October 2009

Metadata: Implications for Academic Libraries

Israel Yañez

Sacramento State University, israel.yanez@csus.edu

Metadata: Implications for Academic Libraries

Israel Yañez
Library
Sacramento State University
Sacramento, California

Introduction

This paper surveys the literature available on metadata and its implications for academic libraries. Definitions of metadata are offered. Metadata schemes are presented, including the scheme most prominently in use among academic libraries: Dublin Core. The paper examines the types of metadata projects prevalent in academic libraries, and the types of collections for which metadata is used. Implications beyond the technical aspects of metadata include organizational changes, and changes in roles and responsibilities necessary to implement projects involving the use and adoption of metadata. Also, implications for the skills set required of aspiring cataloging and metadata librarians are explored. Finally, this paper looks at the importance of undertaking metadata projects and the benefits for academic libraries in pursuing digital initiatives.

Almost every book and scholarly article on metadata includes a section dedicated to defining metadata. Caplan (2003) reminds us that while metadata is a term used in library science, it actually has its origins in computer science. The term metadata, as adopted by the computer science and library and information science field, simply means "data about data" (Greenberg, 2005).

Caplan further expands the definition, for her target librarian audience, as "structured information about an information resource or any media type or format" (2003, p.3). Other sources go a bit further adding the element of retrieval, use and management of an information resource to the concept of metadata (Understanding Metadata, 2004).

Miller (2004) offers a definition that seems to be aligned with FRBR's (Functional Requirements for Bibliographic Records) user tasks of "find, select, identify and obtain" (International Federation of Library Associations and Institutions, 1998). Miller tells us that metadata is the "extra baggage" associated with a resource that aids a user in finding that resource (find); discover where, and by whom it was created (identify); decide whether the resource is of value to the user (select); and conclude whether there is feasible access to the resource (obtain).

While the most basic definition of metadata ("data about data") can be applied to traditional library metadata such as the information on the cards in a card catalog and the information in a bibliographic record displayed in an online public access catalog (OPAC), when metadata is mentioned today, it usually alludes to data that facilitates the description, discovery, and retrieval of networked electronic resources (Hudgins, Agnew, & Brown, 1999). And while AACR2 is a metadata standard (Smiraglia, 2005), when metadata standards and schemes are mentioned today, more than likely reference is being made to metadata schemes such as Government Information Locator Service (GILS), Encoded Archival Description (EAD), Text Encoding Initiative (TEI), or Dublin Core, to name several examples.

Major Metadata Schemes

GILS is a service used to locate and access information resources generated by federal agencies. The general public can search the GILS records for thirty-two federal agencies via the search interface offered at GPOAcess (Government Information Locator, 2007). EAD is an "extensible markup language (XML) Document Type Definition (DTD) and the international standard for XML encoding of finding aids" (McCrory & Russell, 2005, p.99). With EAD, one can create collection-level finding aids and individual cataloging records (Single Item Metadata) for resources in academic libraries' special collections (Hudgins, Agnew, & Brown, 1999). EAD is modeled on TEI. TEI is a well established DTD. It is extensively used in the digitizing of texts, such as literary works, ensuring standardization and facilitating the sharing of texts in library collections (Hudgins, Agnew, & Brown, 1999). TEI provides a set of guidelines which stipulate encoding methods for machine-readable texts, mostly in the humanities, social sciences and linguistics (TEI: Text Encoding, n.d.).

Dublin Core is a metadata standard formed over a series of workshops attended by professionals from the computer science and library science worlds, as well as other professions (Hudgins, Agnew, & Brown, 1999). The name Dublin Core is derived from Dublin, Ohio, where the first workshop was held in 1995. That initial workshop produced the thirteen basic core elements. Today, fifteen elements make up the level of Dublin Core use which is called Simple. When the sixteenth element is used, along with refinements such as qualifiers, the level of Dublin Core use is called Qualified (Coleman, 2005).
The Dublin Core Metadata Initiative (DCMI) is instrumental in developing metadata standards and it holds an annual conference where professionals from the library world and other fields gather. The next annual conference will be held in Korea, in October 2009. The DCMI website offers helpful documentation, including a user guide by Dianne Hillmann (2005).   The DCMI website also presents a Library Application Profile. Application profiles "emerged within the Dublin Core Metadata Initiative as a way to declare which elements from which namespaces are used in a particular application or project" (n.d.). Application profiles vary according to the needs of the field of expertise or domain. The DCMI website also offers a guide for creating DC application profiles.

