The Florida Geographic Data Library: Lessons Learned and Workflows for Geospatial Data Management

Crystal Goodison, Alexis Guillaume thomas & Sam Palmer

The Florida Geographic Data Library: Lessons Learned and Workflows for Geospatial Data Management

CRYSTAL GOODISON, ALEXIS GUILLAUME THOMAS,
and SAM PALMER
University of Florida, Gainesville, Florida, USA

The University of Florida GeoPlan Center has been organizing, standardizing, and distributing geospatial data since 1998 through the Florida Geographic Data Library (FGDL), a collection of Florida geospatial data from various agencies, as well as data developed in-house. With funding from the Florida Department of Transportation (FDOT), FGDL serves as a public distribution mechanism for hundreds of Florida geospatial data layers. Since 2004, FGDL layers have formed the data foundation for FDOT's interagency application known as efficient transportation decision making (ETDM), which facilitates the review and permitting process for proposed transportation projects. Data layers are standardized and inspected to ensure that the data are adequate for use in analysis and decision making. To efficiently manage the geospatial data and workflows, a quality assurance/quality control (QAQC) process and relational database schema was developed to track the standardization procedures, lineage, versioning, and archiving of the data layers. This article details the history of the Florida Geographic Data Library, challenges of managing a large geospatial data repository, development of a QAQC process and database schema, and lessons learned.

KEYWORDS   GIS, geospatial data, data management, relational database management systems, metadata

Address correspondence to Alexis Guillaume Thomas, University of Florida, Geoplan Center, 131 Architecture Building, Gainesville, FL 32611. E-mail: lex@geoplan.ufl.edu

BACKGROUND

The University of Florida (UF) Geo-Facilities Planning and Information Re-search Center, or GeoPlan Center, was established in 1984 as a response to local and statewide needs for a teaching and research environment in geographic information systems (GIS). The center is affiliated with the university's Department of Urban and Regional Planning, a part of the College of Design, Construction and Planning. The GeoPlan Center works to support land use, transportation, and environmental planning in the State of Florida by providing geospatial and planning expertise, data, training, and education to the stakeholders involved in the planning process.

The GeoPlan Center has been organizing, standardizing, and distributing geospatial data since 1998 through the Florida Geographic Data Library (FGDL), a collection of Florida geospatial data from various state, federal, and regional agencies, as well as data developed in-house. FGDL has long been recognized as Florida's vector geospatial data clearinghouse. FGDL data is Web accessible, free of charge, in a standard geographic projection, accompanied by Federal Geographic Data Committee (FGDC)-compliant metadata, and typically covers the geographic extent of the State of Florida. With funding from the Florida Department of Transportation (FDOT), FGDL serves as a public distribution mechanism for hundreds of geospatial data layers.

Florida Geographic Data Library History: Impetus and Development

Development of the Florida Geographic Data Library (FGDL) began as an effort to improve public access to GIS databases created by government agencies. Starting in the 1970s and extending into the 1980s, many government agencies were developing detailed GIS databases, using significant resources, to support specific agency programs (Coppock and Rhind 1991). However, only a small percentage of institutional GIS data were being shared with others outside of these institutions (Onsrud and Rushton 1995). This need for data sharing was one major reason that President Clinton established the National Spatial Data Infrastructure (NSDI) in 1994 via a presidential executive order. The order not only established the NSDI to promote geospatial data use and applications, but also explicitly called for a National Geospatial Data Clearinghouse to electronically link data producers, managers, and users (Clinton 1994). In 1996, the Florida Department of Environmental Protection (FDEP) contracted with the GeoPlan Center to gather, standardize, and distribute GIS data for the state of Florida and to assist FDEP with environmental education, public outreach, and information distribution.

Major issues addressed in the development of FGDL included standardization of both file format types and geographic map projection. In the late 1990s, the availability of technical expertise and computer hardware required

for conversion of data sets to the same format and map projection was time consuming and costly (Zwick, Lambert, and Carr 1998). Data layers needed to be in the same projection to overlay, view, and analyze, but projection "on the fly" did not exist at that time. Having a GIS data library with all data layers in the same file format and map projection allowed GIS users to more quickly utilize the data.

The ArcView shapefile format, developed by Esri, was chosen to distribute vector data in FGDL, because it was an openly published data format, efficient, and widely used, and supported across multiple platforms and software packages. A common raster format, Esri ArcInfo GRID, was chosen for raster data, while ERDAS/LAN and GeoTIFF formats were chosen for imagery data. In addition, a standard, statewide projection was needed for all FGDL data layers. The Albers Conical Equal Area map projection was chosen, as it covered the entire state and was the standard projection already used for FDEP data.

In 1998, the GeoPlan Center released FGDL Version 1. Sixteen state and federal government agencies contributed data to the initial development, including the FDOT and the Florida Department of Environmental Protection, both of whom provided funding. Initial FGDL data was distributed via CD-ROM, with one county per CD-ROM. Online data distribution methods were investigated, but Internet download speeds at the time were too slow for practical use. After the initial release of FGDL, subsequent FGDL data was released in versions, which allowed snapshots of the library to be distributed efficiently.

By 2005, Internet download speeds became faster and more reliable; hence the Metadata Explorer Web application became the primary distribution mechanism for FGDL data. The Metadata Explorer, developed by Esri, is driven by an ArcIMS Metadata service that stores FGDC-compliant metadata and allows users to search, discover, and download data through keyword searches. An updated Web portal with data visualization has recently been developed and is in the testing phase; live roll-out is scheduled for summer 2015. The new portal was built using Open Geoportal, and is a collaboratively developed, open-source Web application for searching, displaying, and downloading geospatial data.

The Efficient Transportation Decision Making Process

In 1999, Congress passed the Transportation Equity Act for the 21st Century (TEA-21), which included provisions for environmental streamlining. That legislation prompted the state of Florida to develop a more efficient transportation planning and environmental review process, named the efficient transportation decision making (ETDM) process, which protects the built and natural environments. To facilitate the ETDM process, a Web application named the Environmental Screening Tool (EST) was designed to

support agency participation and community involvement throughout the life cycle of each transportation project. The EST provides tools to input and update information about transportation projects, perform GIS analyses, gather and report comments about potential project effects, and provide information to the public. The user community includes staff from seven FDOT district offices, twenty-six metropolitan planning organizations (MPOs), approximately twenty-six state and federal resource agencies, and the general public.

Central to the Environmental Screening Tool (EST) are GIS analyses of the potential effects of proposed transportation projects on the human and natural environments. These analyses are used to prescreen transportation projects and inform stakeholders of potential problems. In 2001, the Geo-Plan Center began working with the FDOT to support the development and implementation of ETDM and the EST. The GeoPlan Center and FGDL were logical partners, with the GeoPlan Center's experience managing geospatial data and the wealth of Florida data ready to use in the FGDL. With the FGDL as its foundation, the GeoPlan Center began building an environmental GIS database from which hundreds of analyses could be run. The FGDL website was chosen as the official mechanism through which EST analysis layers would be distributed to the public. In addition, data used in the EST needed to be archived and retrievable for a period of up to 30 years, per FDOT legal requirements.

