Developing A Geo-Spatial Search Tool Using A Relational Database Implementation of the FGDC CSDGM Model

Presented by Kit Na Goh

Introduction

- Executive Order 12906 was issued on April 13, 1994 with the following definition:
  - –A) "Geospatial data" means information that identifies the geographic location and characteristics of natural or constructed features and boundaries on earth. This information may be derived from, among other things, remote sensing, mapping, and surveying technologies. Statistical data maybe included in this definition at the discretion of the collecting agency.
- The "National Geospatial Data Clearinghouse" means a distributed network of geospatial data producers, managers and users linked electronically.
- GIS Center at ISU plays the same role as a Data Clearinghouse to facilitate data sharing.

Continued Introduction

- The GIS Training and Research Center maintains over 11,000 datasets within its spatial library allowing remote users to freely access both raster and vector GIS datasets.

- The current search engine of the GIS Center does not reveal files stored within ZIP files (the current bundling and compression method used at the GIS Center).

- We propose to develop and deploy a relational database containing geospatial metadata documentation for all datasets within the spatial library and an intelligent web interface to help the GIS Center's clients more efficiently find required geo-spatial data.

Why use FGDC standard?

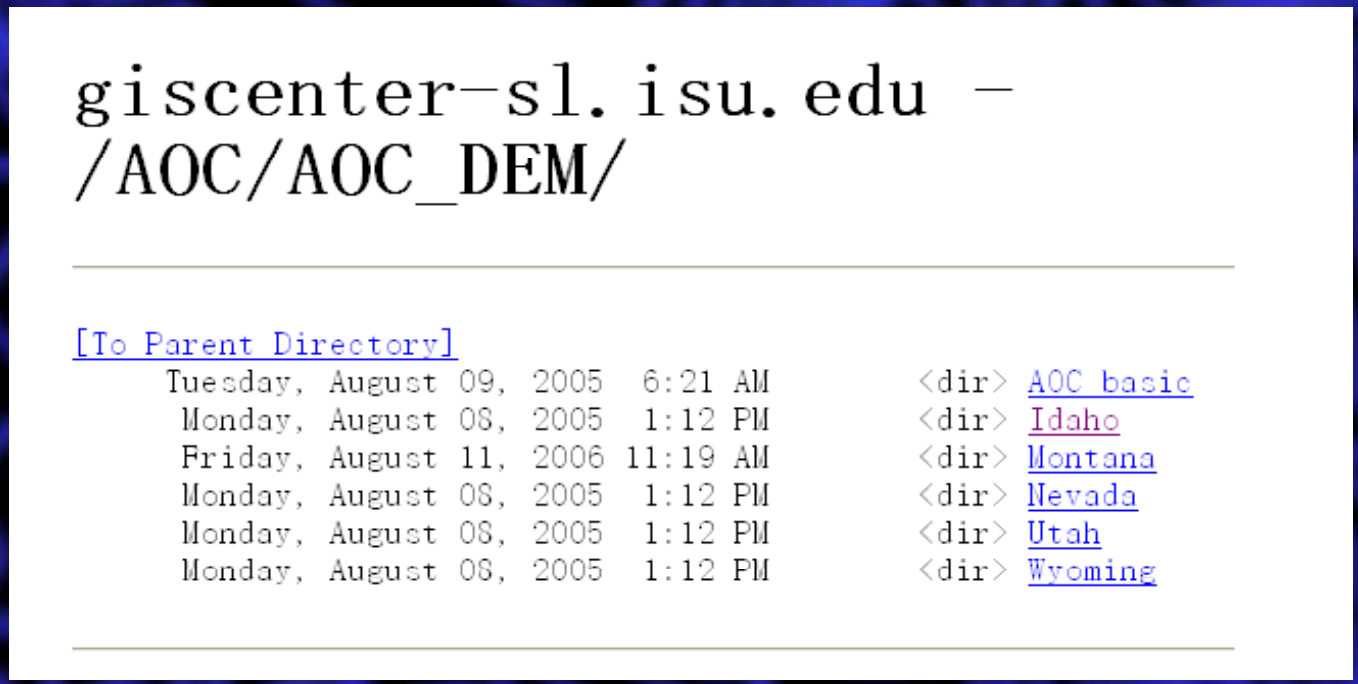
- Since June 8, 1994, The FGDC (Federal Geographic Data Committee) approved the metadata standard 'content standards for Digital Geospatial Metadata (abbreviation is FGDC standard or format).