Uses for Metadata in Academic Libraries

Now that libraries are well into the "digital age," it no longer seems enough to only describe, organize, and provide access to traditional print material. Academic libraries, in particular, may have a special collections and archives unit; within these units we may find the types of collections typically organized and described using a metadata scheme. According to a 2005 survey of Association of Research Libraries (ARL) by Boock and Vondracek, digitization projects typically begin within special collections and archives departments. These departments contain the materials to be digitized. Some libraries even transfer the responsibility of digitization to a new digital library department, rather than have special collection and archives departments absorb the new responsibilities (2006).

Image databases are common digital library projects at academic libraries. Sacramento State University's Special Collections and University Archives maintain the Japanese American Archival Collection (JAAC) ImageBase. The digital collection consists of 1400 images in a database of selected photographs and images of artifacts related to the internment of Japanese Americans during World War II (University Library, California State University, Sacramento, n.d.). The digitization project is a collaborative effort between the Special Collections and the Cataloging departments. These collaborative efforts are common in academic libraries that undertake their first digitization project.

The JAAC ImageBase was built using CONTENTdm, an OCLC product that includes a user-friendly interface that allows library staff to contribute metadata records without the need of Dublin Core expertise. CONTENTdm also includes a searching interface, allowing end users to easily search and access collections.

Another name in the digital collection management field is Content Pro, a product from Innovative Interfaces, an integrated library systems vendor. Like CONTENTdm, Content Pro includes a collection building and management interface as well as an interface for the end user. Both CONTENTdm and Content Pro are OAI-PMH-compliant. OAI-PMH stands for Open Archives Initiative Protocol for Metadata Harvesting, an interoperability standard for repositories.

Academic libraries primarily target archival and special collections when undertaking digitization projects. In these digitization projects, historical documents/archives are a top priority for 38.7% of academic libraries, and photographs a top priority for 24.2% (Institute of Museum and Library Services, n.d.). These digitization projects typically imply the use of some metadata scheme to describe the collection.

ETDs

Electronic theses and dissertations (ETDs) is another type of collection at academic libraries that involve the use and adoption of metadata standards. ETD-MS is a metadata standard developed in 1997 by the Networked Digital Library of Theses and Dissertations (NDLTD) and used for ETDs. ETD-MS uses DC elements in addition to one extra element specifically for theses and dissertations (Networked Digital Library of Theses and Dissertations, 2008). The extra metadata element in ETD-MS is used to record the thesis degree, with refinements for the name of the degree, level of education, discipline, and granting institution.

There is a considerable amount of literature available on ETDs. University Microfilms (UMI) and Virginia Tech both were instrumental in the early development of ETDs. A workshop held in 1987 by UMI was the first time ETDs were given serious consideration. Virginia Tech first began accepting ETDs in 1995 (McCutcheon, Kreyce, & Maurer, 2008).
A related growing trend among academic libraries is the implementation of institutional repositories (IRs) to house the intellectual output of a particular academic community (Piorun & Palmer, 2008). This output, of course, includes theses and dissertations. ETDs are either scanned from the physical format, or submitted electronically by the student.

Institutional Repositories

One such repository is Humboldt Digital Scholar (Humboldt State University Library, n.d.). Humboldt State University is one of the smaller campuses among the 23 in the California State University system. Humboldt Digital Scholar (HDS) started out as a pilot project in 2004 and became a permanent service two years later (Wrenn, Mueller, & Shellhase, 2009).
As with most institutional repositories at universities, HDS' chief focus in building the collection were the theses produced by Humboldt State's graduate students.   The process for submission of an ETD involves the student describing the thesis in metadata fields and uploading a PDF of the thesis.
Cataloging staff at Humboldt State's library receive an automatic email notification when a submission is made. The thesis remains suppressed from public view until approved by the cataloging staff. Thesis advisor names are authority controlled according to AACR2r practice. Subject keywords are provided by the students and are not reviewed by the cataloging staff. After the thesis is approved, the student is notified and the URL of the electronic thesis is added to the record of the print thesis in the library catalog (Wrenn, Mueller, & Shellhase, 2009).
The HDS case exemplifies the trends and implications of metadata for academic libraries. Students provide provisional metadata.   Cataloging staff review and complete the metadata record. The record for the thesis is then housed in an institutional repository, itself an emerging trend among academic libraries looking to collect, store, and preserve its institution's scholarly output.

