Fashioning the Data-Management Solution

E. Christine Santi and Tara Morgan
Florida Fish and Wildlife Conservation Commission
Fish and Wildlife Research Institute

Abstract:

The Fish and Wildlife Research Institute (FWRI) is a division of the Florida Fish and Wildlife Conservation Commission, a state agency dedicated to managing Florida's fish and wildlife resources. The FWRI GIS data-management team has maintained the Marine Resources GIS (MRGIS) data collection for more than ten years. This collection of data has grown significantly over the years and now spans two ArcSDE Oracle geodatabases, numerous CD and archive data holdings, and a terabyte of imagery. The MRGIS data and metadata, like two sides of a zipper, must be kept together to keep the whole functional. This presentation will discuss how FWRI's GIS data managers keep the zipper intact and keep the data from "being caught with their pants down." Topics will include presentation of data to users, metadata documentation, database maintenance, and data archiving.

Introduction:

The Coastal and Marine Resource Assessment (CAMRA) subsection of the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute (FWRI) has maintained a large GIS data collection for more than ten years. Although this collection was originally dedicated to marine resources, it is growing to encompass data for all of Florida's coastal, inland, and freshwater environments. This collection of data now spans two ArcSDE Oracle geodatabases, numerous CD and archive data holdings, and a terabyte of imagery. Since 1986, managers have made data management a priority and dedicated staff and resources to maintain the data. We offer a model of the benefits of having dedicated data-management staff providing information to the public. Dedicated GIS data-management and metadata professionals allow our GIS analysts to concentrate on output and production. In our organization, data and information are brought together like the two halves of a zipper.

The two fabric sides of the zipper are the physical, visual GIS data, such as shapefiles, and the important intangible information, such as data history and purpose. Metadata, as the slider of the zipper, bring all aspects of data management together. The data-management professionals, as the teeth of the zipper, hold everything in place. In our organization, the MRGIS is the pull tab, the most accessible part of the zipper. Just as breaking the pull tab of a zipper makes it difficult to get into and out of clothing, if the MRGIS is damaged, it is difficult to access stored information. Like breaking the slider of a zipper, if the metadata are damaged, we are unzipped, and the data are unusable—how embarrassing! Detailed metadata increase the efficiency of the analysts, assist in the timely answering of public information requests, and document data maintenance activities. Metadata bind the data together wherever they go, through multiple databases and eventually into archive.

GIS Data Management at FWRI, a Brief History:

Beginning in 1986, CAMRA management staff members invested in GIS and remote sensing as tools to analyze Florida's natural resources. Since then, FWRI staff members have worked to develop the MRGIS to compile, analyze, and display complex scientific observations in a simplified, map-related form. FWRI staff members use the MRGIS to create maps and analyses that are used to focus research efforts, create resource-protection strategies, support the development of management initiatives, and educate both decision-makers and the public. These aspects of research have continued to receive support in the form of dedicated staff and funding.

The initial development of the MRGIS began in 1986. In the early 1990s, as the MRGIS began to grow and the number of analysts increased, CAMRA dedicated staff members specifically to the GIS data-management needs. Early on, a "GIS librarian" position was created to develop the beginnings of a centralized enterprise GIS within ARC/INFO's spatial data-management software, LIBRARIAN. The MRGIS centralized data repository became the one consistent source of data for all users. The advantage of having a dedicated GIS librarian to manage the repository was immediately apparent because it allowed the other analysts more time to do their own research.

Since its inception, the MRGIS has gone through several "new looks." The most significant change came when the data maintained in LIBRARIAN in a tiled structure were migrated to ESRI's ArcSDE (Spatial Database

Engine) and stored as seamless feature classes. When CAMRA implemented Oracle as the relational database management system (RDBMS) in 2001, the number of staff dedicated to GIS data management increased. A full-time position dedicated as a junior Oracle database administrator and ArcSDE administrator was hired to manage the GIS and remote sensing data warehouse infrastructure. The role of metadata has also increased from that of nearly an afterthought to an integrated part of the GIS data-development process. Metadata once seemed like an optional accessory to the data, but it is now a requirement of the data's "dress code." Data cannot leave FWRI unless they are properly adorned by metadata.

