The FDEP Library Catalog
Christopher L. Judd

1233 Cross Creek Circle Apt. B
Tallahassee, Fl. 32301
Email: Chris.judd@dep.state.fl.us
Phone: 850.245.8297
Fax: 850.245.8263
Florida Department of Environmental Protection
Florida State University

ABSTRACT

This presentation will discuss the current multiphase project that will make use of Information Science fundamentals to build a new interface to the Florida Department of Environmental Protection's (FDEP) library of 300 Geographic Information Systems (GIS) layers. The current interface, DataMiner, is a custom FDEP ArcGIS, Environmental Systems Research Institute's (ESRI) desktop GIS product, extension that serves as a gateway to the GIS library. This tool employs very broad categories to organize the layers and provides a pick list, a list of options for user selection. This allows layers with predefined symbology, a method of representing map features and attributes with colors and shapes, to be added to a user's ArcGIS map. The upgraded library catalog will encapsulate the browse and search features of a library catalog, allowing a more granular data taxonomy. A portion of the project will be dedicated to properly cataloging and organizing the data utilizing Information Science principles. Additionally, the upgrade will allow user-contributed layers and reference data sharing services such as ArcIMS, ESRI's Internet Mapping Server that will allow FDEP users to consume data from both internal and external entities.

Keywords and phrases: library, taxonomy, data distribution, data categorization, data organization, Florida State University, Florida Department of Environmental Protection

1 INTRODUCTION

The FDEP GIS Library Catalog is a product of a Directed Independent Study (DIS) for the Florida State University (FSU), College of Information (CI) sponsored by the Florida Department of Environmental Protection. The purpose of the project is to develop a data distribution system, based on library science principles that will allow for better search and retrieval of the FDEP's collection of disparate data. The project is a multiphase initiative, currently in its first phase; this paper will focus on providing a background of the DEP environment, present what has been accomplished in the project, discuss lessons learned and methodologies used in the process and discuss future plans. This discussion will include an overview of the project management methodologies, candidate needs, candidate requirements, and classification taxonomy.

The FDEP is a state agency employing approximately 5000 personnel. Headquartered in Tallahassee, Florida the department has numerous office locations including park sites, district offices, and a multitude of field and remote offices. There are approximately 350 GIS users in the agency that are concentrated mainly in Tallahassee, and a small number of users are sprinkled around the state (Davis & Judd, 2003). The agency utilizes GIS in decision-making, data analysis, and environmental planning capacities. GIS is employed in regulatory, conservation, and enforcement activities of the department.

The technical environment at FDEP includes a plethora of technologies. The department has two data centers outfitted with redundant power and climate control systems, both are located in Tallahassee. In the networking arena, Tallahassee users are fortunate to be connected to a fast and reliable local area network. Offices in Tallahassee are connected via a Metropolitan Area Network (MAN), which provides ample connectivity. Remote users are connected to enterprise database and application servers in Tallahassee by fractional T1 lines. The centralized GIS infrastructure includes ESRI ArcSDE, spatial data engine, and Oracle Corp.'s software for a spatially enabled Relation Database Management System (RDBMS). A cluster of ESRI ArcIMS (internet mapping servers) is provided as a resource for the agency to host ArcIMS sites and services. The ArcIMS websites currently hosted by FDEP offer a multitude of functionality, including data editing, a conduit for public information, and internal and external data analysis. A floating license manager is provided for desktop GIS clients using ESRI ArcGIS products. The FDEP enterprise vector database hosts over 300 datasets. The raster imagery database contains almost two terabytes of imagery. A strategic goal of the enterprise is to continue using Oracle spatial for the undergoing data integration efforts.

The current DataMiner is a client-server application. It is a component of the FDEP customized ArcGIS/ArcView extensions (See appendix A). The program is an interface to the agency's spatial data library that uses broad categories to organize the GIS data and provides a pick list that allows layers with predefined symbology to be added to the user's ArcGIS/ArcView map. The program was initially developed for the ArcView 3.x platform with the Avenue programming language and was ported to the ArcGIS 8.x platform and completely rewritten in Microsoft Visual Basic and ArcObjects. Aside from the porting from one platform and programming language to another, the overall functionality of the program remained the same.