- Therefore, the GIS Center at Idaho State University is following the standardization process in the field of digital geographic information and providing the metadata in FGDC compliant standard.

Problem Statement

- Currently, GIS TreC clients need to know the path to the datasets within the spatial library to download them.

- It is easier for professionals that are familiar with GIS metadata but maybe not meaningful to novice users.



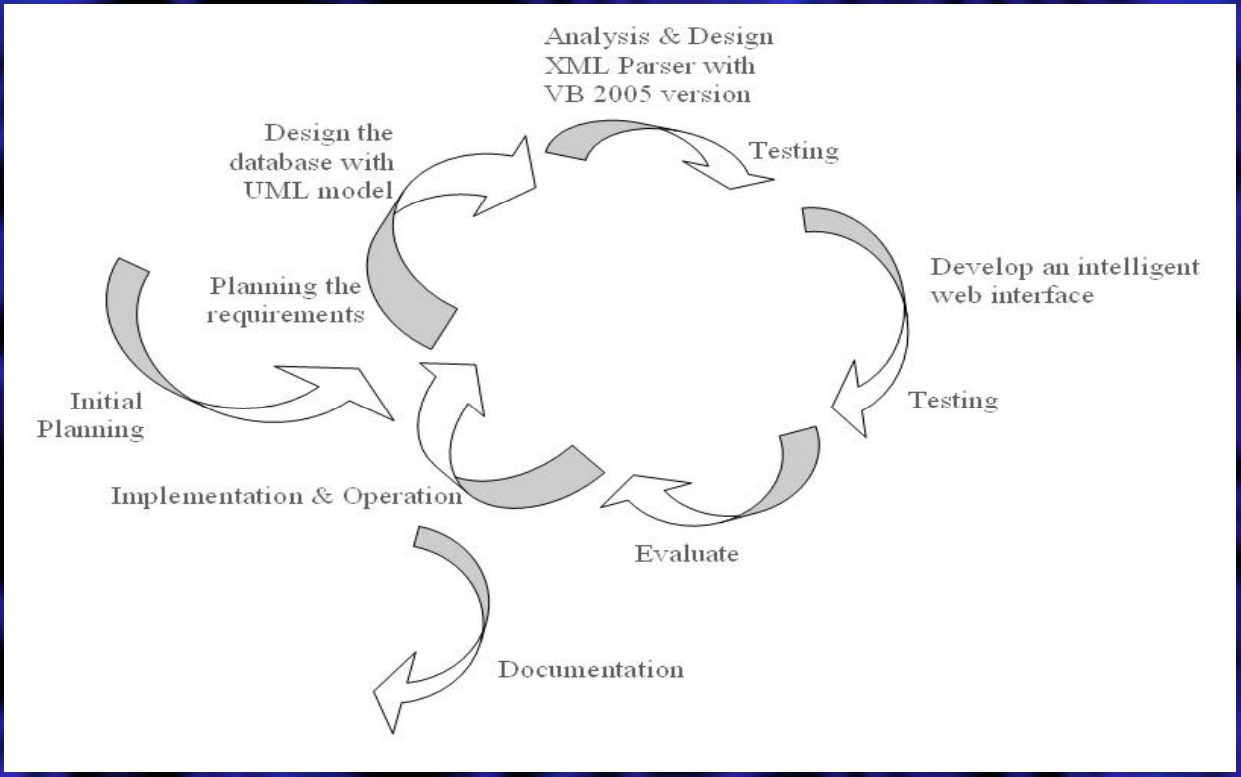
giscenter-sl.isu.edu - /AOC/AOC_DEM/

[To Parent Directory]

Tuesday, August 09, 2005	6:21 AM	<dir> AOC basic
Monday, August 08, 2005	1:12 PM	<dir> Idaho
Friday, August 11, 2006	11:19 AM	<dir> Montana
Monday, August 08, 2005	1:12 PM	<dir> Nevada
Monday, August 08, 2005	1:12 PM	<dir> Utah
Monday, August 08, 2005	1:12 PM	<dir> Wyoming

Process and Model

- The Iterative Development Model will be applied for the design and implementation of the project.
- With this process, I can learn from my mistakes, better handle user responses, re-analyze the design structure, modify and re-implement the solution.



