

URLs in this document have been updated. Links enclosed in {curly brackets} have been changed. If a replacement link was located, the new URL was added and the link is active; if a new site could not be identified, the broken link was removed.

Optimizing Web Access to Geospatial Data: The Cornell University Geospatial Information Repository (CUGIR)

Philip Herold

Information Services Coordinator
Albert R. Mann Library, Cornell University
ph31@cornell.edu

Thomas D. Gale

Programmer/Analyst Librarian
Albert R. Mann Library, Cornell University

Thomas P. Turner

Metadata Librarian
Albert R. Mann Library, Cornell University

Abstract

With the aid of a 1997 Federal Geographic Data Committee CCAP Award, Cornell's Albert R. Mann Library recently established the Cornell University Geospatial Information Repository (CUGIR), a Web-based clearinghouse containing geospatial data and metadata related to New York State. The staff at Mann Library has established an efficient model for spatial data distribution. This paper describes the processes, problems, and solutions involved in the creation of a geospatial data distribution system.

Introduction

Libraries with collections of digital geospatial data, and those that serve and support the use of geographic information systems, are utilizing the Internet and World Wide Web as an efficient and flexible means to distribute collections of geospatial data. Creating a Web-based distribution system requires librarians to understand and address a wide array of issues surrounding geospatial data, including metadata and standards, partnerships and liability, and data organization and technical infrastructure. Although some of these issues mirror those associated with other, more commonly known and understood library resources, geospatial data contain attributes that require special attention and an understanding of both cartographic and geographic concepts.

Staff at the Albert R. Mann Library at Cornell University began looking at ways to disseminate geospatial data from Mann's collections via the World Wide Web in 1995, and in 1998 established a Web-based clearinghouse for New York State geospatial data and metadata. Building a clearinghouse entailed creating partnerships with local, state and Federal agencies, understanding how to interpret and apply the Federal Geographic Data Committee (FGDC) Content Standard for Geospatial Metadata, and designing a search and retrieval interface and a flexible, scalable data storage system. These tasks brought both anticipated and unforeseen challenges. This paper will examine the data dissemination model that Mann Library has adopted, and will explore the tasks and challenges that model has presented.

Why Build a Data Clearinghouse?

Since the release of the U. S. Census Bureau's TIGER/Line 1990 files in 1991, Mann Library has made strong efforts to support the use of geospatial data and geographic information systems by University faculty, students, and staff. In the early 1990s GIS applications were not widely used because there was a lack of available digital data covering fundamental aspects of geography, software applications were immature and difficult to use, and GIS technology was relatively new and its applications were not well known. In the past eight years all of these problems have largely been mitigated.

There remain, however, several impediments to the successful utilization of GIS and geospatial data. One difficulty is the high degree of technical understanding that accompanies using sophisticated and powerful GIS applications. A second issue is the requirement that users understand important cartographic and geographic concepts related to GIS. A third obstacle is the relative difficulty in accessing geospatial data sets required by users to complete projects using GIS. It is the third impediment that poses the greatest challenge to many libraries, because geospatial data is a specialized resource, and a relatively new addition to library collections.

Mann Library makes efforts to alleviate all three of these impediments, by offering workshops, self-paced tutorials, thorough documentation, and flexible consulting services designed to help users achieve the technical and conceptual understandings necessary to use GIS in their work and study. However, even for users with the requisite understanding, providing ready access to the geospatial data needed by Mann's users is problematic because there is a relative scarcity of geospatial data in usable digital formats. Most digital geospatial data are derived by converting existing analog map information into digital formats through digitizing, scanning, or geocoding processes. Most often, digital geospatial data are produced by local, state, and Federal government agencies, where the creation and distribution of this data is typically slow and scattershot. The result is that many fundamental data sets either do not yet exist, or are incomplete. The difficult task of libraries is to identify, acquire, and provide access to those data sets that are complete.

To provide fast, easy access to geospatial data in a well-organized fashion, Mann Library staff designed a Web-based system for data distribution. In our first attempt at this, in 1996, Mann staff worked with the Cornell Institute for Social and Economic Research (CISER) to convert parts of the U. S. Census Bureau's TIGER/Line 1992 files (Herold 1996). Six separate coverages (transportation, hydrography, and four sets of census and political boundaries) were converted for each of New York State's sixty-two counties and organized into a Web site with browsing tools, help, and non-standardized metadata. Users could select a county by name or from an image map and then download geospatial data describing that county.