In December 2004, the EST web application went live for use. As transportation projects were reviewed and analyzed using FGDL data, the data themselves came under scrutiny and evaluation for currency, relevance, and completeness. With input from ETDM stakeholders, data gaps were identified, and GeoPlan staff worked to close those gaps. Periodic data review and feedback was necessary to ensure that the data layers included in the EST analyses were the most appropriate and current for each agency reviewing the transportation projects.

Challenges of Managing the FGDL

As the popularity of GIS grew in the 1980s, the number of GIS databases created by agencies and organizations grew rapidly (ESRI 1990). Obtaining data from individual state, federal, and local agencies was time consuming; coordinated data sharing was needed to more efficiently distribute data. This need was addressed by establishing the National Spatial Data Infrastructure (NSDI) and an online National Geospatial Data Clearinghouse to promote sharing of geospatial data. Since then, geospatial data clearinghouses have experienced rapid growth and development, as well as challenges in meeting the demands of users and rapidly changing technology (Kelly, Haupt, and Baxter 2008, 802). As the FGDL became a popular place for people to get Florida GIS data, FGDL became recognized as the statewide vector geospatial

data clearinghouse. Other data clearinghouses (including Geospatial One-Stop and its predecessors), library websites, and GIS centers started listing FGDL as the primary source of Florida GIS data.

DATA MANAGEMENT AND CURRENCY

With the heightened exposure came heightened expectations from users for the most current and accurate versions of data. At the same time, data originators were producing more data sets, which resulted in greater numbers of data layers being processed for FGDL. FGDL Version 1 started with eighty layers sourced from sixteen data originators. By the fourth version, the number of layers released had tripled to 250 layers sourced from approximately thirty-seven data originators. FGDL experienced the growing pains of keeping up with the demands for more data and the expectation that the data be the most current version available. During this time of expansive growth, data originators were checked ad hoc for the availability of new and updated data layers. Hence, some data layers were not released to FGDL in a timely fashion, and lag times existed between when a data layer was published by the data originator and when those data were released via FGDL. Feedback from data users included their confusion about which version of the data was most current, while feedback from data originators indicated their frustration about outdated versions of the data being hosted on FGDL.

METADATA STANDARDS COMPLIANCE

FGDL has always required metadata for all data layers, for which an in-house format (FGDL format metadata) was created. A considerable amount of time was spent creating FGDL format metadata because metadata from the source data originator was often incomplete or absent. Metadata creation is a complicated and time-consuming process that has been considered by some data creators as a low priority and a necessary evil (Batcheller 2008, 387). The FGDC (the interagency group that assists in implementing the NSDI), recognized the importance of standardized geospatial metadata for data sharing and distribution and hence began offering financial assistance for metadata-creation efforts. In 2001, the GeoPlan Center was awarded an NSDI Grant to convert FGDL format metadata to the FGDC's Content Standard for Digital Geospatial Metadata, which was the then-current national standard for documenting geographic data. This was a massive effort, which included learning the FGDC metadata format and then writing FGDC-compliant metadata for over 250 FGDL data layers.

METHODOLOGY

Development of FGDL's Quality Assurance/Quality Control Process

After its inception in 1998, FGDL experienced rapid growth and demand for its data. As a result, the number of data layers processed and distributed via FGDL increased. During that time, FGDL experienced challenges with managing large numbers of data layers, data currency, growing data size, efficient distribution methods, and compliance with documentation standards. To address these challenges, a systematic process was developed for compiling, standardizing, updating, and distributing FGDL data layers. The FGDL quality assurance/quality control (QAQC) process was designed and developed to meet these needs.

Goals for the QAQC process were to provide a systematic, replicable method to:

- standardize data to FGDL protocols (including map projection, metadata, and file format)
- perform QAQC procedures on the data
- provide opportunities for peer review of the data
- facilitate timely updates to data layers
- avoid duplication of effort
- communicate the status or progress of data
- track data lineage

Because of the complexity of the process, a relational database management system (RDBMS) was chosen to store, manage, and relate information about the QAQC process. Much like the benefits of centralizing disparate GIS data into one data clearinghouse, the information regarding the data itself would benefit highly from a centralized database that could store, track, and facilitate the various workflows. The database design was driven by the goals of the QAQC process. This methodology section will discuss the QAQC process and the accompanying database schema.

The QAQC process consists of four stages known as (1) get, (2) examine, (3) load, and (4) enter (see Figure 1). The process is executed by the QAQC team, which is comprised of a QAQC data manager, a GIS database administrator, two senior GIS analysts, and multiple junior GIS analysts. The status of work completed by team members is communicated through e-mail and reinforced through personal communication to avoid duplication of effort. As data pass through each stage, different procedures are performed to check data quality, modify the data to meet established standards, document data lineage and modifications, and publish the data for public download. Each stage of the process has a corresponding table in the database and each data layer has a unique record in the database. The database schema is updated and viewed using a system of interactive Web forms accessible
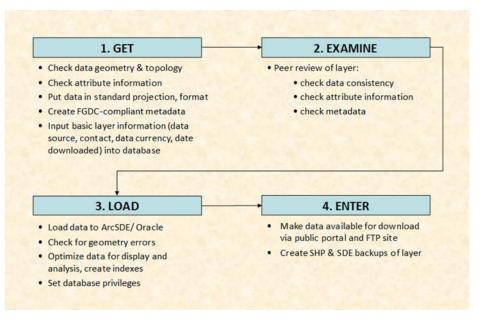
FIGURE 1 Core stages of the QAQC process.

through an internal intranet. The Web forms not only capture the pertinent information from each stage, but also serve as a reminder of the steps necessary to complete each stage. The database triggers e-mail notifications at various points in the QAQC process, which communicates the status of the data layers being processed and helps avoid duplication of effort among the QAQC team members. In addition, the database enforces constraints defined by the database administrator, which assists in standardizing the data layers.

The get stage includes physical acquisition of the data and standardization of the data so it meets FGDL protocols. In the examine stage, the data is reviewed for consistency by a more senior analyst. In the load stage, the data is loaded into the spatial database (ArcSDE) and any loading issues are documented. In the enter stage, the data is published to the FGDL website for public distribution, and backups of the data are created. Peer review is built into subsequent stages of the QAQC process to catch errors or potential issues. The QAQC process and database are designed so that each data layer is tracked for its entire life cycle: from acquisition through the QAQC process, while it is a publicly accessible data layer for download and analysis through ETDM, and finally to its deprecation.