The project proposes adding functionality to the next generation of the DataMiner tool. These new features include a design that will allow for exposure of controlled vocabulary and free text search of metadata. The project also plans to make enhancements, including:
- ArcGIS layer file format to support multiple symbologies for layers
- User contributed layers and symbology
- Data sharing services such as ArcIMS that will allow FDEP users to consume data from internal and external entities
- Local file based spatial data
- Support for the ArcGIS 9.x platform

2 USER NEEDS AND REQUIREMENTS

The candidate needs and requirements are formal project management deliverables for this system. These deliverables aggregate the functionality required by the users (candidate needs) and what each of the needs demand in terms of product development (candidate requirements). After the proposed candidate needs are collected and candidate requirements determined, they are submitted to management for approval. The candidate needs and requirements for this project are broken down into six categories; browse functionality, search functionality, supported data types, data manipulation functionality, data export functionality, and catalog management.

2.1 BROWSE AND SEARCH FUNCTIONALITY

The browse and search functionality of the data catalog will allow users to efficiently navigate and search data based on a series of category schemes. The users needed a way to be able to search by a combination of different criteria to yield precise results. The system also needed to accommodate public Internet users.

A requirement of the system was the ability to search the data in a number of different aspects. This required the application to be designed to accommodate searches based on geographic areas selected by the user as well as political boundaries such as districts, counties, and parks. The system shall be able to be queried by controlled vocabulary as well as free text searches of all metadata elements. The system would also need the capacity to search by the date the data was updated and the data steward.

Historically the FDEP has made data that is deemed "publicly available" accessible on an unrestricted File Transfer Protocol (FTP) server. This application would need to be able to allow public users access to download data and view metadata for these layers. This would require limiting access to restricted data and maintaining a metadata viewer for external access.

2.2 SUPPORTED DATA TYPES

The objective of the FDEP Library Catalog is to support GIS data types compatible with the ArcMap 9.x lyr file format. Data that is not compatible with lyr format, is in a floating media format (CD, DVD or tape) or is not of digital nature (such as a printed map) is out side the scope of this application and will not be included in the catalog.

The system will support the following data formats: file based vector and raster data that is compatible with ArcGIS and lyr files; data services including local and remote ArcIMS services and ArcSDE vector and raster data. Examples of data formats that will be unsupported include: non-spatial data such as database tables and Excel spreadsheets; printed maps; PDF maps; floating media containing GIS data such as CDROMS; downloadable only file based data which could be found on HTTP or FTP sites. Native Oracle Spatial layers not registered with ArcSDE will also not be supported.

2.3 DATA MANIPULATION AND EXPORT FUNCTIONALITY

The current DataMiner application incorporates the functionality to select data from a series of categories and add the specified layer to an ArcGIS map. The FDEP GIS Library Catalog will also need to be able emulate this utility. This need will borrow some of the functionality from the legacy application DataMiner, which is the basis for the FDEP Library Catalog.

An identified need of the system was to build in the capability to export data. This need will require the identification of data that is eligible for export. Proposed requirements require data to be less than 200,000 records and must be in vector format. The automated exportation of raster data is not being entertained due to technical limitations.

2.4 CATALOG MANAGEMENT

Management of the catalog will require documented procedures, protocols to communicate with the user community, and a computer system to maintain the data. Written procedures will provide a consistent method of adding, deleting, and updating records in the catalog. Printed methodologies will ensure data are classified in a specified manner, mandating these decisions to be agreed upon unanimously by two parties in a blind manner or turned over to third party for additional review. Establishing communication protocols will keep users aware of addition and depreciation of data to the library. A computer system will be required to effectively implement a dynamic library catalog. The development of this system, outlined in the candidate needs section of the project, will be essential in maintaining the currency of the catalog. The system will determine what data is exposed to internal and public data consumers.