The success of the New York State TIGER/Line system served as an impetus to develop an expanded and improved Web-based service. In 1997, Mann Library was awarded a one-year grant from the FGDC's Competitive Cooperative Agreements Program (CCAP) to build a clearinghouse node as part of the National Spatial Data Infrastructure (NSDI) Federal Geospatial Clearinghouse. The FGDC's CCAP program is designed to provide seed money (up to \$40,000 in 1997) to institutions that undertake one of several types of initiatives towards building, on a local, regional, or national level, the infrastructure for creating, distributing and sharing geospatial data or standards.

Mann Library's clearinghouse node is one of over 90 such nodes located around the world (most located in North America), containing searchable metadata records describing geospatial data sets. All nodes are located on data servers using either the Z39.50 or a compatible information retrieval protocol. As a result, they can be linked to a single search interface called the Geospatial Data Clearinghouse (Federal Geographic Data Committee Geospatial Data Clearinghouse Entry Points) where the metadata contents of all 90 nodes, or any subset in combination, can be searched simultaneously. In addition, most clearinghouse nodes have their own Web sites and customized browsing and searching interfaces.

The CCAP program requires funded agencies to establish partnerships with outside agencies. Mann Library, which services Cornell's College of Agriculture and Life Sciences, College of Human Ecology, and Divisions of Biological Sciences and Nutrition, is primarily interested in working with agencies that produce and own geospatial data related to agriculture, environmental sciences, and selected social sciences. We approached the New York State Department of Environmental Conservation, the owner of many key data sets related to agriculture and the environment, and the Cornell Soil Information Systems Laboratory, where soil survey maps are currently being digitized from analog media, about forming data sharing partnership agreements.

In developing an NSDI Clearinghouse Node, Mann Library and its partners proposed to the FGDC the following objectives:

- to establish and manage a National Geospatial Data Clearinghouse Node; to be accessible remotely, both through the Cornell University Library Web site and online catalog and through the NSDI Clearinghouse;
- to inventory, document and provide access to geospatial data holdings of Mann Library, the New York State DEC, and the Soil Information Systems Laboratory, in accordance with existing FGDC-endorsed Content Standards for Digital Geospatial Metadata;
- to develop a plan to acquire, disseminate, and assist in the development of new data products in the agricultural and environmental sciences; and
- to create an on-line, ANSI/ISO Z39.50-compliant database of geospatial metadata; to be browsable and searchable (by fields, coordinates, and free-text keywords or phrases).

Cornell's Clearinghouse Node would serve to further NSDI objectives by:

- participating as a node of the National Geospatial Data Clearinghouse;
- providing standardized documentation of data adhering to FGDC Content Standards for Metadata;
- initiating data collection and sharing within the State of New York; and
- providing standardized means of on-line access to geospatial metadata and data utilizing a platform-independent information retrieval protocol.

The development of CUGIR has been accomplished through a team-based model of work and cooperation. Project staff were selected from each division within Mann Library, including Public Services, Technical Services, Collection Development and our Information Technology Section. The primary working group consisted of five regular members, each coordinating work within his or her area of specialty. Other Library staff participated on an as-needed basis. Primary responsibilities for the overall coordination of clearinghouse development were held by a Public Services Librarian with significant experience using and advising in the use of geographic information systems and geospatial data.

Data Definition, Identification, and Preparation

To develop a clearinghouse node or data repository, developers need to define the nature of the data to be collected and create a plan to develop or acquire that data. We began by creating a working collection development policy that established the criteria for data selection. In creating this document the working group addressed a number of issues that would prove essential to developing a clear collection scope and would create an identity for and give purpose to the clearinghouse. The document also serves to define the philosophy of CUGIR, specifically regarding issues of use and access. There are no downloading access restrictions imposed on CUGIR data. CUGIR collects and makes available geospatial data and metadata describing the agricultural, environmental, biological, and social characteristics of New York State. CUGIR also collects data described by the FGDC as "Framework" data -- data that has wide applicability in spatial analysis and use of geographic information systems. This type of data includes transportation, hydrographic, soils, elevation, cadastral, and other commonly used data themes. It is our intention to make data available in the most widely used data formats and in multiple formats when feasible, given the limitations and restrictions of existing resources. The collection development policy clearly states these guidelines and is included on the CUGIR site.