The distinction between new and updated layers is integral to the QAQC process. New layers are data layers that have not yet been included in the FGDL. Updated layers are those that have been included in FGDL before and are being updated with different spatial features, attributes, or a significantly different data structure. Each FGDL release of an updated layer is

considered a "version," which captures a unique snapshot of the entire data layer. This versioning method does not capture feature-level changes, since most of the data are not actively changing on a frequent (daily or weekly) basis. Database records of each layer snapshot or version are related through unique identifiers that allow the QAQC team to quickly evaluate update frequencies of the data and which data characteristics have changed. These database relationships also allow the QAQC team to track the lineage of the data layers. The next section discusses the details of each QAQC stage and its corresponding database components.

THE GET STAGE

Most data acquired for FGDL is downloaded from various state and federal government agency websites. Hence, the first step in the QAQC process is the physical acquisition of the data and their processing, which is done in the get stage. During this stage, the data layer is inspected, projected, standardized, and documented according to FGDL protocols. It is important to note that while most modern GIS software packages utilize projection-on-the-fly technology, FGDL data are still projected to Albers to ensure consistent geoprocessing results and reduce processing overhead from both desktop and server applications. The get stage begins when a junior GIS analyst is either assigned a data layer or identifies a layer that needs to be updated. The analyst locates the data, which are usually on a website or FTP site, but also can reside on a disk (DVD/CD-ROM). The analyst downloads a copy of the data and saves it in their "downloads" folder on the shared network drive, which is accessible by the entire QAQC team. That original copy of the data will remain unchanged. Results of subsequent processing of the data are saved in another location, the analyst's "data output" folder on the shared network drive.

The analyst then performs QAQC procedures to standardize and check the data, including the following:

- Project the data to the FGDL Albers projection
- Check geometry and topology
- Check features for consistency and accuracy
- Check attributes for completeness, spelling errors
- Rename the data layer to FGDL filename format: THEME_MONYY (Ex: SCHOOLS_FEB14, which would represent school locations published February 2014)
- Creation of FGDC compliant metadata
- Addition of DESCRIPT field to describe features
- Addition of FGDLAQDATE field to record the date the data were acquired or downloaded from the source

Information about the data layer is saved in a database table via a Web form that the analyst fills out after completion of the get steps. Captured in the database is basic information about the layer such as the file name, data source (originator), publication date, analyst working on the layer, when the layer was downloaded, feature type, file format, scale, and QAQC status of the layer. Other pieces of information captured in the database are essential for tracking currency, updates to the layers, and data lineage; in particular:

- Date the data was last checked for update: This is the date the data was received or when the source website was checked. In some cases, an update to the data is not available, but the date checked is still documented
- Update cycle of the data (if known): This indicates how often a data layer is updated by its source and assists GeoPlan in scheduling checks for updated data
- Field changes: For data layers that are updated regularly, field name changes or attribute value changes are documented. Tracking these field changes allows for consistency between older and newer versions of the data, especially for those fields used for analysis in ETDM
- Record and feature changes: For updated data layers, changes in the number of records and/or features are documented. Tracking these changes allows for consistency between versions of the data and allows data users to analyze how features change between versions. Some changes in the data may be an indication of errors in previous versions or increased accuracy in newer versions.

During the get stage, each data layer is assigned a primary key (unique identifier) in the database. This unique ID is used to track the data layer's progress throughout the QAQC process. After the get form is submitted, an e-mail notification, triggered from the database, is sent to the QAQC team notifying them that the get stage has been completed for that layer.

THE EXAMINE STAGE

After the get stage is the examine stage, in which the data are peer reviewed. The get and examine stages are always done by two different QAQC team members to ensure a fresh perspective on the quality of the data. The examine stage is completed by a senior GIS analyst, the examiner, who performs a basic QAQC check of the data to ensure that they are consistent with FGDL standards. The metadata is also reviewed and checked for consistency with GeoPlan and FGDC standards. In addition, the examiner has access to the information submitted in the get form and reviews that information as part of the examine stage. If there are problems with the data layer, the examiner

is able to send it back to the get stage for the analyst to further investigate or correct.

The data checks performed in the examine stage are saved in a corresponding database table, which is updated through a Web form. Additional pieces of information captured in the examine Web form include layer access (public or restricted), geographic extent, whether the data are to be used in ETDM, and specific changes to fields and features of data layers used in ETDM. This information is used to administer and distribute the data and identify any inconsistencies with data used for analysis in ETDM. The data layer is assigned a foreign key (unique identifier) in the examine database table. The foreign key corresponds to the primary key in the get database table, and the relationship is enforced through database constraints. After the form is submitted, an e-mail notification, triggered by the database, is sent to the QAQC team notifying them that the examine stage has been completed for that layer.

The Load Stage

In the get stage, the data are thoroughly checked and in the examine stage, the data are peer reviewed. The next step is the load stage, in which the data are loaded into the ArcSDE spatial database and finalized for public distribution. The load stage is completed by the QAQC data manager. The data loaded into the database is the final snapshot of the data that will be released to the public for download via FGDL and used to analyze proposed transportation projects through the ETDM process. Any loading errors or issues that might arise due to geometry errors or anomalies are rectified and documented. Also during this stage, an AUTOID field is created for each layer, which is a unique identifier for tracking individual features. After the data are loaded to SDE, privileges are set for data access, depending on whether they are public or restricted, and the data are optimized for display and query through the creation of attribute and spatial indexes. Many of the data layers used in the ETDM process are spatial views, which are stored queries of other SDE data layers. Spatial views allow for custom subsets of one or more data layers without duplication of data storage. If spatial views related to the data layer are needed, they are created at this stage after the data layer is loaded to SDE.

The steps performed in the load stage are saved in a corresponding database table, which is updated through the load stage Web form. This form includes several reminders to prompt the QAQC data manager of special tasks or issues associated with the data layer. In particular, if spatial views are associated with the data layer, then the Web form displays a reminder to create the needed spatial views. If the data layer is an updated layer, then the information saved in the miscellaneous comments of the prior version of the data layer will be displayed on the form. The miscellaneous

comments field is an important field used to communicate special circumstances, peculiarities, or abnormal issues with data layers. Sometimes data issues are discovered when a layer is being used for ETDM analysis and a geoprocessing error occurs. To rectify the issue, additional steps might be needed (perhaps dissolving features, repairing geometry, etc). These issues are noted in the miscellaneous comments field and displayed throughout the Web forms and in e-mail notifications to the QAQC team. These notifications alert the team about past issues with the data layer that may recur with the current version. In addition, when completing the load stage Web form, the QAQC data manager has access to the information submitted in both the get and examine forms for review.

After the load stage Web form is submitted, the data layer is assigned a foreign key (unique identifier) in the load database table. The foreign key corresponds to the primary key in the get database table and the relationship is enforced through database constraints. After the form is submitted, an e-mail notification, triggered by the database, is sent to the QAQC team notifying them that the load stage has been completed for that layer.