3 CLASSIFICATION TAXONOMY

The objective of the classification taxonomy is to categorize data in order to ease retrieval of information. The current classification system used by the DataMiner application is insufficient for the large quantity of disparate data in the collection. The proposed data taxonomy allows for more granular categorization of the data and is based on International Organization for Standardization (ISO) and Federal Geographic Data Committee (FGDC) standards.

The FDEP data library contains approximately 260 datasets that currently categorizes data in the DataMiner application. The existing categorization scheme is loosely based on the different divisions in the agency, base data sub categories and a number of external sub categories [See appendix A]. Basing categories on agency division names and external business unit names is problematic for many reasons such as, a business name does not necessarily expose the type of data offered in the category and some division names (e.g., Water Resource Management, could lead the user to believe all data pertaining to water would be put in the category).   The current categorization has provided the following "base map data" subcategories: boundaries, imagery, indexes, water, and transportation. The categories are too broad and lack enough description for the data administrator to decide where the data is best placed. Additionally, many of the data overlap into the different categories, this should be avoided with a good taxonomy "Everything has a place and only one place." (Pack, 2002, p. 28).

The proposed data taxonomy will be based largely on the MetroGIS data categories (MetroGIS, Twin City Metropolitan Area of Minnesota [MGISTCMAM], 2003, p. 1-3), which is based on the ISO 19115 Topical Category Definitions and the FGDC Topic Category names. The proposed data taxonomy will also borrow some ideas from the Cornell University Geospatial Information Repository (CUGIR) data categories ("CUGIR data categories", n.d.). [Please see appendix B for the data taxonomy.]

4 FUTURE PLANS

Future plans for the FDEP GIS Library Catalog project include the completion of the current phase, following that development, will be the testing and deployment of the final product into the user community. At the time of writing this paper, the completion of the current and first phase is near. Outstanding deliverables for this phase of the project include a high level list of implementation technologies and project wrap-up.

The next phase of the project will spawn subsequent DIS and will include additional project management deliverables. A low level Detailed Resources Specification (DRS) will give a road map for the application developers to begin materializing the plans and designs into software. This development will include programming the creation of web-services and database objects to support the application. Testing will encapsulate the creation of test plans, integration and unit testing. Before deployment, a comprehensive user manual will be developed to help orientate users to the new system.

I also plan to present this paper at the 2005 ESRI International Users Conference. A large portion of the FDEP user community belongs to the Seven Hills Regional User Group (SHRUG) and it is planned to also present some of the products of the project at this year's SHRUG workshop.

5 CONCLUSIONS

The planning of the system revealed a set of needs and requirements for the system that spanned from catalog management to functionality and user interfaces. Management of the catalog will require documented procedures, protocols to communicate with the user community and a computer system to maintain the data. A documented need of the system it the capability of supporting exporting data from the database. The application is designed to support GIS data types compatible with the ArcMap 9.x lyr file format for flexibility and functionality. The browse and search features of the data catalog will allow users to efficiently navigate and search metadata based on a series of category schemes.

The initial stages of the project show promise in providing solutions for many data organization problems. GIS users have responded well to the concept of the data catalog and are anticipating its completion. The next step of the FDEP GIS Library Catalog project includes the completion of the current phase, development, testing and deployment of the final product into the user community. Armed with improved data taxonomy, and a host of new features, users will be able to find data with greater precision and efficiency than before.

