatial and non-spatial databases and the integration of linear referencing tools in ArcGIS 8.3 the FtMF set out to develop a replacement system applying these new technologies to better manage its data.

In developing a new system our focus was to meet the data management, analysis and reporting needs of the FtMF Fish and Watershed team. In addition we facilitated input from key partners (two additional groups) in developing these tools to expand the utility of a new system to provide additional benefit to those organizations in managing their fish survey data and facilitate easier exchange of data between them.

A secondary objective of this initiative was to test and demonstrate to our partners how Geodatabase technology and specifically a standard modeling approach could be applied to natural resource data management.

They key requirements of the new system were identified to be as follows:

- Consolidate all survey related data in a single database system.
- Provide a better interface to enter survey data from field sheets.
- Provide more rigorous constraints on the data as it is entered to minimize errors.
- Provide a more flexible data model to facilitate expansion and simplify querying and reporting.
- Ensure correct placement of survey locations relative to existing GIS layers.
- Enable better mapping capabilities of fish related parameters.
- Provide access to the non-spatial data without requiring GIS software. This was an interim requirement as access to ArcGIS 8.x tools was limited among partner agencies at the time of development but was expected to become available eventually.
- The database is to house about 100,000 to 110,000 spatial features including river centerlines, river shorelines, lakes, watersheds and survey locations.
- The spatial database need not be multi-user. That is, in each organization only a single user will be writing spatial data at any given time.

Based on these requirements and available software funds and our objective of testing and demonstration we decided to develop the system as a Personal Geodatabase using Microsoft Access 2000.

Overall we also wanted to focus on the integration of spatial and non-spatial systems. We wanted to emphasize the elimination of complex spatial reference data in a tabular database and replace it with spatial features and GIS functionality. In the historical databases we worked with, the weakest and most complicated components were those that tried to store spatial references, and in some cases network references (hydrocoding), in tabular databases. Our approach was to demonstrate eliminating the need to store large amounts of spatial references in traditional tables. We hope that this approach will facilitate change in an apparently deeply ingrained way of thinking that continually segregates GIS data and business data.

Methodology

The requirement of access to the system without GIS software posed a challenge. The development therefore was divided into two streams of spatial and non-spatial, although references between the two were retained throughout the process.

Non-spatial database development.

Traditional database development strategies were applied in planning and developing the non-spatial component. The following tasks were undertaken to develop a system in Microsoft Access 2000.

- Needs analysis. This involved several iterative meetings with the FtMF fish and watershed team and its partner agencies.
- Database design. A data model was developed using traditional Entity Relationship methods in Microsoft Visio Enterprise software.
- Database creation. The database schema was generated in Microsoft Access 2000 directly from Visio.
- Forms development. The interface for non-spatial data entry and data navigation was developed in Microsoft Access 2000.
- Historical data loading. Historical data were loaded into the new schema using append queries and VBA scripting.

Spatial database development

We built the spatial component (Personal Geodatabase) of the database on ESRI's ArcHydro spatial data model framework. We chose this route to test the concept of a standard modeling approach to natural resource management among our partners. It is hoped that based on the success of this system there may be additional interest in trying to standardize models in an effort to better monitor natural systems across multiple jurisdictions.
The Geodatabase technology allowed us to take advantage of topology rules and network tracing functionality. These are key functions that were not available in earlier versions of ArcInfo. These functions ensured accurate placement of survey locations on the stream network and enabled us to perform rudimentary network analysis.

The steps taken in building the Geodatabase were based on datasets, tools and knowledge provided in ESRI's *ArcHydro – GIS for Water Resources* book. This package provided the ArcHydro schema and the tools to generate the schema in a Personal Geodatabase. The general steps taken to generate the ArcHydro data model were as follows:

- Apply the Schema Wizard to generate the schema within a Personal Geodatabase.
- Drop the HydroNetwork feature class in order to allow loading of hydro edge features.
- Load hydro edge features from a topologically correct stream network layer. It was not necessary to fix the stream network, as a topologically correct one was readily available.
- Rebuild the network manually using the populated edge and junction features in ArcCatalog.

We did make changes in the form of additions only to the basic ArcHydro model to meet our custom requirements. We retained all objects and fields from the original model to facilitate loading of new datasets. The changes we made were as follows:

- HydroRoute systems feature class under the Hydrography feature dataset. One of the key requirements of the system is the ability to map from specific survey locations in the network to the outflow. Route systems that travel from the stream headwaters to the outflow were created manually. This enabled the creation of point and line events from the survey start location to the outflow.
- Topology rules. This forces all new features to be snapped to an edge in the geometric network.
- Additional columns. We added columns to most of the feature classes to enable us effectively join to the non-spatial entities and to carry forward valuable attributes from our coverage datasets.
- Relationship class between the UserPoints and HydroLineEvent. This relationship is required to tie all events back to the non-spatial tabular database.
- Topology class to ensure all UserPoint features fall on the HydroEdges.