Once CUGIR's scope was clearly defined, staff identified data sets for preparation, documentation, and inclusion in the clearinghouse. We received an inventory from NYSDEC and met with representatives in June 1997 to discuss plans to create metadata for, and select for inclusion, several data sets at the state, county, and 7.5-minute quadrangle levels. We also received a status report from SISL indicating that several counties and quadrangles were in progress with several others awaiting Federal certification. We also created an inventory of data sets in Mann Library's collections that met criteria for inclusion. Documentation and preparation of data sets to be included were prioritized, with Mann Library holdings placed directly after NYSDEC data.

Data preparation was one of the more significant activities and accomplishments of the CUGIR team. Although most data sets coming to CUGIR from agencies outside Mann Library were in the agencies' native formats and required no conversion, there was a significant amount of data conversion that took place in-house. An Arc/Info programmer was hired to perform the conversion of raw TIGER/Line 1995 data into both Arc/Info coverage (which was packaged in Arc/Info interchange (export) format for distribution) and shapefile formats. This programmer converted eleven coverages, including roads, railroads, hydrography, landmarks, and county, minor civil division, place, census tract, census block group, census block, and unified school district boundaries. The coverages were developed for each of New York State's 62 counties (a total of 682 unique geographic themes) in two formats (a total of 1,364 files derived from TIGER/Line 1995). The shapefiles were then archived using UNIX tar and compressed using the public domain software Gzip (GNU zip). Similar geospatial data processing was carried out for several USGS-produced framework-level Digital Line Graph (DLG) small-scale themes for New York.

It should be noted that the data conversion was performed in a way that is both scalable and replicable. Arc/Info AML (Arc Macro Language) scripts were created to automate conversion processes and run them on batch files. These need only be rerun to regenerate the same types of files in the future, and we anticipate that they can be used to convert data from the Census Bureau's 1997 release of TIGER files. AMLs created for CUGIR's data sets will be shared with others who wish to do their own conversion of TIGER. It should also be noted that conversion is a complicated and time-consuming process. It required considerable amounts of time and energy to create AMLs that ran successfully, and to include in them data improvements such as the creation of keycode fields (concatenations of FIPS codes identifying unique polygons) for census designated areas including block groups and blocks

Standards and Metadata

Sharing geospatial data over the World Wide Web involves communication between remote users and data providers. For this process to be successful, metadata and information retrieval standards serve as the couriers between these groups and are central to the entire process. Basic information about geospatial and other forms of data should answer the following questions. What information is produced? Who created it? For what purpose was it created? What was the process used to create it? This data about data is called metadata (Federal Geographic Data Committee Metadata). In June, 1994, the Federal Geographic Data Committee (<http://www.fgdc.gov/>) established a metadata standard for describing geospatial data that can be used to assist in this process. The Content Standard for Digital Geospatial Metadata defines a minimal set of information that must be recorded about all geospatial data as well as an optional list of more detailed information. The FGDC describes the standard as a method for communicating between producers and users:

"The standard was developed from the perspective of defining the information required by a prospective user to determine the availability of a set of geospatial data, to determine the fitness of the set of geospatial data for an intended use, to determine the means of accessing the set of geospatial data, and to successfully transfer the set of geospatial data. As such, the standard establishes the names of data elements and compound elements to be used for these purposes, the definitions of these data elements and compound elements, and information about the values that are to be provided for the data elements." (Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata).

The Content Standard defines seven basic types of information that potential users might need to know: Identification Information; Data Quality Information; Spatial Data Organization Information; Spatial Reference Information; Entity and Attribute Information; Distribution Information and Metadata Reference Information (Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata). Of these areas only Identification Information (basic information about the file such as originator, abstract, and purpose) and Metadata Reference Information (information about the production of the metadata) are defined as being mandatory for all records. All the other areas of the standard are mandatory if applicable. Within each section are sub-fields that can be defined as mandatory, mandatory if applicable, or optional. This flexibility allows metadata creators to determine the level of detail that they can provide or support based on perceived user needs. It guarantees that at least basic metadata will be recorded about each data set. Hart and Phillips (1998) provide a useful overview of metadata creation.

