Metadata Use in Research Data Management

by Christie Wiley

Research Data Access & Preservation

EDITOR'S SUMMARY

Prompted by federal requirements, academic institutions and librarians have become more involved in data management planning. An e-research implementation group was formed at the University of Illinois to develop and promote data initiatives in a variety of ways, including updating a data management website to provide education and awareness to students and faculty. Part of the site focuses on documentation and metadata, describing traditional and nontraditional data types and sources, including lab notebooks, data files and database contents, workflow and operating procedures. The site details the importance of data documentation for the immediate research project and for the institution, explains metadata and provides links to metadata standards and resources.

KEYWORDS

metadata
data set management
research data sets
strategic planning
academic libraries

Christie Wiley is the engineering research and data services librarian at the University of Illinois at Urbana-Champaign. She can be reached at cawiley<at>illinois.edu.

The National Science Foundation and other agencies require researchers to develop data management plans as part of their grant applications. These data management plans can identify types of data being collected, use of metadata and data gathering procedures, as well as policies and mechanisms for sharing data. However, researchers have generally not concentrated on the organization, access, reuse and preservation of data in their day-to-day research. Given these gaps, libraries and universities have been actively discussing the role and participation of librarians in research data management.

Many institutions have formed initiatives, committees and groups to provide support for a wide variety of research data activities: depositing data in institutional and external repositories or data archives; finding relevant external datasets; developing data management plans; creating tools to assist data management; bringing together available technology, infrastructure and tools; and data literacy education/training. To support researchers, some universities – such as Cornell, Johns Hopkins, MIT and Purdue – have established formal scholarly data services, and many other institutions are in the process of developing similar programs [1].
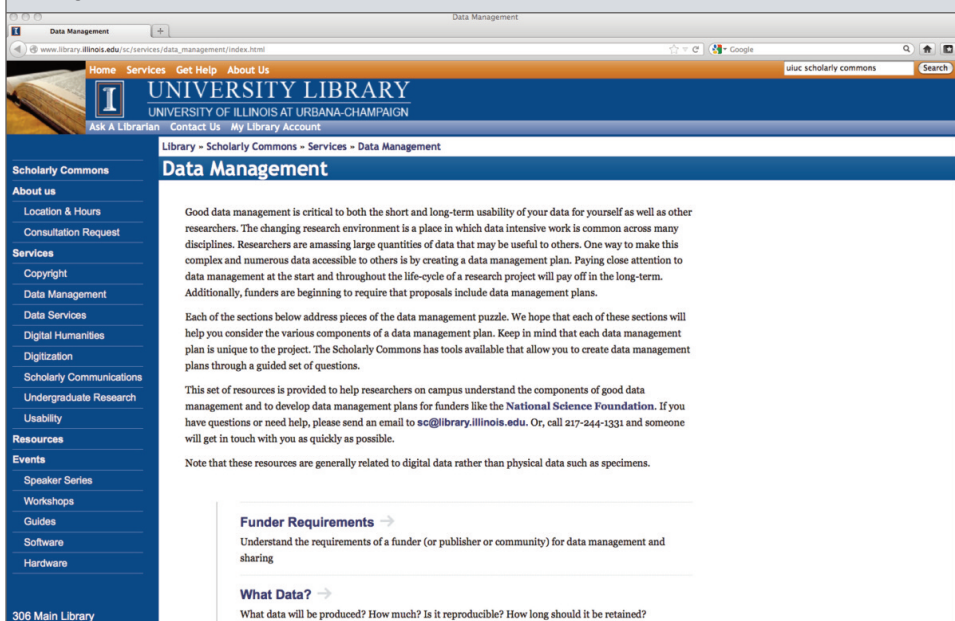
The University of Illinois created an e-research implementation group to bring together subject specialists, research data librarians and functional specialists to advance the library's data initiatives. These are among the initiatives created since the group's formation:

- hosting webinars on the data management plan tool;
- a pilot using EZID to assign persistent identifiers to datasets;
- creating a research data management interest group to provide awareness of data management to the campus and
- starting a research data management blog to provide information on data and data management topics.

In order to support the broader goals of 1) educating local researchers about the research data services that are available to them, 2) discussing ways that people think about data and 3) offering tools to meet their data needs, a smaller group of librarians within the e-research implementation group began a project to update the data management website. The updated website www.library.illinois.edu/sc/services/data_management/index.html now includes information and education regarding the definition of data, intellectual property, data sharing, funder requirements, file formats, privacy considerations, data documentation and metadata, and preservation and storage information. Users can access template guides and website resources needed for the data management plan tool. The goal of this website is to provide education and awareness regarding data management to students and faculty within and outside of the University of Illinois community. Since its launch in March 2014, the site has been accessed at least 129 times.

FIGURE 1. Screenshot of University of Illinois at Urbana-Champaign library's updated data management website

The website section featured at the 2014 Research Data Access and Preservation Summit focused on documentation and metadata. As libraries become more responsible for more data, they are being called on to support data preservation, discovery and analysis. Data types and sources include, but are not limited to, the following:

- bibliographic records,
- digital library collection metadata,
- website resources and
- research data.