THE ENTER STAGE

The last stage in the QAQC process is the enter stage. During this stage, the finalized data is made available for download, and backups of the data are created. A zipped copy of the data is placed on the FTP site for public download, and the metadata is published to the FGDL metadata server, the current public distribution mechanism for FGDL data. The metadata record in the FGDL Metadata Explorer contains a link to the corresponding zipped data on the FTP site. A shapefile version of the data is copied to GeoPlan's in-house library. Shapefile and SDE export backups of the data are also created and copied to a backup server. Metadata accompanies the data in each location including both extensible markup language (XML) and hypertext markup language (HTML) formats. Using Esri ArcCatalog software, the metadata is created in XML, a robust and structured format for storing and transferring data. In this case, the data stored in the XML file is the information about the geospatial data. The XML metadata is then exported to HTML format, the common language for describing Web documents, which can be accessed through a Web browser.

The steps performed in the enter stage are saved in a corresponding database table, which is updated through a Web form. The data layer is assigned a foreign key (unique identifier) in the enter database table. The foreign key corresponds to the primary key in the get database table and the relationship is enforced through database constraints. After the form is submitted, an e-mail notification, triggered by the database, is sent to the QAQC team notifying them that the enter stage has been completed

for that layer. After completion of the enter stage, the data layer receives a new database record and primary key in another table, which tracks all current, published FGDL data layers. This database record is linked to the Metadata Explorer database schema, which is used for public data download.

Integration of the QAQC Process with ETDM and EST

To ensure that ETDM transportation projects are reviewed and analyzed using the most current and relevant data available, tracking of data currency and regular data review have become a critical component of the QAQC process. Agencies that review transportation projects give input on the types of data needed in the environmental screening tool (EST) to evaluate impacts to their resource of concern. For example, the U.S. Fish and Wildlife Service may request that data layers identifying strategic habitat conservation areas be included in EST analyses so that they can evaluate whether those habitat areas, which are an agency priority, might be affected by a proposed transportation project. Feedback from the agency reviewers regarding data issues, data layers that need to be deprecated, and update schedules is communicated to the QAQC data manager via e-mail to keep the data relevant and current. Parts of the QAQC process were designed to capture information regarding update cycles and currency and then communicate information to the data users. The flow chart in Figure 2 displays an overview of the entire QAQC process, including the core stages of the QAQC process (get, examine, load, and enter), the public outlets of the data (FGDL and EST), and the internal and external feedback and review that drive the data update schedules. Some components in the flow chart will be discussed in the next section. When data are published to FGDL and ETDM, key fields from the QAQC database are exposed to the users to clearly indicate when the data were downloaded or received, when the data were added to FGDL, and when data were available for use in ETDM analyses. To facilitate data review and feedback by the QAQC team, additional Web forms and views into the database were created. This section will discuss those forms and views.

Data Queue

Much like a to-do list, the data queue (see Figure 3) is a listing of data layers that need to be checked for updates. The data queue is updated through a Web form, in which the QAQC data manager inputs basic information about layers that need to be checked for potential updates based on their respective update cycles or on feedback received from agency reviewers regarding data issues. Data layers in the data queue are then assigned to a GIS analyst, who checks each data layer's source for an updated version and if available, continues the data layer to the get stage. If the GIS analyst
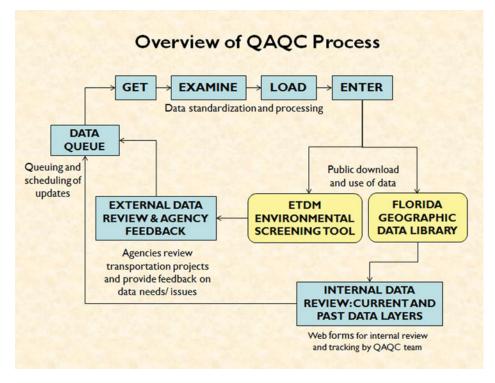
FIGURE 2 Overview of QAQC process.

finds that there is no updated version of the data layer available, then the data layer does not continue to the get stage, but instead the data layer's corresponding record in the database is marked with the date checked. Data layers can also be added to the data queue through the current data layers form, which will be discussed in the next section.

CURRENT DATA LAYERS

The current data layers Web form serves as the centralized location for the QAQC team to access information about all published data layers currently available in the FGDL. These windows into the database display basic layer information with the option to drill down and access more detailed layer information and QAQC process steps completed. Some of this information is contained within the data layer's metadata document, but there are other pieces of information that are important to tracking internal QAQC workflows and data lineage.

The current data layers form is used to facilitate periodic data review by the QAQC team, that can view data layers through a variety of filters. For example, QAQC team members often use "filter by data source" to see all data layers originating from a particular data source. Another useful fil-

QAQC DATA QUEUE

The following layers need to be processed. Click on any layer for more information or to sign up to work on that layer.

Filename	Descriptive Layer Name	Data Source	Priority	Assigned By	Date Assigned	Assigned To	Select
GC_BRIDGES_AUG13	Bridges in the State of Florida	University of Florida GeoPlan Center	5 - Medium	Kate Norris	04-JUN-13		Select
SCENIC_BYWAYS_JUL13	Florida Scenic Highways and Byways 2013	Florida Department of Transportation (FDOT)	5 - Medium	Kate Norris	05-JUN-13	Kate Norris	Select
GC_SOCIALSERVICE_DEC13	Social Service Centers 2013	University of Florida GeoPlan Center	5 - Medium	Kate Norris	05-JUN-13		Select
GC_LASERS_DEC13	Laser Facilities in Florida - 2013	University of Florida GeoPlan Center	5 - Medium	Kate Norris	05-JUN-13		Select
OASDI_2011	Social Security (OASDI) Beneficiaries in the State of Florida by Zip Code - 2011	University of Florida GeoPlan Center	5 - Medium	Kate Norris	05-JUN-13	Kate Norris	Select
POWERLINES_2014	Power Transmission Lines for State of Florida 2014	University of Florida GeoPlan Center	5 - Medium	Kate Norris	05-JUN-13		Select
CELLTOWERS_2014	Cellular Communication Towers in Florida - 2014	University of Florida GeoPlan Center	7 - Low	Kate Norris	05-JUN-13		Select
GC_DAYCARE_JAN14	Registered Daycare Facilities in the State of Florida - 2014	University of Florida GeoPlan Center	7 - Low	Kate Norris	05-JUN-13		Select
ENTERPRISE_ZONES_AUG13	Enterprise Zones in Florida - August 2013	Executive Office of the Governor / Office of Tourism, Trade & Economic Development	5 - Medium	Kate Norris	05-JUN-13		Select
USACE_DAMS_2011	National Inventory of Dams - Florida 2011	U.S. Army Corp of Engineers (USACE)	5 - Medium	Kate Norris	29-AUG-13	Kate Norris	Select
USACE_BND_2011	Army Corps of Engineers Permitting Sections in Florida - 2011	U.S. Army Corp of Engineers (USACE)	5 - Medium	Kate Norris	29-AUG-13	Danny Downing	Select

Add a layer to the queue

Current List of ETDM/ FGDL Layers

Return to QAQC Forms Main page

FIGURE 3 QAQC data queue.

ter can be set to show only layers that need an update. Since the database tracks the update cycle of each data layer and the date when the data layer was last checked for an update, a simple filter was created to compare these two database fields. For example, if a data layer is updated quarterly and more than three months have elapsed since the date last checked, it is likely that an update is needed. In this case, that data layer would be displayed if the filter to show only the layers that need an update is applied. After one or more filters are applied, the resulting table will display basic layer information including file name, description, source, source date, update cycle, date the layer was last checked, layer access, and more (see Figure 4).