Program Development Cycle

- 1. Understand the problem: To improve the current search engine of the GIS Center to help the clients more efficiently find required geo-spatial data and understand the structure of metadata.
- 2. Design a relational database (Rational Rose software will be used to create UML model) to capture the information within FGDC geospatial metadata documents. The database will include the name and the path to the geospatial dataset, along with rich keywords, to better facilitate search.
- 3. Create an XML parser to parse geospatial metadata and populate the relational database described above.

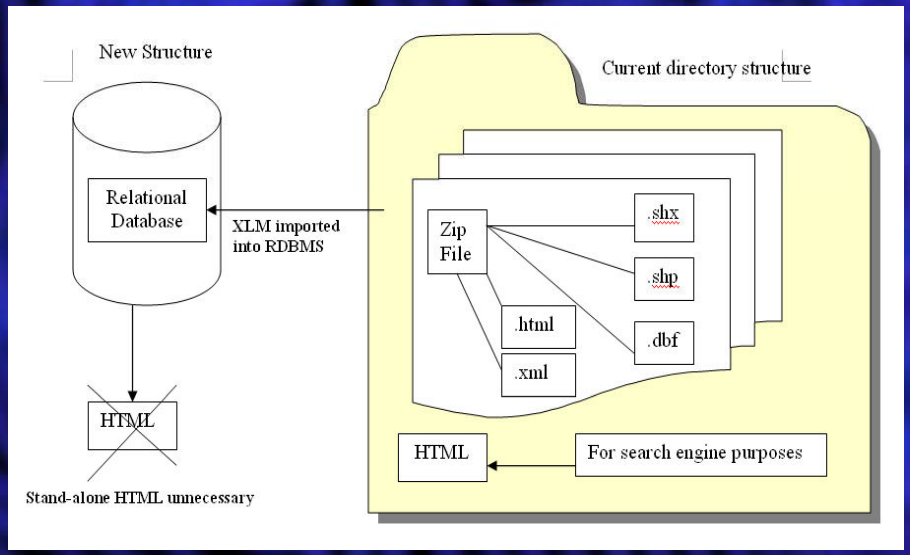
Continued …

- 4. Develop an intelligent web interface (that has the capability to query the spatial data with topology keywords, such as, intersect, contain, and adjacent and view the metadata in XML (Extendable Markup Language) style sheet before downloading) to assist our clients to search for data available in the GIS spatial library.

- 5. Organize and document all materials describing database design, XML parser design, and Active Server Page design.

Current Directory Structure

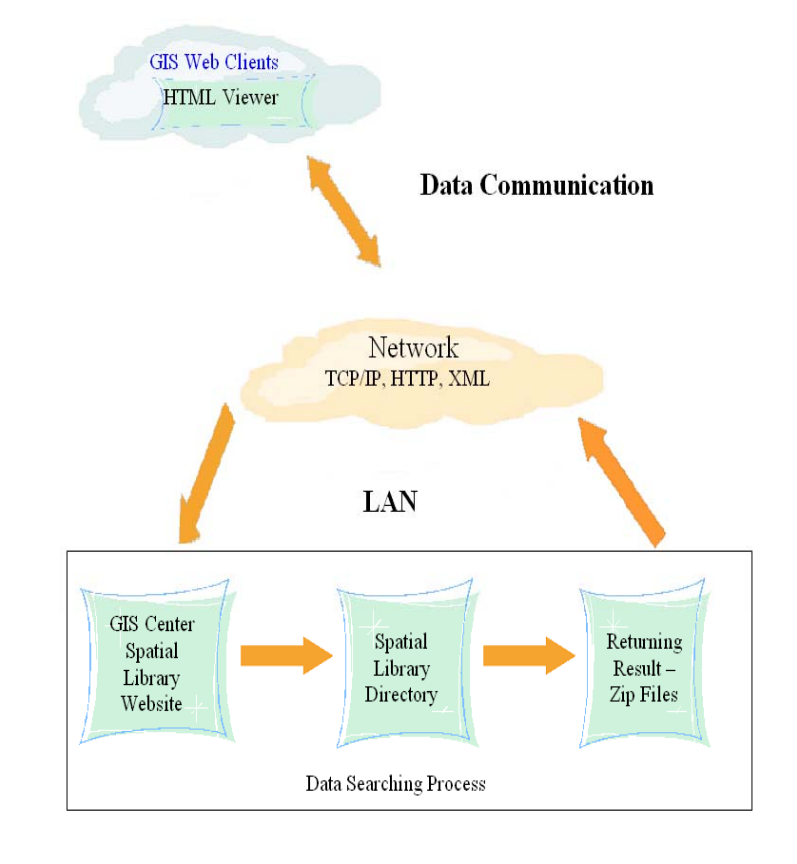
- Currently, there are zip files with HTML file(s) saved inside a folder.
- When creating and populating the relational database, HTML files can be eliminated by storing all fields and attributes from the XML files.
- ASP page will be designed to allow the users to query and view the metadata and access the geospatial objects.



