Faculty Publications, UNL Libraries

Libraries at University of Nebraska-Lincoln

2007

METADATA: THE KEY TO THE PRESERVATION AND DISSEMINATION OF SPATIAL DATA

Adonna Fleming
University of Nebraska-Lincoln, dfleming2@unl.edu

METADATA: THE KEY TO THE PRESERVATION AND DISSEMINATION OF SPATIAL DATA

Adonna Fleming
Geology Library
University of Nebraska – Lincoln
10 Bessey Hall, 0344
Lincoln, NE 68588
dfleming2@unl.edu

Abstract – The preservation and dissemination of digital research data has become a hot topic among the university research community. Of particular concern is the availability of GIS spatial data. Often the biggest roadblock to making GIS data available is that the creator of the project either does not have the time or the skills to create the metadata, a crucial component to making spatial data retrievable from GIS data portals. This paper describes how the University of Nebraska – Lincoln Libraries (UNL Libraries), as part of its GIS program, has stepped-in to help its campus community develop and distribute FGDC/ISO compliant metadata through the UNL Libraries' catalog, as well as state, regional and national GIS metadata portals.

INTRODUCTION

Academic libraries have always been responsible for archiving and providing access to research data. In the print age this role was accepted by the campus community with many academic libraries described as the 'heart of the campus'. In the digital age the role of the library has become more ambiguous to the research community. Researchers are developing data and metadata portals outside the purview of their university library, often without considering long- term access and preservation of the data, or the standardization of metadata records. Yet, the mission of an academic library has not changed and these libraries are still responsible for the preservation and dissemination of research materials regardless of format. Academic library personnel have found they need to be proactive with their research communities in advocating the need for the preservation of digital data, the standardization of metadata records, and emphasizing their role in this process.

The Association of Research Libraries (ARL) and the Cartographic Users Advisory Council (CUAC), among other organizations, have been advocating, with the academic library as a key player, for the preservation of digital data at the national level.

The ARL Steering Committees for Scholarly Communication and for Research, Teaching, and Learning jointly appointed a task force in 2006 to study the effects of e-science on academic libraries. The task force published a report in 2007 which, in part, advocates that, "Libraries be far more engaged in the processes of research, integrating content, tools and services more intimately within scholarly communication workflows"(ARL, 2007). The report continues, "Library expertise in developing systems and standards for digital content is relevant, as are library roles for stewardship and preservation of content"(ARL, 2007).

Reports from various government agencies during CUAC's Spring 2007 meeting focused on digitization issues and concerns. Dr. Brett Abrams, Electronic Records Archivist, National Archives and Records Administration (NARA) reiterated that his agency's mission to preserve the original work and to assist federal agencies in managing their records includes that of geospatial data. He indicated three initiatives of his agency that focuses on geospatial data and map digitization projects. One, the development of a working group (OGCTC) within the Open Geospatial Consortium (OGC) whose goal is to get private industry, international, federal and local government agencies, and universities to work together in developing open standards for geospatial information. Two, working with the *Geospatial One Stop Portal* community to assure access to historical collections, and three, the increased scanning of historical maps. He concluded by advocating the importance of the historical dimension of geospatial data and encouraged research libraries to join working groups that focus on the topics of preserving, archiving and accessing geospatial data (CUAC,

2007). A list of these groups can be found at: http://www.fgdc.gov./

Recognizing the importance of this national trend to preserve spatial data and provide access through the standardization of metadata records, the GIS program at the UNL Libraries includes an access and archive component. The goals of this component are: To advocate the importance of open access, the need for the preservation of GIS spatial data, and the necessity of metadata standards. One objective of these goals is to create and maintain the metadata for Nebraska geospatial data sets created by UNL faculty, staff and students, with the outcome of having that metadata searchable from the state's GIS metadata portal as well as the UNL Libraries' on-line catalog.

GIS PROGRAM AT UNL LIBRARIES

The GIS program at UNL Libraries consists of three components; development, outreach, access and archive. As part of the development component, the UNL Libraries provide four public workstations where patrons can create and store GIS data sets. The workstations offer a variety of GIS development aides including ESRI's *ArcView* and *ArcInfo* software. Patrons also have access to a large scale color scanner and a GPS unit which is available for check-out. Through the outreach component the GIS librarian offers introductory level GIS workshops in the UNL Libraries' electronic classroom using ESRI's *ArcView* software. In addition, the GIS librarian speaks to various campus organizations on the variety of ways to use GIS in their research.