If the QAQC team member wants more information about a particular data layer listed in the filtered table, then they can click on the filename, which will open a second page ("GIS Layer Information") displaying more detailed information about the data layer (see Figure 5).

The GIS Layer Information page lists the internal lineage for the selected data layer, so the analyst can investigate and determine whether the layer needs to be checked for update. The pieces of information included here are important for tracking the data layer throughout the QAQC process and

after, when the data is available to the public via FGDL and ETDM. Important dates, associated with data layer milestones, are displayed on this page:

- Date Added to FGDL: date when the data layer was published to FGDL
- Date QAQC Completed: date when the QAQC process was completed for the data layer
- Date Last Checked for Update: date when the data layer was last checked for update
- FGDL Acquisition Date: date when the data layer was downloaded or acquired
- SDE Load Date: date when the data layer was loaded to ArcSDE database
- EST Live Date: data when the data layer was available for ETDM analyses (via the EST)

The GIS Layer information page is particularly useful for the QAQC team when embarking on a data layer update. Before processing a data layer update, the GIS analyst needs to read the metadata and review the database records of the old version(s) of the data layer. This page includes

Current Data Layers Used in ETDM or FGDL

Filter by Data Source: BAR

Filter by QAQC person: All

☐ Check here to only see layers needing update

Submit

Return to QAQC Forms Main page

Click on the filename of any layer to see more information about that layer.
Click on the column headers to sort that column (first click will sort ascending, second click will sort descending).

FILENAME	DESCRIPTION	SOURCE	SOURCE DATE	UPDATE CYCLE	LAST CHECKED	ACCESS	FGDL	ETDM	WHO_QAQC
SHPO_ALL_SITES_JAN15	Archaeological Sites in Florida - January 2015	BAR	2015-01	Quarterly	14-JAN-15	sensitive	N	Y	Sam Palmer
SHPO_NATL_REGISTER_JAN15	National Register of Historic Places in Florida - January 2015	BAR	2015-01	Quarterly	14-JAN-15	sensitive	N	Y	Sam Palmer
SHPO_BRIDGES_JAN15	Historic Bridges in Florida - January 2015	BAR	2015-01	Quarterly	14-JAN-15	public	Y	Y	Sam Palmer
SHPO_CEMETERIES_JAN15	Historic Cemeteries in Florida - January 2015	BAR	2015-01	Quarterly	14-JAN-15	public	Y	Y	Sam Palmer
SHPO_RES_GROUPS_JAN15	Resource Groups in Florida - January 2015	BAR	2015-01	Quarterly	14-JAN-15	public	Y	Y	Sam Palmer
SHPO_STRUCTURES_JAN15	Historical Structure Locations in Florida - January 2015	BAR	2015-01	Quarterly	14-JAN-15	public	Y	Y	Sam Palmer
SHPO_SURVEYS_JAN15	Field Survey Project Boundaries and Attributes in Florida - January 2015	BAR	2015-01	Quarterly	14-JAN-15	public	Y	Y	Sam Palmer
									Record Count = 7

FIGURE 4 Current data layers web form.

GIS LAYER INFORMATION

Filename (Metadata Link)	SHPO_ALL_SITES_OCT14
Description	Archaeological Sites in Florida - October 2014
Data Source/ Originator	BAR
Data Source Website	dhr.dos.state.fl.us/bar/index.html
GET Last Downloaded From	
Source Date	2014-10
Data Access	sensitive
Feature Type	Polygon
File Format	Shapefile
Data Scale	1:24,000
Misc Notes	This layer is sensitive, dont forget to notify fdot. (Aja Stoppe and David Turley)

UPDATE STATUS	
Date Added to FGDL	2014-11-04
Update Status	2014 - Updated
Old Layer Name	SHPO_ALL_SITES_JUL14
Is name consistent w/ old layer?	Y
Number of changed records	95
Date QAQC Completed	29-OCT-14
Update Cycle	Quarterly
Date Last Checked for Update	21-OCT-14
Update Notes	

ADD TO DATA QUEUE

MARK LAYER AS CHECKED

FGDL/ ETDM Status	
ETDM Layer?	Y
ETDM Analysis Layer?	Y
FGDL Layer?	N

SDE LOAD INFO	
SDE Load Date	27-OCT-14
SDE Load Type	ArcCatalog 9.3.1
Loaded by	Sam Palmer
SDE Load Notes	N/A
Feature Dataset	sensitive
Dynamic Layer	N

LAYER IDs	
PK_FGDL_ID	7640
FK QAQC Theme ID	6562
SDE Registration ID	
SDE Layer ID (etdmsde)	

QAQC INFO	
Get Who	Sam Palmer
FGDL Acquisition Date	20-OCT-14
Data Contact	ftp
Original Filename	FloridaSites.shp
Local file path	Q:\data_output\sam\shpo\201410
Are all fields defined?	Y
Total records with no descript?	10
Metadata Author	Sam Palmer

FIGURE 5 GIS layer information.

information regarding the QAQC process steps completed on the data layer, where the data layer was downloaded or received from, where the data layer is stored locally, what QAQC member processed the data layer (in case questions arise), and any particular inconsistencies or issues to be aware of with the data layer, many of which are captured in the Miscellaneous Comments field.

Once the GIS Layer Information page has been reviewed by a QAQC team member, he or she may choose to look for an updated version of the data layer or add it to the Data Queue for another analyst to check. The GIS Layer Information page includes two forms for facilitating both of these options. The first is a form for adding a data layer to the Data Queue. This is typically done by the QAQC data manager, but GIS analysts also have permissions to execute this task. When adding a data layer to the Data Queue, the form automatically populates some information, and the QAQC team member inputs the rest. Sometimes specific instructions are included for acquiring the data layer and other times the instructions are simply to check whether an updated version of the data layer is available.