How a client currently uses the spatial library to find geo-spatial data?

- Browse to the GIS Center's spatial library website and search through the spatial library's directory of folders and files to find the desired dataset.

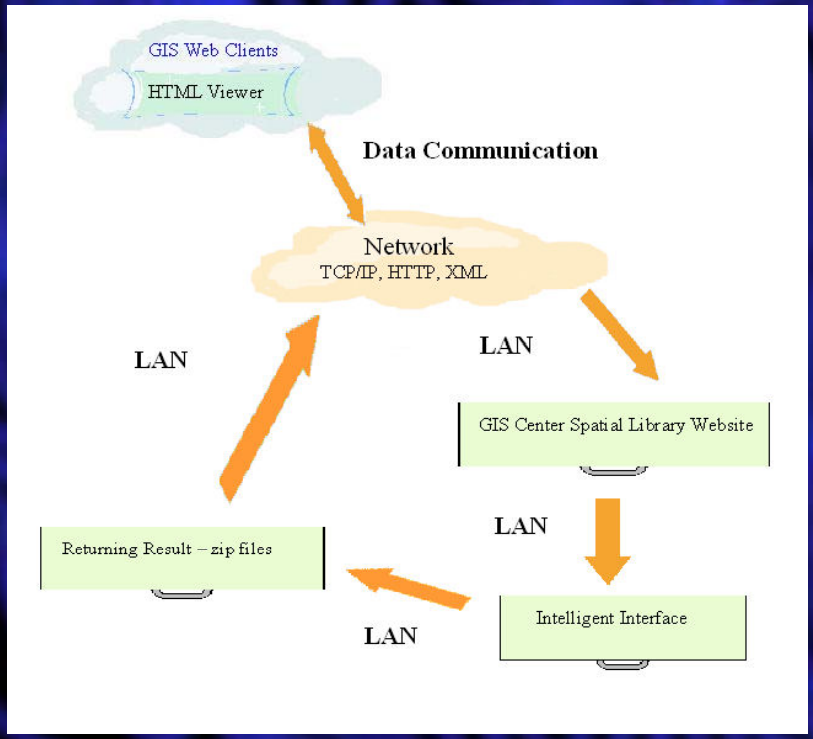
- Once found, the client downloads the data as a zip file to use on their local computer.



The Proposed Method of Data Discovery

- Through the intelligent interface on the GIS Center spatial library website, clients would enter keywords or search phrases to locate a dataset.

- The results would include a link to each dataset which will facilitate rapid download of the files to the client's local computer.



The difference between the current search process and the proposed method of data retrieval

- The clients can obtain the data by filling out the search form at the GIS Center's spatial library website with keywords and search criterion.
- It will return the results that meet the criterion with the data location link. The clients will be able to view metadata in XML format while hovering over the link.
- The clients don't need to know the path to the data files in the spatial library.

Similar Projects that are available

- Ying Teng (a previous master student of the University of New Brunswick) had developed search engines for searching ISO XML metadata files on the Web by geospatial coordinates, dates and strings with a GSDindex (geospatial data index based on R-tree and AVL trees) approach and a relational database (Oracle 8) database approach.

- She created A UML diagram of the Java class structure for R-tree construction and search and another UML diagram for GSDindex range search for her thesis.

Continued .....

- InGeoForum (Information and Cooperation Forum for Geodata in Germany) and InGeo Information Center provides a theme-overlapping web-portal to metadata servers and metadata information system.
- A theme-overlapping semantic network for geodata builds the basis for an information retrieval and a visualization component to access geodata archives.
- Topological relationships between geographic entities (specified search targets and candidates) are considered for spatial ranking and a global ranking algorithm has been developed.