The UNL Libraries began focusing on the access and archive component of their program in 2004 when they became a member of the Metadata Taskforce formed by the Nebraska Geographic Information Systems Council (NGISC). The NGISC is responsible for coordinating GIS data from the various state and local governmental agencies including those affiliated with the university. The purpose of this taskforce was to develop a metadata clearing house as part of the state's GIS data portal. In order to make the metadata searchable from this and other metadata portals, such as the *Geography Network* www.geographynetwork.com or the *Geospatial One Stop* (GOS) http://gos2.geodata.gov/wps/portal/gos, it must meet the Federal Geospatial Data Committee

(FGDC) Content Standard for Digital Geospatial Metadata (CSDGM). As an incentive to get UNL GIS programs to contribute their metadata to the clearing house, the UNL Libraries offered to write and maintain it for them on the clearing house portal. (Nebraska's metadata clearing house portal can be viewed at this url: http://www.dnr.state.ne.us/databank/geospatial.html)

BUILDING SUPPORT

In order to convince GIS users on campus that the UNL Libraries had the expertise to manage their GIS metadata, the GIS librarian attended a week long GIS metadata training class sponsored by the United States Geological Survey (USGS), and was certified to create metadata to FGDC/ISO standards. The GIS librarian then developed a PowerPoint presentation defining the role of academic libraries and made several presentations to various departments around campus. The presentation emphasized that libraries were always in the 'business' of organizing, classifying and providing access to information and correlated the similarities between the traditional library catalog and a metadata portal (Figure 1). As a result, in 2005, the UNL Libraries formed a partnership with the Nebraska Conservation and Survey Division (CSD) of the UNL School of Natural Resources to clean-up and maintain the GIS metadata to their various spatial data sets.

PROCESS

Metadata is best when the information is provided by the originator of the GIS data. However, in this case much of the CSD's GIS data was created over a period of years and the principal researchers on the projects were no longer available. In addition, only one of the four people working on the project, the GIS librarian, had a background in geology and was able to interpret the data. To rectify this, the GIS librarian developed a packet consisting of an information sheet, a copy of the original GIS metadata, and a template to use with one of the metadata editors.

The information sheet was designed to include information that was unique to each data set. For example, the title, abstract, original attributes, and a summary of the process would be included. This information was obtained either through an interview with the originator of the

spatial data, or derived from the existing metadata.

Personnel working on the project had a choice of metadata editors; the editor embedded in ESRI's *ArcCatalog*, or *TKME*, a free editor developed by Peter Schweitzer of the USGS. For those using *ArcCatalog* a template or 'dummy' record was created with information that was the same for every data set, such as contact information, legal notices, etc. This record was imported into the editor and then the unique information was filled in. For information on how to create a metadata template in *ArcCatalog* go to this article on the ESRI website: http://support.esri.com/index.cfm?fa=knowledge base.techArticles.articleShow&d=15270. For those using *TKME*, 'snippets', a macro function within the software, were programmed with the repetitive information. For information about creating *TKME* 'snippets' view this memo by the editor's author: http://geology.usgs.gov/tools/metadata/tools/doc/ opinion/announce/snippets.html.

The packet, along with a color coded example of metadata that met FGDC/ISO standards was then distributed by the GIS Librarian to those who were also working on the project. Once the metadata records were completed, the metadata was exported as an xml file and run through *MP*, a parser created by Peter Schweitzer which checks the syntax against the CSDGM standard. Once the errors were corrected, the metadata was returned to CSD's GIS coordinator for final review and then uploaded to Nebraska's metadata clearing house.

Initially four people worked on the project: Adonna Fleming, GIS – Maps – Geosciences Librarian and Margaret Mering, Principal Serials Catalog and Metadata Librarian from UNL Libraries; Les Howard GIS Coordinator from CSD and Rachel Simpson, Natural Resources Data Specialist, Center for Advanced Land Management Information Technologies (CALMIT). Fleming, Mering and Howard continue to work on this project.

Currently the UNL Libraries plan to expand this project by creating Dublin Core level cataloging records in the UNL Libraries' catalog for CSD's GIS data sets.

CONCLUSION

The UNL Libraries continue to stress to the GIS community their knowledge of GIS metadata standards and their expertise in the maintaining and cataloging of metadata records. They expanded their GIS program to include teaching metadata workshops. In 2007 NGISC, sponsored a metadata workshop taught by the UNL GIS librarian and Leslie Bearden from the USGS. In 2008, the UNL GIS librarian and Patrick Wilke-Brown of the Iowa Department of Natural Resources taught two metadata workshops as part of the Mid-America Geographic Information System Consortium's (MAGIC) Metadata Triage program.

METADATA RESOURCES

- USGS website with information about GIS metadata, including information about *TKME* and *MP*. http://geology.usgs.gov/tools/metadata/
- FGDC website with information about metadata standards including the CSDGM workbook, and how to contact metadata instructors in your area. http://www.fgdc.gov/metadata The workbook is available at this url: http://www.fgdc.gov/metadata/documents/w orkbook_0501_bmk.pdf.
- Mid-America Geographic Information System Consortium (MAGIC) website. includes metadata training exercises. http://magicweb.kgs.ku.edu/. Training exercises are at this url: http://magicweb.kgs.ku.edu/magic/projects/ metadata_triage.cfm

REFERENCES

Association of Research Libraries' Joint Task Force on Library Support for E-Science, 2007. Agenda for Developing E-Science in Research Libraries, p. 6. http://www.arl.org/bm~doc/ARL_EScience_fi nal.pdf. Accessed 11 July 2008.

Cartographic Users Advisory Council (CUAC), 2007. Agency Presentation Minutes April 26-27, 2007, USGS, Reston, VA. http://cuac.wustl.edu/. Accessed 11 July 2008.