The general steps taken to load data into the ArcHydro model are as follows:

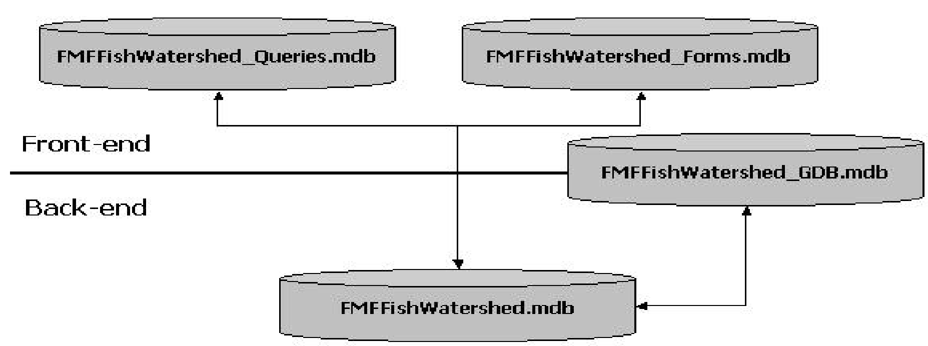
- Load hydro edge features from a topologically correct stream network layer. This step must be completed prior to creating the geometric network therefore this step was listed in the previous section also.
- Use the ArcHydro extension tools to populate downstream length fields on the HydroEdge feature class.
- Create feature class HydroRoutes in the hydrography feature dataset.
- Load lake features into Waterbody feature class.
- Build route systems from headwaters to outflow on those streams where surveys have occurred. Apply the "trace downstream" network function to find the down stream paths.
- Load survey locations into UserPoint feature class.
- Correct all UserPoint topology errors.
- Create events in the HydroLineEvent table representing survey locations. Populate downstream measure as the full distance to the outflow to represent assumed fish presence.

At this time survey points are represented as both UserPoints and HydroLineEvents, that is, they are redundant. It is hoped that once the users are trained sufficiently or interface customization is developed the surveys will be stored only as events.

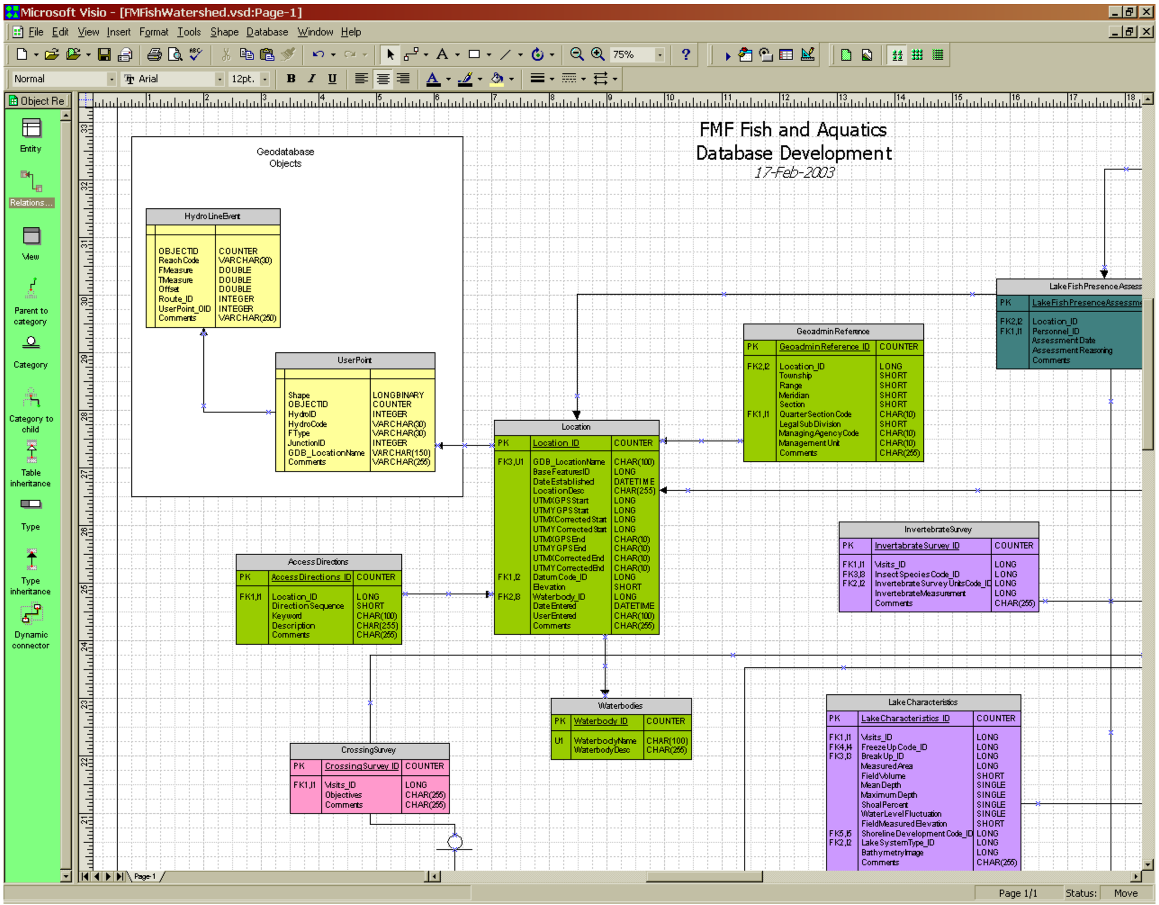
The database system is comprised of several linked Access MDB files. We employed front and back end database design. Each of these files is related via linked tables. The intent of this design is to add a level of security by protecting the back-end tables from being altered and to separate database objects in a logical manner that is more easily managed. The system is comprised of the following MDB files:

- FtMFFishWatershed.mdb - This is the core back-end database. Non-spatial attributes are stored directly in these tables. Tables that reference spatial locations contain foreign keys to the appropriate Geodatabase tables.
- FtMFFishWatershed_Forms.mdb - This is a front-end database. It contains forms used for data entry.
- FtMFFishWatershed_Queries.mdb - This is another front-end database that contains canned queries, simplifying access to commonly reported data.
- FtMFFishWatershed_GDB.mdb - This is the Personal Geodatabase containing the spatial features.

The image below illustrates how the Access databases are positioned.



The image below illustrates how the Geodatabase is linked to the non-spatial Access database from a database design perspective. Currently there is a single field designated in the Locations table called GDB_LocationName that is a foreign key to the Geodatabase table UserPoints.



Challenges

We met some challenges working with very fresh technology in ArcGIS 8.3. The most notable issues were:

- Inability to recreate the ArcHydro Network features using the schema wizard once all other objects were created. The workaround was to create the network feature class manually.
- Inability to add new features to feature classes that contained a unique index. The workaround was to remove the unique index from the table.
- Inability to relate Access views or Access tables with no primary key to feature tables. The workaround was to generate tables and manually add primary key fields.

Each of these problems we encountered has been reported by other developers and will hopefully be resolved in upcoming releases.

Results

At present we have attained an integrated system based on ArcGIS, Microsoft Access technology, and the ArcHydro data model framework. Most historical data are loaded into the back-end non-spatial database. Survey locations are stored in the Personal

Geodatabase as both UserPoint and HydroLineEvent features. We are able to create fish presence maps based strictly on survey data stored on the attribute side although the process to do this still requires the expertise of the GIS staff.
The Personal Geodatabase contains about 85,000 features and the performance remains good to excellent. At this point we do not expect to require moving the system to SDE.
We still have to complete testing, acceptance and training phases to get to an operational system. It is our expectation that the assistance of a GIS specialist will still be required to move survey locations to event features. That is, there remains a learning curve to overcome in the ArcMap interface through training or customization.
It remains to be seen how effective the standard data modeling approach will be in terms of acceptance and functionality. We have and will make an effort to retain all characteristics of the model and only provide modifications in the form of additions; and only make additions when absolutely necessary. It is hoped that this approach will better retain the power of a broad scope solid data model and of data exchange.
We struggled at times with the new technology in terms of learning and bugs. We found the best prerequisites and resources to assist with the process to be as follows:

- Knowledge of Geodatabase design and structure
- Knowledge of database design
- Knowledge of linear referencing
- ESRI press release Arc Hydro – GIS for Water Resources
- ESRI User Forums

Future Development

We are nearing the completion of a stream reach and watershed dataset that we plan to load into the ArcHydro model. These new datasets will be the foundation for building Resource Selection Function (RSF) models to predict fish presence. These results will also be stored in the Geodatabase and be provided to partners for development planning purposes. We will also look at customizing some of the ArcMap interface to more easily allow end users to enter new survey locations and to build sophisticated maps themselves.

Author Information

Christian Weik, GIS Coordinator, Foothills Model Forest
Box 6330, Hinton, Alberta, Canada
Phone 780.865.8290
Fax 780.865.8331
Email Christian.Weik@gov.ab.ca
Web www.fmf.ab.ca