ACKNOWLEDGEMENTS
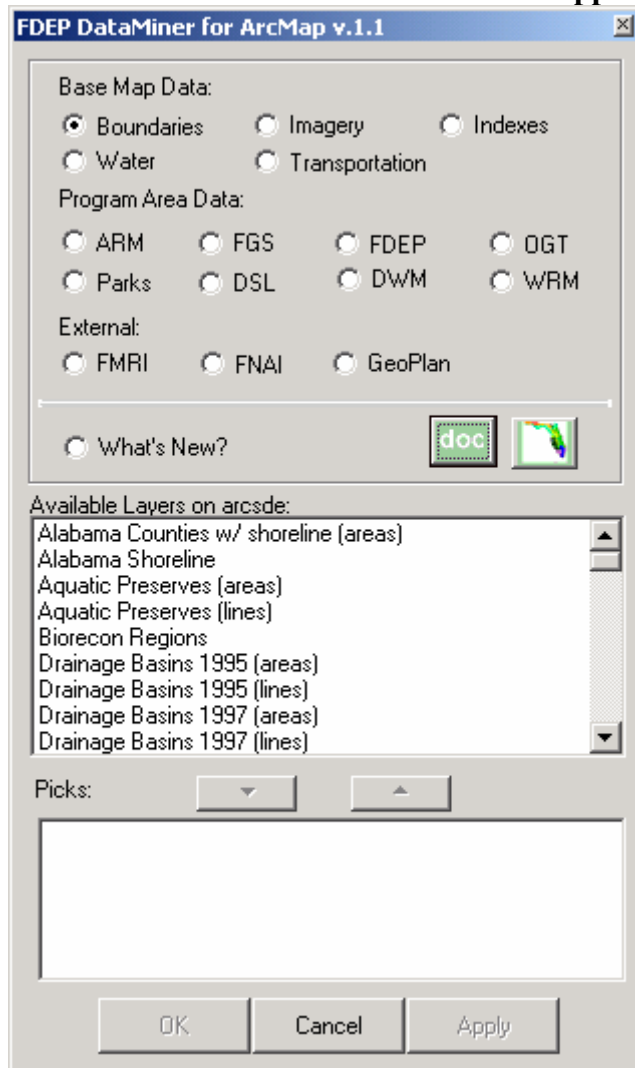
Appendix B.

* Example of FDEP Data Taxonomy

Category Name	Definition	DEP ID	Status	Keywords	ISO Code	ISO Topic Category Name	ISO Definition	FGDC Topic Category Name
Agriculture And Farming	Information related to the rearing of animals and/or cultivation of plants	010	active	agriculture, irrigation, aquaculture, plantations, herding, pests and diseases affecting crops and livestock	001	Farming	rearing of animals and/or cultivation of plants	Agriculture and Farming
Biology And Ecology	Flora and/or fauna in natural environments	020	active	wildlife, vegetation, biological sciences, ecology, wilderness, sealift, wetland ecology, habitat	002	Biota	flora and/or fauna in natural environment	Biologic & Ecologic Information
Administrative and Political Boundaries	Information related to political and administrative boundaries and related non-boundary information	030	active	political and administrative boundaries, counties, cities and townships, voting districts and polling places, neighborhoods, zip codes	003	Boundaries	legal land descriptions	Administrative and Political Boundaries
Atmosphere And Climate	Processes and phenomena of the atmosphere	040	inactive	Processes and phenomena of the atmosphere	004	Climatology, Meteorology, Atmosphere	processes and phenomena of the atmosphere	Atmospheric and Climatic Data

* Based on MetroGIS data taxonomy (MetroGIS, Twin City Metropolitan Area of Minnesota, 2003, p.

References

CUGIR data categories. (n.d.). Retrieved March 18, 2005, from

http://cugir.mannlib.cornell.edu/browse_lis/browse_lis.html

Davis, D. M., & Judd, C. L. (2003, June). *The Challenges of Building an Enterprise Quality Web Mapping System for State Government.* Retrieved February 2, 2005, from

http://gis.esri.com/library/userconf/proc03/p1051.pdf

MetroGIS, Twin City Metropolitan Area of Minnesota. (2003, April 30). *DataFinder Theme Categories.* Retrieved March 18, 2005, from

http://www.datafinder.org/documents/DataFinder_ISO_Compliant_Theme_Categories.pdf

Pack, T. (2002, March). Taxonomy's role in content management. *EContent, 25*(3), 26-31.

Retrieved January 6, 2005, from Wilson Database, Article No. 200206006672003.