It is important to note that the FGDC Content Standard is a *content* standard. It defines the content of the record rather than defining the method for organizing this information in a database or on a server, transferring files or displaying material to users (Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata). Other standards are used to define those processes. The FGDC Content Standard has created a Standard Generalized Mark-up Language (SGML) document type definition. SGML is an international standard that can be used to make digital materials accessible regardless of the specific system used to store the material (Cover 1997). By using SGML, metadata records can be easily indexed and shared using a variety of software. In addition, using server software that supports the Z39.50 protocol enables records in one collection to be seamlessly searched by other systems that employ the same protocol. Lynch (1997) discusses the value of the Z39.50 protocol for digital initiatives.

By Making use of the FGDC Content Standard, SGML and Z39.50, the work done by CUGIR can be easily searched, accessed and used by remote users.

Working with the FGDC Metadata Content Standard

During the grant timeline, metadata activities fell into four categories: staff training, creation of metadata for Mann Library-produced data, collaboration with data partners to create metadata, and planning for inclusion of data sets in the future. To accomplish these tasks, we had to decide who would work on the metadata records, what sort of training they would get and how the records would be created.

Choosing who in an organization will deal with the creation of metadata is an important starting point. Equating the creation of metadata records to the cataloging of books, Schweitzer (1998) suggests that metadata experts, not just data experts, should be involved in the process:

"Data managers who are either technically-literate scientists or scientifically-literate computer specialists [should create metadata records]. Creating correct metadata is like library cataloging, except the creator needs to know more of the scientific information behind the data in order to properly document them. Don't assume that every biologist or -ographer needs to be able to create proper metadata. They will complain that it is too hard and they won't see the benefits. But ensure that there is good communication between the metadata producer and the data producer; the former will have to ask questions of the latter."

Schweitzer observes that it is not practical for every data specialist to be as familiar with the metadata record structure as is necessary to produce metadata effectively. Therefore, he suggests adjusting workflow so data producers send basic information to data or metadata managers to create metadata. This is the approach that we followed at Mann Library to develop CUGIR.

Technical Services staff at Mann Library completed the metadata work. Learning and using the FGDC metadata standard fit with other work in the department since staff

are trained to work with complex metadata structures for other library work. The Metadata Librarian in our Technical Services department was the primary staff member designated to work with geospatial metadata. However, other Technical Services staff members have been given basic training in record structure and GIS and geospatial data concepts. This training was provided by a workshop given by Mann Library's GIS specialist in the summer of 1997. Following this introductory session, five staff members from Technical Services took part in the satellite videoconference: "A Practical Guide to Metadata Implementation for GIS/LIS Professional" (Hart & Phillips 1998). This conference provided an excellent introduction into the metadata record structure and to the tools that could be used to create metadata. Since catalogers were working on the creation of metadata, staff focused on metadata records in relation to one another in a database rather than solely on the content of individual records. This focus reflects a different perspective than that of the data producer. Larsgaard (1996) describes the complexity of cataloging geospatial data and the development of the metadata schema.

Mann Library created metadata for data sets that were produced at the library from TIGER/Line files. As part of that process, important areas of the record were highlighted for mandatory inclusion in CUGIR even though they were only deemed mandatory if applicable by the FGDC standard. All areas of the record had at least basic information. In addition, theme and place keyword types were identified for mandatory inclusion. For instance, data types and attributes are always included as theme keywords and FIPS codes and state, county or quadrangle names are always included as place keywords. This approach allows us to assume consistency within the database for searching and retrieval purposes.

The data sets that were created at Mann Library were at the county-level. Most of the information for the records was the same. Changes were predictable and involved differences in data set title, file name, bounding coordinates and place keywords. In addition, the county-specific information was the same for the ten coverages created. Coverage differences involved data set title, file name, abstract and theme keywords. To reduce the amount of time required to generate approximately 600 metadata records for these files, the Programmer/Analyst wrote a script to generate these files. The Metadata Librarian created a template metadata record, a file with the county - level changes and a file with the coverage-information changes. The script produced the 600+ records from these three files.