The range and variety of researcher needs further complicate the process of creating and publishing data. Librarians often work with researchers to help them make their data more accessible, just as they have done for more traditional bibliographic materials.

Library bibliographic data includes information about printed and manuscript textual materials, computer files, maps, music, serials and continuing resources, visual materials and mixed materials. Bibliographic data commonly includes titles, names, subjects, notes, publication data and information about the physical description of an item [2].
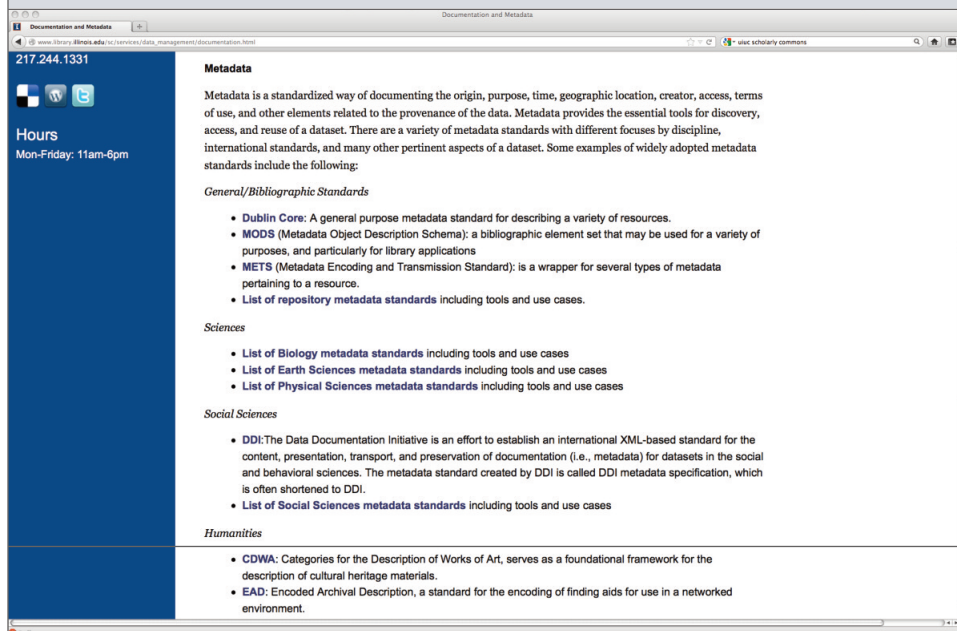
Research data is collected, observed or created for purposes of analysis to produce original research results. Research data categories can include observational, experimental, simulation, derived, compiled and reference. The types of research data may include text, word documents, laboratory notebooks, questionnaires, audiotapes, photographs, slides, data files, database contents, collections of digital objects, models, methodologies and workflows, and standard operating procedures and protocols.

The data documentation and metadata section of the website provides users with information about why data documentation is important, as well as the types of data level documentation that is needed for research data. The metadata section provides a definition of *metadata* and lists website links for general/bibliographic, sciences, social science and humanities metadata standards.

Multiple metadata standards exist within the various subject disciplines, although many standards collect similar information. Therefore, it is important that researchers carefully consider which metadata standard will best suit their data and their research needs. Factors to consider when choosing a metadata

WILEY, continued

FIGURE 2. Screenshot of the metadata section of the data management website

standard are the type of data the research produces, organizational guidelines and the resources available to create metadata. Metadata helps researchers avoid duplicating data, better share information and promote their work in various fields of study. It provides users with the ability to search, retrieve and evaluate datasets. Metadata allows users to find data and decide if data meets a particular information need. It also allows users to discover, process and use a dataset. Metadata provides value to an organization and institution because it helps protect the organization's investment in the data. It creates an institutional memory and advertises an institution's research efforts, thus creating partnerships and collaborations through data sharing [3].

The data management website is useful because it provides education to students, faculty and the community within and outside of Illinois about data management and importance of metadata and documentation to proper data management and data sharing. Reactions to this website have all been positive. Librarians, subject specialists and individuals from various organizations and institutions have stated that they would review the website or recommend it as a resource for others to use. Future plans are to continue assessing the use of the data management website and other campus research data management resources and to survey faculty to determine how they use the website in order to determine and possibly implement additional improvements. ■

Resources Mentioned in the Article

[1] Corrall, S., Kennan, M.A., & Atzal, W. (2013). Bibliometrics and research data management services: Emerging trends in library support and research. *Library Trends, 61*(3), 636-674. Retrieved from http://muse.jhu.edu/journals/library_trends/v061/61.3.corrall02.html

[2] U.S. Library of Congress. (1999, rev to April 2014). *MARC 21 format for bibliographic data.* Washington, DC: Library of Congress. Retrieved from www.loc.gov/marc/bibliographic/ecbdhome.html

[3] *DataOne*: www.dataone.org

CONTENTS   TOP OF ARTICLE   < PREVIOUS PAGE