The second form included on the GIS Layer Information page is "Mark Layer as Checked". Typically an analyst will complete this form if they have done the research on a data layer and found that no update is available from the data source. On completion and submittal of this form, the database triggers two actions: (1) an e-mail notification, which is sent to the QAQC team relaying that the layer has been checked, by whom, and the date; and (2) an update to the "date last checked for updates" field with the current date for the data layer in the database.

PAST DATA LAYERS

Past data layers can be viewed using similar Web forms (see Figure 6) as the Current Layers forms. These forms are driven by a database table that contains a unique record for all historic data layers in FGDL and ETDM. This table is large and will continue to grow. The information that is displayed in the Past Data Layers forms is very similar to the Current Data Layers, but includes some additional fields:

● Date Deleted: The date the data layer was deleted or deprecated from use.
● Delete Status: Indicates whether the layer has been replaced by an updated version of the data layer or deleted without replacement

Being able to generate a list of past data layers is vital for managers that need to report back to stakeholders that have questions about data layers from previous project analyses. If a dispute arises about a past or present

C. Goodison et al.

Old Layer Info

Filename (Metadata Link)	SHPO_ALL_SITES_JUL14
Description	Archaeological Sites in Florida - July 2014
Data Source/ Originator	BAR
Data Source Website	dhr.dos.state.fl.us/bar/index.html
GET Last Downloaded From	
Source Date	2014-07
Data Access	sensitive
Feature Type	Polygon
File Format	Shapefile
Data Scale	1:24,000
Misc Notes	This layer is sensitive, dont forget to notify fdot.
UPDATE STATUS	
Date Added to FGDL	2014-08-11
Update Status	2014 - Updated
Old Layer Name	SHPO_ALL_SITES_APR14
Is name consistent w/ old layer?	Y
Number of changed records	117
Date QAQC Completed	28-JUL-14
Date Last Checked for Update	21-JUL-14
Update Notes	
Delete Status:	Replaced by SHPO_ALL_SITES_OCT14
Date Deleted:	05-NOV-14
FGDL/ ETDM Status	
ETDM Layer?	Y
ETDM Analysis Layer?	Y
FGDL Layer?	N
SDE LOAD INFO	
SDE Load Date	28-JUL-14
SDE Load Type	ArcCatalog 9.3.1
Loaded by	Sam Palmer
SDE Load Notes	N/A
Feature Dataset	sensitive

FIGURE 6 Past data layer info.

layer, detailed information about the layer is tracked and accessible in the Current or Past Data Layer forms. If the dispute cannot be resolved using information from the Web forms, then the original data layer likely needs to be investigated. Locating the original copy of the data is an easy task, as a link to it is stored in the database.

RDBMS AND SPATIAL DATABASE SOFTWARE

The original QAQC database schema was built using MySQL and the database forms were written in Perl. When the GeoPlan Center began working with
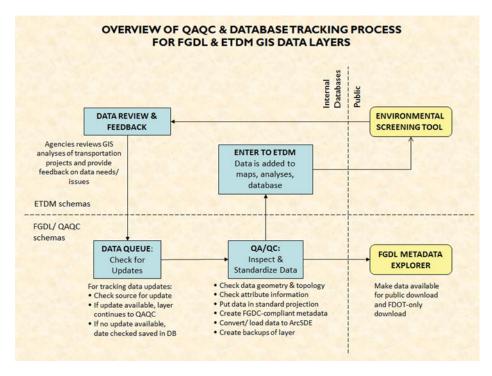
FIGURE 7 Overview of QAQC and database tracking process for FGDL & ETDM GIS data layers.

the FDOT to implement ETDM and the EST, the database was migrated to Oracle for the RDBMS, and Esri ArcSDE was chosen as the spatial database format. A spatially enabled RDBMS was needed to store transportation project information, GIS data, and GIS analyses, and communicate information to stakeholders. During the development of the EST, various software configurations were tested. A configuration using Oracle RDBMS with ArcSDE on a Unix system was chosen as it offered the best combination of performance, compatibility, stability, cost, and ease of deployment. Since the QAQC team utilizes Esri ArcGIS desktop software for QAQC work, ArcSDE was a logical extension to the server level. The ETDM GIS servers are located at the University of Florida and managed by the GeoPlan Center. Additionally, the Web forms for updating the QAQC database are written using PHP, a server-side scripting language. PHP was chosen because it integrates well with html and is relatively simple to learn.

It should be noted that there are multiple database schemas used for the FGDL QAQC process and ETDM/EST. While the schemas are related through primary and foreign keys, the business processes and workflows for each are different (See Figure 7).

DISCUSSION

SUCCESSES OF THE QAQC PROCESS

The QAQC process has been in use for over fifteen years. The core stages of the process have remained the same, while the intricacies have changed and evolved to meet the needs of FGDL and ETDM users, geospatial data and metadata standards, and changing technologies. For example, changes to the process and database were made after GeoPlan began working with FDOT on the ETDM process to accommodate better data tracking and updates. The early challenges with FGDL have largely been addressed through the implementation and enhancement of the QAQC process and database. The process has created a more efficient workflow that saves time and reduces confusion, while improving data organization, management, currency, and confidence in data quality.

One particular component of the QAQC process that has significantly improved the FGDL workflow is the tracking of update cycles of individual data layers. This tracking was implemented after ETDM went into production (late 2004) and has significantly streamlined the data update process and yielded time savings for the QAQC team. Instead of checking the source data in an ad-hoc manner, the QAQC team knows exactly when to look for an update. The database also contains instructions on where to obtain the data, along with any special instructions regarding the data. This coordinated scheduling of data updates has allowed the QAQC team to release data more frequently and in a timely manner that more closely matches when the data originators publish the data.

The increased data update frequency has resulted in fewer inquiries about the currency of data. While this reduction is difficult to quantify, the QAQC team can evaluate anecdotally through inquiries and feedback with FGDL and ETDM users. In the past, data currency was a primary concern amongst the data inquiries and feedback from users, in particular from the source agencies whose data was being distributed through FGDL. Data currency is no longer a primary topic in the data inquiries. While the QAQC team still receives occasional inquiries regarding data currency, the version of the data in question is almost always within the already planned update schedule.

The QAQC process has also enabled the QAQC team to more effectively manage FGDL's growth. As mentioned in the background, the first growth increases were not well managed. FGDL started in 1998 with 80 data layers, tripled after two years, and quadrupled to 330 by the fourth year. Following this initial explosion, FGDL maintained a minimum of 350 data layers, while steadily growing. Since 2011, the library has stabilized at about 450 layers. In these recent years, the QAQC team is not necessarily processing more data
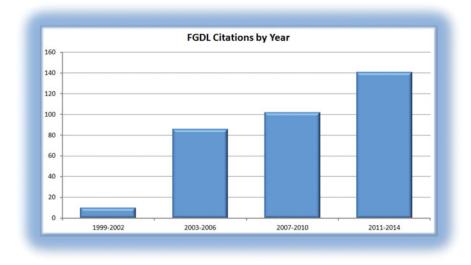
FIGURE 8 FGDL citations by year.

layers, but rather doing a more efficient job of keeping the existing library up to date. Approximately 30% of FGDL's data layers are updated annually, with many layers being updated quarterly.