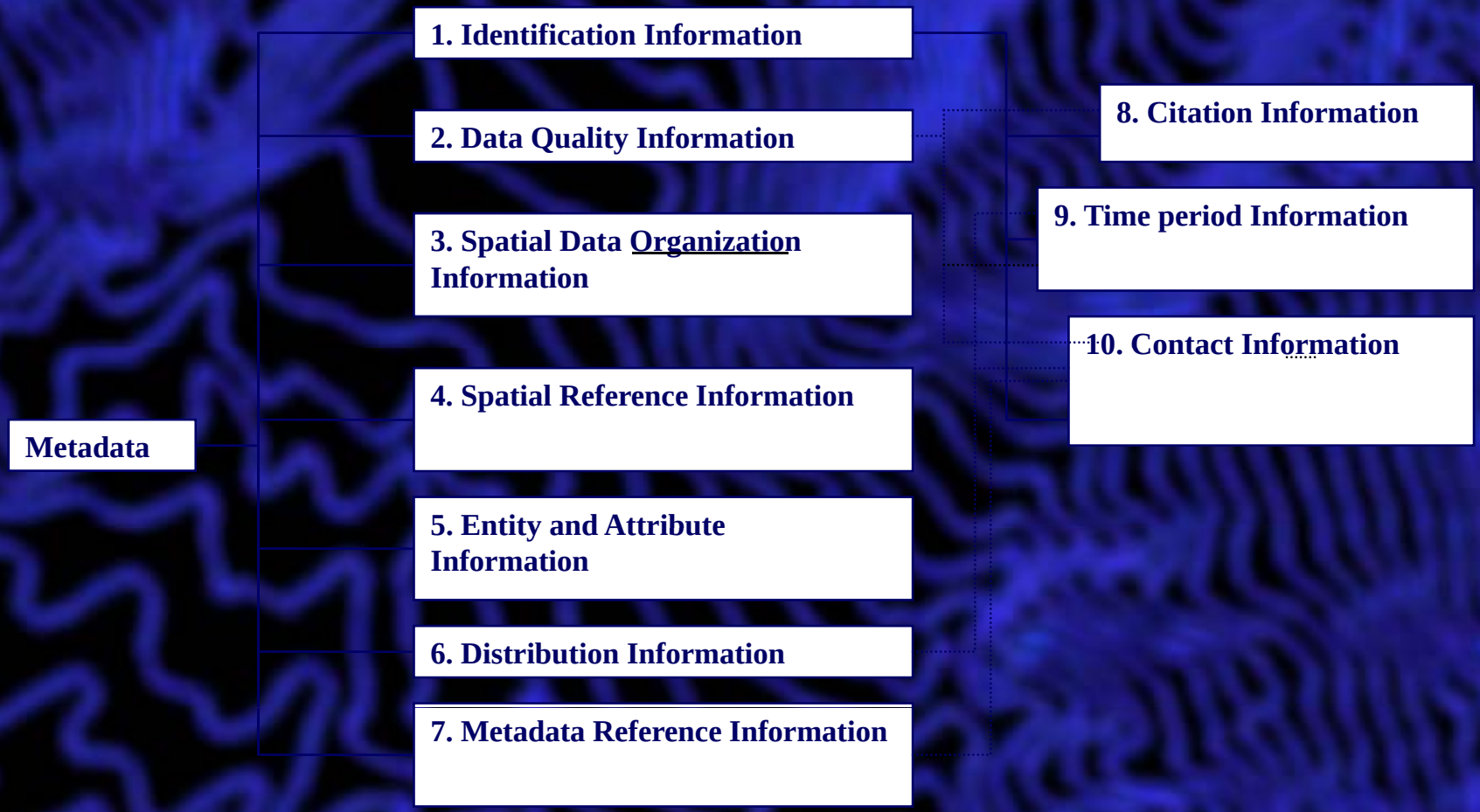
Current Database Status

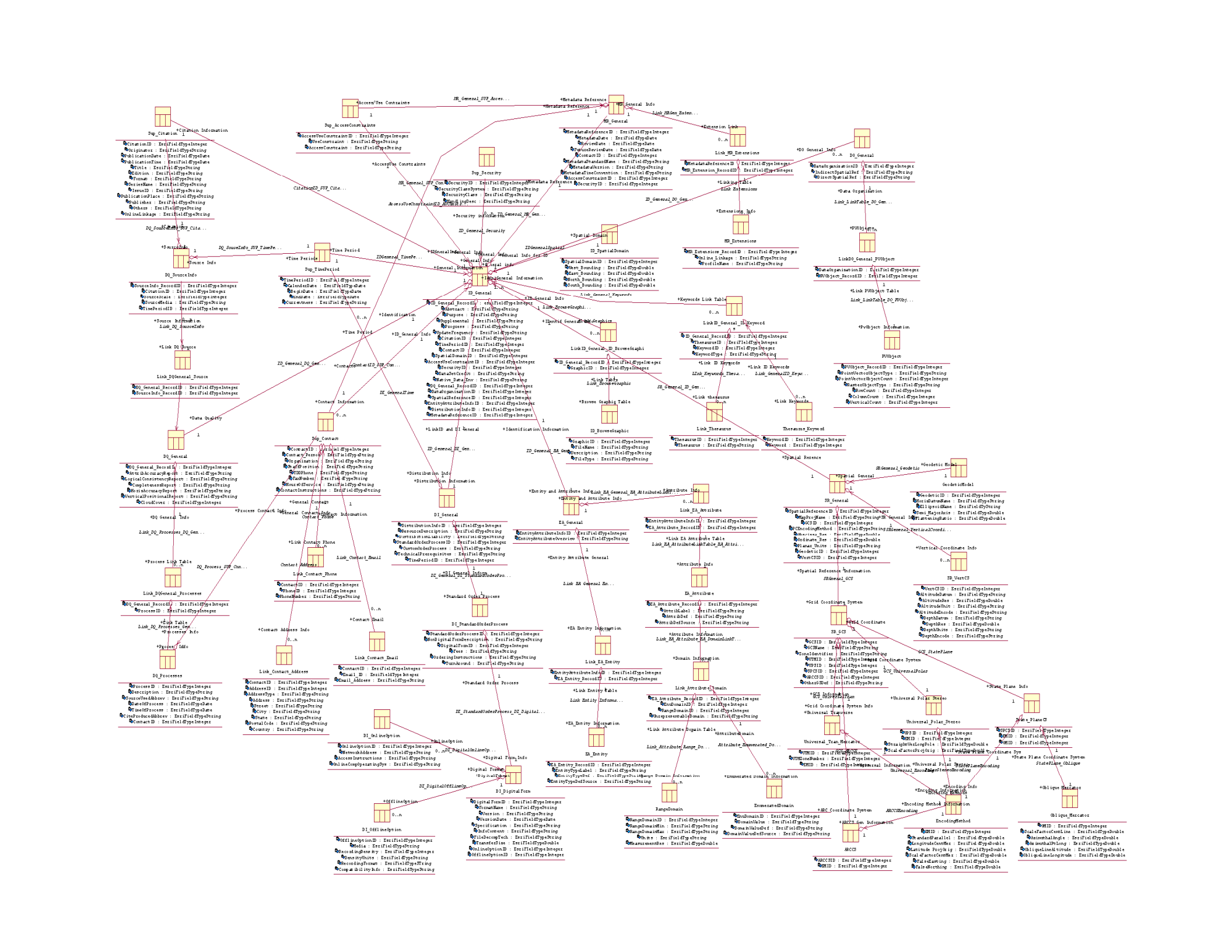
- Rational Rose Software was used to create a UML (Unified Modeling Language) to facilitate all the fields and attributes of FGDC metadata documents.

- With the UML model that has been created, a relational database (Microsoft Access) is populated from the UML CASE Model via the case schema tool in ArcCatalog.