Work with Data Partners on Creation of Metadata

Mann Library also worked with data partners to produce metadata for data sets distributed through the Clearinghouse. This process was different but complementary to the process used to create metadata for materials produced at Mann Library. Fields were identified for mandatory inclusion and patterns were built into theme and place keywords. However, the process used to work with data partners was iterative and used the appropriate experts at the appropriate times. The data creators worked on content issues by providing a basic metadata record for review by the Metadata Librarian. The record covered all basic information related to the data set and the data experts were able to focus on the record and data set in question. The Metadata Librarian reviewed the record for format and consistency within the context of the database. Corporate names, theme and place keywords, and title formats are among the issues examined by the Metadata Librarian. This activity is consistent with the work done by catalogers in the Technical Services unit of Mann Library. In addition, Technical Services staff is familiar with using metadata-supporting documentation. The Metadata Librarian then returned the records to data partner staff for revisions and final work. This process enabled the data partners to focus on data provision and basic metadata information while the Metadata Librarian was able to provide metadata expertise and consultation. Future work will also involve the creation of MARC records for these data sets in the library's online catalog. These records will then be passed to two national databases-RLIN and OCLC. They will then be retrievable by searchers of those databases, and the records can also be downloaded by other libraries for their online catalogs.

During this project, metadata was created using NBII's MetaMaker (National Biological Information Infrastructure 1997), mp (USGS July 1998), and cns (USGS October 1998). These products were very useful in understanding the record structure and its requirements. It was also helpful that they worked jointly so several different software interfaces did not need to be used.

Data Organization and Technical Implementation

With data partners and funding secured, data identification, acquisition, and preparation tackled, and the process of generating standardized metadata begun, our staff began the work of building a system to distribute geospatial data and metadata. Prior to receiving a CCAP grant to become a clearinghouse node, we considered dissemination mechanisms that included the use of CD-ROM, Internet, NSDI clearinghouse node, or some combination. Although stipulations in our grant proposal limited our choice of system to a Web-based clearinghouse system, a number of implementation questions quickly followed. Specifically, we needed to address the questions of whether to build or buy software, what hardware we required, how files would be handled and how statistics would be gathered and analyzed.

Software-Build, Buy or Both?

System builders need to identify the most appropriate software infrastructure for a system that will best suit users' needs and will take advantage of current resources. The answer to the question of whether to build and/or buy software develops after enumerating specific requirements, examining time and human resource constraints, and reviewing existing software choices.

Our list of requirements revealed that the system needed to have a Web-based metadata searching facility and a geographic browsing facility supported by an interface that would integrate well with other clearinghouse nodes. Our time frame was set at something less than one year as determined by the CCAP grant. Since we had determined that much of the data conversion would be conducted in-house, we needed to limit the amount of funds that could be allocated to programmers for system development.

Our examination of geographic data distribution sites revealed that there were essentially two choices for our software architecture. First, we could build a Z39.50 database that would house and index the metadata for our system and integrate customized fields that would allow for extremely flexible Web-based browsing when combined with CGI scripts on our Web server. The second option would be to take the tested, popularly implemented indexing and searching freeware called Isite (Center for Networked Information Discovery & Retrieval) and create our browsing system separately. The first choice would allow us to create a very customizable and flexible interface that users could use to browse geographically and to search our metadata. However, this option was rejected for two reasons. First, it would require considerable resources and time to build and test it from the ground up. Secondly, despite the fact that Z39.50 protocol would be supported by such a system, it would be time intensive and difficult to attain the level of integration with other Clearinghouse Nodes that accompanies the FGDC-endorsed Isite software product.

By using the established Isite package, we had the advantage of using a tested, documented, and well-supported free product that worked well with existing nodes. Isite has facilities for simultaneously searching local and remote nodes that use the same software. Also, the FGDC Web site offers the ability to search all clearinghouse nodes that are using the Isite software simultaneously from their site (Federal Geographic Data Committee Geospatial Data Clearinghouse Entry Points). In addition, opting for

the Isite solution meant that the short development time and limited human resources could be focused on Web design and browsing facilities rather than on the creation and development of an entire Z39.50 database and information retrieval system. The disadvantage to the Isite system was that it would be difficult to integrate our homegrown browsing facilities given that the Isite product is continually being developed and upgraded.

Given time and human resources constraints and specified system requirements, one needs to make the determination to build or buy part or all of the software that will power a geospatial information dissemination system. In developing CUGIR, our circumstances warranted both build *and* buy (or rather borrow -- Isite is freeware). We elected to develop our own browsing system and to run the Isite metadata indexing and searching facilities in parallel. The browsing facilities consist of HTML pages containing maps and lists of geographic regions that interact with our data files via Perl CGI scripts (Cornell University Geospatial Information Repository 1998a, 1998b). This system has worked well, and the use of a file naming convention provides a high level of integrity between the systems.