Recently, some FGDL data was independently reviewed by an outside entity and this offered a good opportunity to evaluate the library's data confidence and currency. The FDOT hired a private consultant to comprehensively review seventy-four environmental data layers in FGDL/ETDM to ensure that the data was current and appropriate for use in the Environmental Screening Tool. Of the seventy-four layers reviewed, sixty-two represented the most current version of the data and twelve needed to be updated with a more current version. All twelve layers needing an update were already scheduled for update or were in the process of being updated by the QAQC team. The consultant identified five new data layers, which were not currently used in the EST and needed to be added for particular analyses. Those five layers were immediately added to the Data Queue to initiate the QAQC process. Using the Web forms that access and query the database, the QAQC team was quickly able to verify what data layers were in the process of being updated and respond to the consultant's comments.

The QAQC process has also resulted in more people accessing and utilizing FGDL data. While it is difficult to quantify the numbers of FGDL users past and present, two proxies can be used: FGDL citations and website activity. The first proxy is the number of researchers utilizing and citing FGDL data in their work, including academic journal articles, conference proceedings, books, technical reports, dissertations, and graduate theses. As
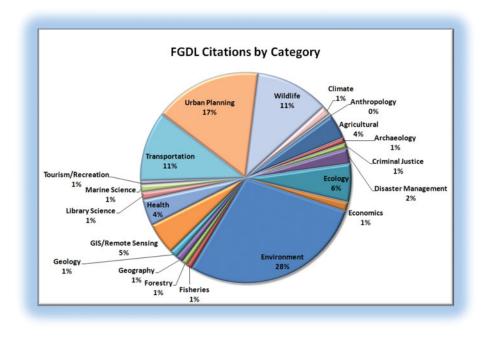
FIGURE 9 FGDL citations by category.

can be seen in Figure 8, the number of FGDL data citations has grown over the past sixteen years.

Additionally, Figure 9 displays the breakdown of FGDL citations by category of research or discipline. The diverse number of categories demonstrates a broad spectrum of FGDL users and the growing number of citations indicates increased utilization and confidence in FGDL data.

A second proxy, analyzing historical website statistics, shows an increase in hits and visits to the FGDL website. Because of Web server, ftp server, and analytic software changes, complete historical statistics of downloads and website activity since FGDL's inception is not available. However, statistics were available for a five-year period between 2004 and 2008, before and after some significant improvements were made in 2005 and 2006, namely; upgraded the Oracle database schema that included the tracking of update cycles and support for ETDM, and the implementation of the Metadata Explorer for data downloads. Analyzing the statistics for this 5-year period show a marked increase in website hits, unique visits, and pages and files requested. This increase can be attributed to better data currency and download distribution, both of which were database driven components.

The development of a systematic QAQC process, implemented and supported through a relational database system has proved to be an efficient method for maintaining and distributing a large geospatial data collection.

DATA CLEARINGHOUSES: PAST, PRESENT, AND FUTURE

Early geospatial data clearinghouses physically collected and organized data into a centralized online repository for download, typically via FTP. Modern geospatial data clearinghouses have evolved from a purely physical collection of data to a digital collection, which connects to other data repositories via metadata harvesting and Web services. A major benefit of this modern model is that it allows for metadata collections from numerous data nodes to be quickly and efficiently aggregated.

The FGDL clearinghouse model still implements a physical collection of geospatial data, which is by design. First and most importantly, FGDL data serves as the foundation for FDOT's ETDM application. Not only is the data physically loaded into a database designed for analyzing the effects of proposed transportation projects, but the data is first vetted and standardized to ensure the accuracy of the resulting analyses. These analyses are used specifically to inform the transportation planning process and the GeoPlan Center is contracted to use its expertise with geospatial data to determine the fitness of the data. Inaccurate, incomplete, and inconsistent data undermines the validity and credibility of decisions made using that data. While it is nearly impossible to ensure that all data layers are 100% accurate, the experience of the QAQC team has caught geometry errors, projection issues, and other inconsistencies that could cause feature shifts and result in inaccurate GIS analyses for the ETDM process. Also, the physical repository of data is needed to meet FDOT's data retention policy, where data used in ETDM projects must be retrievable for a period of up to 30 years.

This growing repository must be managed over time to meet changing technologies and formats. Currently, ArcSDE and shapefile backups are maintained, as ArcSDE is the spatial database used and shapefile is an open format supported across multiple geospatial software packages. Because of the retention policy, naming conventions have been established, great emphasis has been put on documentation, and database schemas have been designed to facilitate data layer tracking and lineage.

From the outside, the QAQC process and database developed for managing FGDL may seem cumbersome and complex. From the inside, the process has yielded more efficient workflows and the database components have facilitated data input and information retrieval. The process detailed in this article could serve as a model for others managing large geospatial data collections. Furthermore, maintaining standard documentation and a relational database can greatly contribute to geospatial data preservation and archiving.

HISTORICAL DATA

As GIS has become a mature and widespread technology, GIS data have proliferated and many repositories have accumulated large collections of

data. With the exception of land use and census data, only current versions of
data layers are typically distributed in FGDL. However, past versions of data
layers have great potential value for documentation of features or conditions
at specific points in time, change analyses, and more. In particular for ETDM,
an agency reviewer may need to see the specific version of a data layer that
was used for an analysis on a certain date. Access to historic data layers is
important, but the QAQC team has not yet tackled the task of distributing
these data on a mass scale. It is not by chance that the design of the QAQC
database, the file naming conventions, and the physical warehouse of data
layers support an infrastructure for distribution of historic data. The primary
obstacle at this point is funding for additional hardware needs and personnel
time to facilitate the data distribution. The future of data clearinghouses
should include historical archives of data and the QAQC team hopes this
will become a funding priority for agencies utilizing geospatial data.

Lessons Learned

Managing a geospatial library for over fifteen years has yielded an efficient
workflow, quality vetted data layers, and many lessons learned. Here are
some lessons learned for others managing large geospatial data collections:

- Not all data are created equal. While there is a need to develop tools to
  assist with systematic identification of data issues, there is no substitute for
  an experienced GIS analyst manually examining the data.
- Standardize file names. Some naming pointers include these:
  ○ Use unique file names for every piece of data
  ○ Use file names that at minimum include "what" (theme) and "when"
    (publication date) identifiers. When possible, include a "where" identi-
    fier. Because the majority of FGDL covers the entire state of Florida, "FL"
    is not typically included as an identifier in our case.
  ○ Keep file names consistent between different versions of the same data
    layer, but change the date identifier to indicate the publication date.
    Do not use the same exact filename for updated data layers. A clearly
    and uniquely named file will reduce user confusion when reconciling or
    comparing different versions of data.
  ○ Examples of file names:
    ○ PAR_CITYLM_2011—City Limits Derived from 2011 Florida Parcel Data
    ○ PAR_CITYLM_2010—City Limits Derived from 2010 Florida Parcel Data
- Capture time stamps of important data layer milestones. In particular, cap-
  turing when the data were downloaded or acquired and when the data
  were released for public download via FGDL. When dealing with large
  numbers of data layers from a large number of sources, it can be easy to
  mix up versions of data.
- Never underestimate the importance of good metadata! Although it is time
  consuming, good metadata is critical to so many issues—it is the founda-

tion of data discovery and data portals, offers credibility to data analysis, reduces user confusion, and aids in data preservation.