Basic Information on Metadata

←Metadata→← Main Section →← Supporting Section →





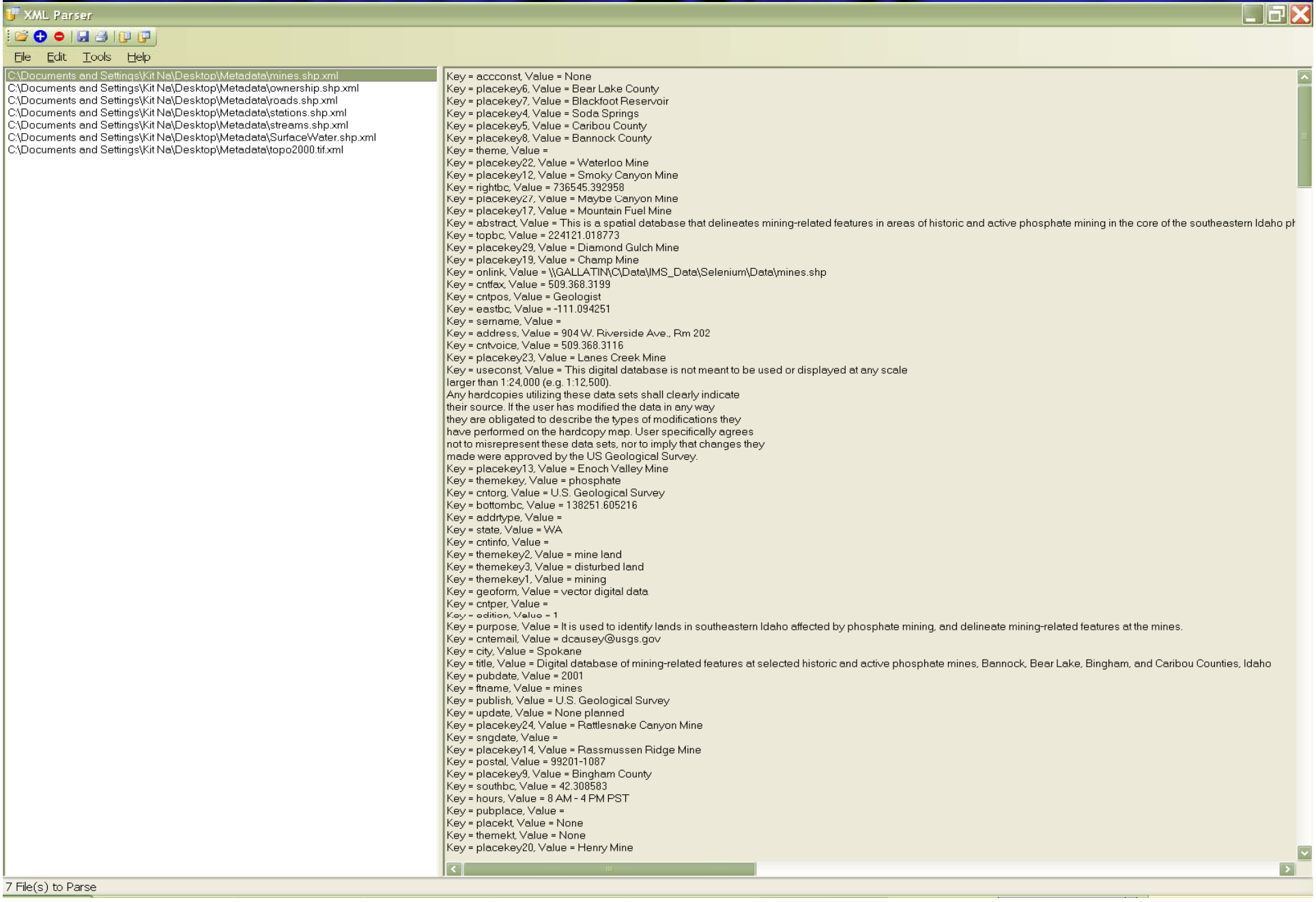
XML Parser Status

- With the XMLTextReader function in Visual Basic 2005, an XML parser application is being created to import the metadata to the relational database.

- Currently, the application is able to parse the fields and attributes and able to connect to the database.

- When the XML Parser application runs, it will store all the fields and attributes in the hash table as keys and values.

The screenshot of current parser



Future Improvement

- XML Parser will be able to write the FGDC compliant metadata to it correctly.
- ASP will be designed to allow clients to query and search the spatial data more efficiently.
- Reiterate the program development cycle to create an error free and efficient design.
- We are also considering the effectiveness of using IBM DB2 for better facilitating search if within time limit.

Conclusion

- To facilitate better discovery and delivery of geo-spatial data for the GIS TreC's clients, I propose the development of a robust relational database containing FDGC-compliant geospatial metadata.

- The database will be coupled to an intelligent web interface.

Any questions or suggestions?