Hardware

Once the software has been chosen, hardware that will support that software must be chosen. Hardware purchases depend on the type of dissemination system chosen as well as the software available or developed. Distributing material via CD-ROM will involve either contracting with a vendor to press the CDs or purchasing a CD writer. In the case of CUGIR, we had a server in place from an earlier geospatial data system implementation that would support Isite software and could sustain anticipated traffic. We needed only to purchase additional disc space to sustain the indexes generated by Isite and to house the data. Disc space is relatively inexpensive and the use of a SCSI port allows us to add more discs easily and efficiently as we acquire them by chaining the components together.

Another dissemination option is to form a partnership with a clearinghouse node. This is a viable option when the quantity of data to be shared is small or there are insufficient funds to purchase or build software or equipment. In this case, data suppliers should consider establishing a partnership with a clearinghouse node, such as CUGIR. If the data is within the clearinghouse's scope, the site developers will likely accommodate this material either free of charge or with a nominal fee.

Hardware decisions should include a system to backup your data, metadata, HTML documents, scripts, and programs regularly. CUGIR uses an 8mm magnetic tape backup system that is run on a weekly and monthly basis. The system and schedule used depend on the frequency of updates to data and metadata. Scripts and HTML files can usually be backed up by keeping local copies on the developer's machine, but maintenance of substantial amounts of data requires a more robust backup system.

File Handling

When maintaining large volumes of files from varying sources, it is useful to adopt a naming convention for metadata and data files. The naming convention allows for the quick and easy means for identifying and organizing data and metadata within the site.

Each unique data file at CUGIR and its corresponding metadata file begins with the same prefix. The prefix begins with either a 3-digit code or 2-letter 2-digit code that represents the geographic level of the data. For example, 109 represents the New York county number for Tompkins County while AA41 represents the quadrangle code for the 7.5 minute Monticello quadrangle in New York State. Following the geographic code is a two-letter feature code that identifies the theme of the data (e.g., 'hy' represents hydrography data). Finally, the prefix ends with a single letter code that indicates the format of the file (e.g., 'a' represents ARC/INFO export format). For example, the file for railroads in Tompkins County in Arc/Info export format is 109rra.e00.gz. There may be a second extension that is required by software to process the file, and the final extension is always indicates the means used to compress the file (Z = UNIX compress, gz = GNU compression). When distributing data files over the Web, compression is a necessity because data files are quite large. To ensure that users can open files, it is important to adopt common compression methods (e.g., UNIX compress or GNU zip). More details on the file naming convention used in CUGIR can be found within CUGIR (1998c).

The file naming convention provides an authoritative means of naming files that arrive from a variety of producers. Fortunately, the partner organizations of CUGIR have adopted, in part, the use of FIPS (Federal Information Processing Standard) codes and either the NYS Department of Transportation or USGS quadrangle codes in their naming of files. Use of these codes provides a base from which Perl scripts can be written to rename and move files around the site quickly. This convention is aided by having a fairly standard geographic coding system (such as FIPS) at its core.

Statistics

Once a system is in place, it is important to track which and how many data sets and metadata files are being disseminated. These statistics will be used in the future to acquire funding and to identify the most used data files for planning purposes. Web server logs provide a basic tracking mechanism for the number of downloads from CUGIR. However, the information from these logs is in no way comprehensive and often requires considerable work to strip away unwanted data. We utilize the log file analyzers, Analog (<http://www.analog.cx/>) and Webalizer (<http://www.usagl.net/webalizer/>) to customize the output of Web log statistics in HTML format. Another option for developers is to pipe entries from Web server logs into a database as they are generated. This system dynamically generates highly customizable reports. A recent article written in WebTechniques magazine (Stein 1998) details a method to do this with Apache Web server and MySQL database.

Future Considerations

When Mann Library constructed CUGIR it made a long-term commitment to providing geospatial data about New York State. The partnerships we formed with data providers are not one-time data acquisition agreements, but relationships that will grow and mature as these partners continue to expand and update the range of data sets they produce. Our focus is now on the long-term relationships with data partners, on finding ways to increase the amount of geospatial data and metadata in CUGIR, and on enhancing the access we provide to the data already within the repository.

We continue to contact data producing agencies whose data is not currently available via the Internet, encouraging them to place their data and metadata in CUGIR. We also continue to provide free metadata and data consulting services to new partners in order that they may begin the difficult process of creating standardized metadata describing their data products.