In Humboldt States' case, a record for the print thesis presumably already exists in the library catalog, since Wrenn et al. (2009) mention that the link for the electronic thesis is added to the record for

the print version in the library catalog. It would be interesting to research the workflow implications for cataloging departments who first must have to catalog the hard copy thesis and then have to contribute to the metadata record for the electronic thesis. From a FRBR-ized point of view, is it necessary to create not only two separate records, but two records under different metadata standards, of the same work? HDS was a catalyst in initiating a California State University system-wide DSpace repository: ScholarWorks. Scholarworks will be hosted by the CSU Chancellor's office and will be running DSpace and Manakin. DSpace is a an open-source platform for accessing, managing, and preserving scholarly works in digital format, including text, still images, moving images, mpegs and data sets (About DSpace, n.d.). Manakin is a user interface developed by Texas A&M University which supports the ability of each community (in this case, each CSU campus) and collection to establish a unique design that might extend outside of DSpace and into an existing institutional web presence (Texas A&M University, n.d.). Humboldt State plans to continue local hosting of their HDS operations while they contribute to ScholarWorks until they are comfortably satisfied the migration was a success and are happy with the flexibility and customization they require (Wrenn, Mueller, & Shellhase, 2009).

Digitization

Piorun and Palmer tell us that one of the implications of digitization projects is the fact that "skills and experience gained from a small project can be applied to larger-scale projects" (2008, p.223). Piorun and Palmer also mention the development of relationships in the institution and a new avenue for outreach as an additional implication of implementing an institutional repository.

As mentioned earlier, digitization projects at academic libraries have implications regarding the organizational structure of the library. In other cases, the implications have to do with the collaboration among library departments, distribution of digitization responsibilities, or the redefining of roles and responsibilities for library staff. In the summer of 2002, Sutton conducted an informal survey of special collections units of twenty libraries with holdings of three to five million volumes (2004). Sutton asked three questions, among them whether the library had implemented organizational changes as a result of digitization projects. Forty-seven percent reported they dealt with digitization projects in an ad-hoc fashion. Thirty-three percent responded they had a position in the special collections unit dedicated to managing digitization projects.   Twenty percent indicated they had established a new digitization unit, either within special collections or reporting to the same entity as special collections (2004).

Sutton's survey was done in 2002. It would be interesting to investigate whether there are more recent statistics regarding digital library departments or digitization projects within academic libraries. Also of interest would be to research whether there is a correlation between libraries that establish a new digital library or digitization unit, and that particular library's degree of embracing metadata and understanding its purpose.

Staff and Organizational Roles

To consider roles and responsibilities in metadata creation at academic libraries, Fleming et al. conducted a survey of libraries and their workflow regarding metadata creation (2008). They concluded that the metadata creation and maintenance had been distributed among librarians and paraprofessionals working in various library units, both technical services and non-technical services. The metadata schema most widely used by both groups was Dublin Core, followed by EAD (2008).
What skills might be required of a metadata librarian? Can a catalog librarian strongly versed in MARC-centric workflows easily become a metadata librarian with a mere name change in position title? What are the implications for aspiring cataloging librarians?

Chapman (2007) examined the skills and responsibilities required for the position of metadata librarian at research libraries. He targeted research libraries which placed the metadata librarian position within a traditional cataloging department or technical services unit. Chapman describes four roles

4

present in these metadata librarian positions: collaboration, research, education, and development (2007).

This paper has already mentioned the aspect of collaboration as typical in metadata and digitization projects. Chapman adds that collaboration, for the metadata librarian, is both internal and external. The metadata librarian must work with the technical services staff to develop procedures and practices, as well as stay abreast of developments in the metadata and standards communities to be well prepared for cross-institutional collaboration (2007).

Skills for head of cataloging positions at academic libraries are also reflecting the implications of metadata. Among the findings of Zhu's study of job advertisements for head of cataloging positions at academic libraries was a rise in the requirement of knowledge of non-MARC metadata schemes and digital resources (2008). A comparison in emerging job titles found in Zhu's study and those found by Khurshid (2003) five years earlier show a rise of the emerging term "metadata" among job titles in advertisements. Khurshid found the term metadata in almost 5 percent of job titles advertised while Zhu found the term in approximately 16 percent, a rise of about 11 percent in five years. While Khurshid's study included all cataloger positions and Zhu's study focused on head of cataloging department positions, the comparison between the two studies is still valid.