Data Acquisition:

FWRI acquires the data in the MRGIS from many sources, including state and federal agencies such as Florida's water management districts, the Florida Department of Environmental Protection (DEP), The National Oceanic and Atmospheric Administration, and others. Our grant budgets also allow in-house data creation and contracted data creation. A primary thrust of the MRGIS program is ongoing data development to support the Florida Marine Spill Analysis System (FMSAS), which was developed to aid the United States Coast Guard (USCG) and the DEP's Bureau of Emergency Response. The FMSAS is used to map coastal resources (such as beaches, seagrass beds, and mangrove shorelines) that may be threatened by a spill and to provide rapid analyses of the possible effects of a spill on those resources. These analyses assist decision-makers in developing response and clean-up strategies, prioritizing response efforts, and assessing damage after a spill occurs. Additional data are also acquired to represent coastal and marine habitats and other basemap features to support the Boating and Angling Guide Series and other cartographic work.

Dedicated data-management staff members are vital at this stage of data development because much time and effort are required to ensure that we acquire and maintain quality data for the entire state of Florida. At this stage, metadata are especially important to document procedures applied to the data and to give credit to source organizations. Sometimes different agencies maintain data for different parts of the state, and it sometimes requires much staff time to get the data into a format that is comparable across the entire state.

For example, FWRI maintains a data set of the USCG Aids to Navigation (ATONs) for the southern U.S. Atlantic and Gulf of Mexico regions. We generate this data set from the geographic coordinates and light characteristics published in the Coast Guard's annual *Light List* series. Our area of interest includes both the seventh and eighth USCG districts. We receive these data in spreadsheet format. The challenge with these data is that each district maintains its section of the *Light List* independently and that each spreadsheet records the information differently. We must work to get these data into compatible formats, merge all the spreadsheets, and then use the Add X,Y tool to convert the information to a GIS file.

We acquire as much metadata as possible from the sources that provide the data. If the source does not have metadata available, we try to work with the source agency to create a Federal Geographic Data Committee (FGDC)-compliant metadata record. If metadata cannot be created, we do not use the data.

Data Quality:

The quality of the data is very important to GIS data-management staff members, who are responsible for ensuring that all of the data's zippers are zipped. While we do not produce survey-quality data, most of our data are used in cartographic products and analyses that require documentation of positional and attribute accuracy. Much of the data we acquire from other agencies have already undergone quality-assurance tests that are documented in the metadata. Therefore, we complete only a cursory review of most data received from these agencies. When data are created in-house, we check them for logical consistency, attribute accuracy, and general positional accuracy. This involves on-screen review of the data and comparison to background source images (nautical charts or photography) using ArcMap. In the past, we have had dedicated Quality Assurance and Quality Control staff members on the data-management team, but currently, the analysts do their own review, with assistance from the GIS librarian on request.

Data Storage:

The MRGIS is stored in two Oracle 9i version 2 geodatabases. We use ArcSDE 8.3 as the middleware to serve the enterprise geodatabase to the analysts across the local area network. FWRI GIS data-management staff members will continue to redesign the MRGIS as upgrades to ArcSDE and Oracle become available. ArcSDE is implemented on the same server as the Oracle geodatabases. The designated FWRI database server is a Dell

PowerEdge 6450 Server, which has 8 GB of RAM, 300 GB of internal storage. In addition to the resources available on the database server, CAMRA has one terabyte of storage, which is used to archive historical project data.

The upgrade to Oracle 9i version 2 allowed us to streamline data storage by taking advantage of locally managed tablespaces with uniform extents. FWRI is responsible for mapping the statewide natural resources in Florida; therefore, the spatial extent of our data varies from specific regions of Florida to the entire state. The multiple spatial extents and different levels of detail impelled data-management staff to implement tablespaces with three different uniform extents; this increased the overall efficiency in disk space used. After assessing a sample set of the MRGIS as shapefiles, we determined three size categories of our data: small, medium, and large. Data falling into the small category are loaded into the tablespace with 128 KB uniform extents. Data falling into the medium category are loaded into the tablespace with 512 KB uniform extents. Data falling into the large category are loaded into the tablespace with 2 MB uniform extents.

Data Organization:

FWRI maintains two ArcSDE instances. One stores the newest, "designer" data, whereas the other is a historical database, storing older, "outmoded" data, which are still useful for time-series studies. With more than 600 feature classes, we also implemented feature datasets as a means to categorize the data in the same manner as libraries in the old LIBRARIAN system. The MRGIS consists of thematic data, including bathymetry, benthics, benchmarks, boating, boundaries, environmental sensitivity index, elevation, grids, hydrography, land cover and land use, pollutants, shoreline, soils, transportation, and wetlands.