- When updating metadata from another data source, retain all original information and simply add new process steps.
- Retain internal metadata about your geospatial data collection. FGDL's internal metadata is the QAQC database schema which maintains a wealth of data about the geospatial data. Documenting issues (geometry errors, changes in features, etc.) over time has helped identify problems or inconsistencies with the data and streamlined the workflow so data layers can be processed more efficiently and effectively. Furthermore, when issues are captured in a database, loss of institutional knowledge about the data is minimized.

CONTINUING CHALLENGES

While much has been learned, challenges still persist, including issues with data formats and open standards, varying data quality, size and storage, hardware upgrades, and data migration. Focus on these issues is needed to ensure that FGDL's geospatial data collection remains relevant for all stakeholders in a constantly changing geospatial world. Particular challenges include:

- Data formats and open standards: The growth of the geospatial industry over the past few decades has been accompanied by a growth in specialized GIS software packages. Some of these software packages use their own proprietary formats, which are not compatible with other software. This presents a challenge for the entire geospatial community, and especially data originators and clearinghouses that are distributing data to the public: in what format should data be distributed? Maintaining and distributing more than one data format can be time consuming and logistically complicated. Esri's shapefile has been the vector file format used in FGDL. While the shapefile is not an open data standard, it is an openly published data format and one of the most commonly used vector formats, supported across many software applications. However, as the size of individual data sets has grown, the capacity of the shapefile's database file (DBF) component has not been enough to store some dense data layers. Because of these DBF size limitations, some data in FGDL has been released via Esri File Geodatabase format. File Geodatabase is the recommended vector data storage format for ArcGIS users, but it is not fully supported on other software packages. While there is an API for the file geodatabase, only certain elements of the format are accessible to developers. Additional limitations of the DBF component of the shapefile include a limit on the length of field names, number of total fields, and number of characters in a field. The ability to distribute larger data sets in open standards that are compatible across software packages is a primary issue of the QAQC team.

Fortunately, the Open Geospatial Consortium (OGC), an industry consortium of companies, government agencies, and academics, is working to make geospatial information and services interoperable so that they can be utilized by a variety of applications. Following OGC's leadership and established standards will be critical to addressing these challenges.

- Maintaining backups: Even though disk space is relatively cheap, thought must be given to organizing, managing, and maintaining backups with a continually increasing geospatial data repository. Multiple backups should be maintained to protect from data loss, equipment failures, media degradation, and external events such as natural disasters that can damage hardware. Multiple physical locations of hardware are often chosen to protect from these external events. Additionally, because storage media can degrade over time and because data formats can change over time, a plan should be in place to regularly migrate backups to more current platforms.

- Data quality: Because of the widespread use of GIS and the variety of GIS software applications and file formats, data quality varies. Additionally, changes in personnel, budgets, software, and hardware at source agencies that create data can lead to changes in data processing workflows. A data layer that was previously updated frequently may not be updated anymore, a hosting website might change, a data field might disappear, or polygons may shift. These changes can lead to increased or decreased data quality. For data regularly utilized in FGDL and ETDM, the QAQC team attempts to document the changes in data quality by capturing the information in the database and the metadata. Usually the documentation is qualitative in measure, noting any differences between the current and the prior version of the data or reporting any observed, consistent shift in features.

- Hardware upgrades and data migration: Because computer hardware has a life span, hardware should ideally be upgraded before the end of its life and before the compatibility of software updates lags too far behind. Migrating large geospatial data repositories takes planning and coordination to insure maximum availability of data. Data migration is also a good time to reevaluate data storage assumptions to meet the latest best practices.

CONCLUSIONS

Managing a geospatial data library for over fifteen years has yielded an efficient workflow, quality vetted data layers, and many lessons learned. Using a systematic process to standardize data and a relational database for tracking workflows has been a critical component to successfully managing and maintaining a current and relevant data collection. The QAQC process has allowed our QAQC team to hone their expertise in Florida GIS data and hence, the FGDL has gained a reputation as the go-to source for vector geospatial data in the state of Florida. None of this would be possible without the support of the FDOT, which has invested in GIS infrastructure for their

environmental review and permitting processes. FGDL's relationship with FDOT enables this value-added, vetted data library approach, which benefits not only FDOT and its decision making, but also the public and others who use FGDL data. Funding from a state agency or another entity is crucial to maintaining a curated data repository like the FGDL.

FGDL's next generation data portal (Open Geoportal) will host Web services for Florida data layers and connect to other data collections for data discovery. However, the cornerstone of the next generation FGDL will still be the Florida data vetted by the GeoPlan Center.

ACKNOWLEDGMENTS

REFERENCES

Batcheller, James K. 2008. Automating geospatial metadata generation—An integrated data management and documentation approach.*Computers & Geoscience 34*(4): 387–398.

Clinton, William J. Executive Order 12906 of April 11, 1994. Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure. https://www.fgdc.gov/nsdi/nsdi.html

Coppock, John T., and David W. Rhind. 1991. The History of GIS. In: *Geographical Information Systems: Principles and Applications (1*st *ed.)*. Paul A. Longley, Michael F. Goodchild, David J. Maguire, and David W. Rhind (Eds.), New York: John Wiley and Sons: 21–43.

Environmental Systems Research Institute, Inc. 1990. *Understanding GIS: The ARC/INFO Method*. Redlands, CA: ESRI Press.

Kelly, Maurie C., Bernd J. Haupt, and Ryan E. Baxter. 2008. The Evolution of SDI Geospatial Data Clearinghouses. In: *Encyclopedia of Data Warehousing and Mining (2*nd *ed.)*, John Wang (Ed.), Hershey, PA: Information Science Reference: 802–810.

Onsrud, Harlan J. and Gerard Rushton. 1995. *Sharing Geographic Information*. New Brunswick, NJ: Center for Urban Policy Research.

Zwick, Paul D., J. David Lambert, and Margaret H. Carr. 1998. *Florida Geographic Data Library: The Design and Implementation of a State-Wide GIS Data Depository*. Paper presented at the annual Esri International User Conference, San Diego, CA.