Our plans include making a number of enhancements to our data-browsing interface, including adding increased customization to the data theme and geography selection tools. We also plan to undertake a CUGIR user survey to better understand the ways in which people search and browse for geospatial data and metadata. With results from the user study combined with an analysis of our access logs, we will attempt to refine CUGIR's interface to make it easier to locate and retrieve data sets and metadata.

References

- Center for Networked Information Discovery and Retrieval.** [Homepage]. [Online]. Available: <http://www.mcnc-rdi.org/> [February 4, 1999].
- Cornell University Geospatial Information Repository.** 1998a. *Browse by Map*. [Online]. Available: <http://cugir.mannlib.cornell.edu/mapbrowse.jsp?series=counties> [February 4, 1999].
- Cornell University Geospatial Information Repository.** 1998b. *Browse by List*. [Online]. Available: <http://cugir.mannlib.cornell.edu/browse.jsp> [February 4, 1999].
- Cornell University Geospatial Information Repository.** 1998c. *Help & FAQ*. [Online]. Available: <http://cugir.mannlib.cornell.edu/help.jsp> [February 8, 1999].
- Cover, Robin.** 1997. *SGML: Answers to Basic Questions*. [Online]. Available: <http://www.isgmlug.org/whatsgml.htm> [February 4, 1999].
- Federal Geographic Data Committee.** *Content Standard for Digital Geospatial Metadata (CSDGM)*. [Online]. Available: http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf [February 4, 1999].
- _____. *FGDC Metadata*. [Online]. Available: <http://www.fgdc.gov/metadata> [February 4, 1999].
- _____. *Geospatial Data Clearinghouse Entry Points*. [Online]. Available: <http://clearinghouse.esri.com/> [February 4, 1999].
- _____. [Homepage]. [Online]. Available: <http://www.fgdc.gov/> [February 4, 1999].
- Hart, David and Hugh Phillips.** June 10, 1998. *Metadata Primer -- A "How To" Guide on Metadata Implementation*. [Online]. Available: <http://www.lic.wisc.edu/metadata/metaprim.htm> [February 4, 1999].
- Herold, Philip.** 1996. Moving Geospatial Data to the Web: GIS at Mann Library. *Library Hi Tech* 14(4): 86-87.
- Larsgaard, Mary Lynette.** 1996. Cataloging Planetospatial Data in Digital Form: Old Wine, New Bottles-New Wine, Old Bottles. In: *Geographic Information Systems and Libraries: Patrons, Maps, and Spatial Information. Papers Presented at the 1995 Clinic on Library Applications and Data Processing, April 10-12, 1995*. (ed. By Ed. Linda C. Smith & Myke Gluck). Urbana-Champaign, IL: Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.
- Lynch, Clifford A.** April, 1997. *The Z39.50 Information Retrieval Standard, Part I: A Strategic View of Its Past, Present and Future*. [Online]. Available: <http://www.dlib.org/dlib/april97/04lynch.html> [February 4, 1999].
- National Biological Information Infrastructure.** January 5, 1999. *NBII MetaMaker Version 2.22*. [Online]. Available: <http://www.nbi.gov/datainfo/metadata/metadata.symposium/leake/sld010.htm> [February 4, 1999].
- Schweitzer, Peter.** October 28, 1998. *Frequently-asked Questions on FGDC Metadata*. [Online]. Available: <http://geology.usgs.gov/tools/metadata/tools/doc/faq.html> [February 4, 1999].
- Stein, Lincoln.** 1998. Webmasters Domain: The Joy of SQL. *WebTechniques: Solutions for Internet and Web Developers*. Vol. 3, No. 10. [Online]. Available: <http://www.webtechniques.com/> [February 4, 1999].
- United States Geological Survey.** October 5, 1998. *Tools for Creation of Formal Metadata: cns: A Pre-parser for Formal Metadata*. [Online]. Available: <http://geology.usgs.gov/tools/metadata/tools/doc/cns.html> [February 4, 1999].
- _____. July 20, 1998. *Tools for Creation of Formal Metadata: mp: A Compiler for Formal Metadata*. [Online]. Available: <http://geology.usgs.gov/tools/metadata/tools/doc/mp.html> [February 4, 1999].

Feedback

We welcome your comments about this article. Please fill out [this form](#) for possible inclusion in a future issue.



Contents