Initially the naming convention of the MRGIS was difficult for analysts to understand. The names were shortened due to the character limit of coverage names when coverages were stored in LIBRARIAN. In the two ArcSDE geodatabases, we've taken advantage of the longer character limit that Oracle allows for table names, and we developed a standardized, more intuitive naming convention for the data. Because data are arranged alphabetically in ArcCatalog, we decided that the most important keyword indicating the main feature of the data set should be used as the first part of the name. Of course, name changes were all documented because we have technical reports referencing the old names, and we still work with people who have received data in the old coverage format. The naming history and a brief description of each data and metadata set are stored in a stand-alone Access 2000 database. Data-management staff members regularly maintain and update this database.

In addition to standardizing the naming convention, FWRI implemented feature datasets as a means of categorizing and organizing the data thematically. FWRI data-management professionals have learned that precision of the feature dataset—also known as the scale or the system units, depending on the software report viewed—is an important concept to remember when creating feature datasets. According to ESRI (Childs *et al*., 2003), precision represents the denominator of the fractional value, in the units of measure of the coordinate system, of how close together vertices can be located in a feature in a given feature dataset. If this is too high, it could limit the spatial extent that can be stored in the feature dataset, but if it is too low, some features may not import properly. We have seen that some polygons are self-intersecting on import when the precision is set too low. The ArcSDE command line tools will report which polygons contain errors, but when loading data using ArcCatalog, the error is not reported. If a polygon or line is self-intersecting, it will not draw, and the SHAPE.AREA or SHAPE.LEN items will have a value of zero (support documentation available on request). There are several solutions to this situation, including manually editing the offending polygons, CLEANing the data if it is a coverage, increasing the precision of the feature dataset, or loading the data as a stand-alone feature class.

Database Backup and Recovery:

The data owner loads all feature classes via the ArcCatalog interface, but from most users' perspectives, the MRGIS ArcSDE Oracle geodatabases are currently read-only data warehouses. Only the data owner is allowed to make alterations. We have few transactions between full database backups, so we run the database in NOARCHIVELOG mode for simplicity. When there is a lot of activity adding new data layers, we create an Oracle export file weekly; an Oracle export file is created less often if there are few new data layers added. The MRGIS Oracle geodatabases are part of the corporate Information Technology (IT) backup and recovery plan. The three-member IT section, which manages computing and networking resources for the FWRI employees throughout Florida, continues to back up the server on a weekly basis.

GIS data-management staff members have been pleased with the functionality and dependability of Oracle as the RDBMS since its implementation in 2001. The few interruptions of FWRI's Oracle production databases have

been the result of hardware failure and the corruption of a datafile. The only severe problem the GIS Oracle databases have experienced was due to an unexplainable anomaly. ArcSDE crashed in March 2003 after the database server, which runs both ArcSDE and Oracle, was shut down and physically moved to another location. We were sure that all our data's zippers had burst, leaving only barely recognizable shards of teeth and scraps of fabric. To remedy the problem, we worked with ESRI support for two weeks before we finally found a way to mend the zippers and restore the database to working order (documentation available on request). There were problems with all attempts to restore the data in the ArcSDE 8.1.2 version. Even after restoration from an older backup file, the error was still evident, although it was not present when that prior backup file was created. The only way to rescue the corrupt database was to copy the data to multiple personal geodatabases. The problem was not present in the personal geodatabase format. To fully recover, we upgraded to ArcSDE 8.3 and copied and pasted the feature datasets from the personal geodatabases. This experience taught us how to be better prepared and taught us the importance of keeping the data in a format other than the DMP file. Fortunately, we had and continue to maintain a backup of the MRGIS in the form of shapefiles; this allowed our users access to the data while the server was down.

Data Distribution:

The Florida law commonly called the "Sunshine Law" (Florida Constitution, 2002 and Florida Statutes, 2003) requires state agencies to provide data to the public on request. This includes all of the GIS data kept at FWRI. Dissemination of this data is on a per-request basis. Funding for the GIS librarian position and the MRGIS comes from a grant for Federal Aid for Sport Fish Restoration, and part of that funding is contingent on providing GIS data to the public.

We developed an ArcIMS application to serve the most-requested data sets online. These data are, of course, only presentable with their metadata intact. We took extra care to ensure that the data and metadata presented online are understandable to the public. We accessorized the data with additional attributes to enhance queries. We proofread the metadata to translate the unfashionable GIS jargon in the abstracts into something more meaningful to casual users. The metadata still include all of the important processes deeper within the document, but the abstract section allows the public to get a quick and easy-to-read idea of what each map represents.

Not all data are served on the ArcIMS site—it is still a work in progress. The GIS librarian continues to play an important role as a contact person for all users who request data that are not served online.

Metadata:

In 1994, President Clinton signed Executive Order 12906 (Federal Register, 1994), which required all new geospatial data to be documented according to the FGDC standards. At FWRI, we have a dedicated metadata coordinator, and each section has a metadata technician. In the GIS group, the GIS librarian facilitates creation of metadata within the section. Analysts are responsible for documenting their own projects; the GIS librarian provides them with guidance and assistance. Metadata must be completed according to FGDC standards, and if the data are provided via ArcIMS, the metadata must be written in a manner that is understandable to the public. The GIS librarian and the Web editor carefully review these documents before posting.

In 2000, we migrated our desktop workstations from the UNIX Sun-Solaris operating system to the Microsoft® Windows operating system in order to prepare for the upgrade to ArcGIS 8.0 Desktop. At this point, we had only minimally compliant metadata in UNIX text files (.rpt). As a part of the data-management upgrade, we dedicated three months of three staff members, working 40 hours per week to enhance existing metadata into a fully FGDC-compliant format and to create metadata that did not previously exist.

FWRI's standard metadata software is Intergraph Corporation's Spatial Metadata Management System (SMMS). This tool includes the Biological Data Profile of the Content Standard for Digital Geospatial Metadata, FGDC-STD-001.1-1999, and can be used to document spatial and nonspatial data. This is the ideal tool for many FWRI sections, where GIS is not as well developed as it is in the CAMRA group and where the FGDC biological profile is a prominent aspect of the metadata. In CAMRA, we use SMMS but do not use the biological profile. We also take advantage of metadata in ArcCatalog, which has the added advantage of allowing us to link the metadata directly to the data set in the ArcSDE geodatabase.

When multiple types of software and formats are used with metadata, it is imperative to understand how each type of software treats the metadata and if there are synchronizers that update some of the metadata elements. When converting metadata, it is important to develop a procedure that allows appropriate synchronization of elements but not inappropriate alteration during conversion. When documenting the spatial reference, geographic

extent, and the number and type of spatial objects (Spatial Data Transfer Standard terms), we like the synchronization feature of ArcCatalog. Our group had to learn from trial and error when synchronization is appropriate and when it is not. Occasionally, synchronization has produced inappropriate results; if this occurs, we turn off the synchronizer after the initial synchronization. An example of this is the Online Linkage field, which ArcCatalog synchronizes to the location of the file. When the file is a feature class in an ArcSDE geodatabase, this field shows the login, server, and port number of the SDE instance. When this information is used in-house, it does provide an online linkage. When the data are served to the public, either through the ArcIMS site or on a per-request basis, however, this is a bit more information than the average user would need and a bit more information than we wish to provide. A link to the ArcIMS site is appropriate for data that are currently served in that manner, but for the data that are only provided by specific request, we prefer to leave this value blank. The synchronizer in ArcCatalog 8.3 will not overwrite the ArcIMS link, but it would overwrite a blank value with the server information. Through trial and error, we developed a procedure to prevent the default text from being inserted into blank fields. Specifically, important metadata elements are synchronized once at the initial metadata creation and not synchronized again.

Data Archiving:

In the past, getting GIS data digitized and created was the priority. Today, with the advent of faster and more powerful computing resources, the focus is becoming the improvement of the spatial accuracy and quality. FWRI's GIS data warehouse is extremely valuable for temporal analysis. It is rich in historical vector data and remote-sensing imagery that represents data from as far back as the 1920s. Paper documents produced long before metadata was standard hold a wealth of information about some of the early GIS layers, such as seagrass. Paper is cumbersome, and it is about as attractive to GIS data managers as a model clothed in a cardboard box is to modern fashion designers. It is important for FWRI to capture the information contained in paper documents and convert it to a digital format before the paper is lost or ruined by the vagaries of age and environment. The information contained in the written reports must be transferred to the digital metadata documents for these data layers. Often, researchers are asked to retrieve older GIS data. To use any data correctly, users need to know the context in which the data were created. Using the earliest available data with formal metadata can serve as a baseline for temporal analysis of environmental resource changes. We are currently researching software to manage the documents once they have been converted to a digital format. In addition to the paper documents, some of the older data are archived on various media that are no longer industry standards. We must transfer these data to other archiving media before the current media are no longer readable. The goal is to archive data and FGDC-compliant metadata in a format that will be accessible for many years.

Conclusions and Further Research:

This is an exciting time; data management has come a long way at FWRI. The once "naked" data have been "dressed up" and made presentable with FGDC-compliant metadata. Careful attention has been paid to the documentation of data quality. We have migrated most data to the latest versions of software; we have taken great care to ensure that they arrived intact and are backed-up and maintained. While data-management staff continue to create and acquire new data from quality sources, FWRI will continue to invest in the technology infrastructure necessary to successfully maintain an enterprise GIS. We will continue development of Florida's inland GIS data within the main GIS database structure. To take advantage of additional functionality in the geodatabase, including topology and versioning, we also plan to research predefined geodatabase models.

Yet, there is still room for improvement in our data-management design, so that it can be applied to the Florida Fish and Wildlife Conservation Commission (FWC) as a whole. The FWC is a data-rich agency; however, to date, there has been little synthesis of data between projects. The FWC maintains years of scientific study data and a wealth of nonspatial business data. In the long term, the FWC is dedicated to synthesizing its data into information and to getting that information to users in a timely manner. By harnessing the power of Oracle as an RDBMS, managers hope to bring together the location GIS data with the attributes of the business data, in effect merging the worlds of spatial and nonspatial data. This is where the value of time and resources spent to create metadata will be actualized. Metadata will continue to be valuable to our publicly accessible Internet sites and to the people making management decisions. In the future, metadata will be a critical communication tool as data become available agency-wide through Internet technology known as "portals."

Acknowledgements:
This work was supported in part under funding from the Department of the Interior, U.S. Fish and Wildlife Service, Federal Aid for Sport Fish Restoration Grant Number F-66.

Reference:

CHILDS, C., G. KABOT, P. MCGRAY, AND B. SIMMONS. 2003. Exercise 3-B: Setting the Spatial Reference. Pp 3-16-3-29 *in* Building Geodatabases I (for ArcEditor™ 8.3 and ArcInfo 8.3™). ESRI (Environmental Systems Research Institute), Redlands, CA. Rev. May 2003.

FEDERAL REGISTER. 1994. Executive Order 12906 of April 11, 1994. Federal Register 59(71): 17671-17674 . Available at http://www.archives.gov/federal_register/executive_orders/pdf/12906.pdf. Accessed June 22, 2004.

FLORIDA CONSTITUTION. 2002. Article I, Section 24 Access to public records and meetings. History: Added, C.S. for C.S. for H.J.R.'s 1727, 863, 2035, 1992; adopted 1992; Am. S.J.R. 1284; adopted 2002. Available at http://www.flsenate.gov/Statutes/index.cfm. Accessed May 13, 2004.

FLORIDA STATUTES. 2003. Title XIX Public Business, Chapter 286, Section 286.001 Reports statutorily required; filing, maintenance, retrieval, and provision of copies. History.--ss. 26, 28, 29, ch. 84-254; s. 12, ch. 92-98; s. 104, ch. 92-142; s. 29, ch. 95-196. Available at http://www.flsenate.gov/Statutes/index.cfm. Accessed May 13, 2004.

For more information on the migration of the MRGIS from Arc/Info 7.2 to ArcGIS 8 see:

MacAulay, G. and T. Morgan. 2001. Migrating our GIS into Oracle - A 2001 Relational Odyssey. American Society of Photogrammetry and Remote Sensing Conference Proceedings, April 2001.

Prior to July 1, 2004, the Fish and Wildlife Research Institute (FWRI) was known as the Florida Marine Research Institute (FMRI).

Contact Information:

E. Christine Santi
Assistant Research Scientist, GIS Database Administrator
Florida Fish and Wildlife Conservation Commission, Fish and Wildlife Research Institute
100 8th Ave. SE
St. Petersburg, FL 33701
christi.santi@fwc.state.fl.us or christi.santi@MyFWC.com
Phone: (727) 896-8626
Fax: (727) 893-1679
Email for GIS information: GISLibrarian@fwc.state.fl.us or GISLibrarian@MyFWC.com

Tara Morgan
Senior Systems Analyst
Manatee County Information Services Department, Geographic Information Systems Division
1112 Manatee Avenue West, Suite 702
Bradenton, FL 34205
tara.morgan@co.manatee.fl.us
Phone: (941) 749-3075
Fax: (941) 749-3086