Given the rise of the term "metadata" in advertisements for cataloger's position, what are the implications for aspiring cataloging librarians? Should library and information science (LIS) students hoping to land a job in an academic library ensure that they include a metadata course in their program of study? Hsieh-Yee (2004) suggests three levels of expertise in metadata for LIS students. The first level of expertise is recommended for all LIS students and implies a general understanding of information organization and description. Included in this first level or expertise is an understanding of AACR and Dublin Core as examples of metadata schemes.

The second level of expertise recommended by Chan is targeted at aspiring metadata catalogers. LIS students aiming to become metadata librarians should have a solid knowledge of selected metadata schemes, including the ability to evaluate a metadata scheme. The student should also be aware of the nexus and distinction between metadata and traditional cataloging (2004). The third level of expertise Chan recommends is intended for LIS students aspiring to leadership roles in metadata projects. These students should acquire a strong command of cataloging standards and practices. They should be able to use a variety of metadata schemes and be able to identify the strengths and weaknesses of using a particular scheme for a particular project (2004). In spite of Chan's very detailed recommendations for students aspiring to become metadata librarians, she finds that only a third of LIS programs in North America offer a metadata course and only 19 percent offer an advanced metadata course (2004). While Chan's study is five years old, a review of the literature does not turn up a more recent or updated study of metadata offerings in LIS programs.

Areas for Future Research

This survey presents numerous possibilities for further research. Of particular interest is the idea of institutional repositories and the general contribution they make to the university mission, as well as particular contribution they make to the viability of the academic library. A more specific interest is the metadata scheme (ETD-MS) used for electronic theses and dissertations and the workflow implication for cataloging departments that catalog the print thesis and are also involved in creating metadata for the electronic version (presumably, using AACR2r). Sacramento State University, for example, will soon begin accepting electronic thesis submissions from graduate students. These theses will be contributed to the CSU-system institutional repository ScholarWorks. Currently, thesis manuscripts are received in print from the office of Graduate Studies. They are sent for microfiche filming and hard-cover binding. Once bound, the theses are cataloged with less-than-full records, using to the AACR2 metadata standard.

When Sacramento State University graduate students submit an electronic thesis, they will be providing preliminary metadata. Library staff will be completing the metadata record, according to a metadata scheme (most likely, ETD-MS). This means that for each work (to use the FRBR term), cataloging staff will use two different metadata schemes. Research into automated metadata generation tools might be in order.

Conclusion

Metadata implications for academic libraries seem to manifest themselves in the form of digital initiatives and, more specifically, institutional repositories. Piorun & Palmer remind us that institutional repositories can help academic libraries develop relationships with other campus departments, as well as provide new avenues for outreach into the surrounding community (2008). In times of budgetary challenges, academic libraries need to find ways to demonstrate they are central to the university's mission of teaching and learning. Engaging the campus faculty and graduate students in contributing to the building of an institutional repository should yield concrete results in raising the profile of the library on campus and in the community.

LIS students aspiring to become cataloging librarians should be proactive in seeking out courses that at least offer introductory material on metadata. Those LIS students looking to become metadata librarians will benefit from a course dedicated specifically to metadata. Finally, students seeking to play a leadership role in digital initiatives and metadata projects will require advanced metadata courses.

Academic libraries will do well to embrace metadata and the project possibilities it brings. Digital initiatives, particularly in the form of institutional repositories, help raise the profile of the academic library across campus and in the community. Metadata librarians, with an understanding of both metadata schemes and traditional cataloging, play a significant role in an academic library's digital initiatives. Metadata librarians in academic libraries are a prime example of the 21st century librarian.

Works Cited

About DSpace (n.d.). Retrieved February 24, 2009, from http://www.dspace.org/index.php/Introducing-DSpace/

Boock, M., & Vondracek, R. (2006). Organizing for digitization: A survey. *portal: LIbraries and the Academy, 6*(2), 197-217. Retrieved February 17, 2009, from Project MUSE database.

Caplan, P. (2003). *Metadata fundamentals for all librarians*. Chicago: American Library Association.

Chapman, J. W. (2007). The role of the metadata librarian in a research library. *Library Resources & Technical Services, 51*(4), 279-285. Retrieved February 21, 2009, from WilsonWeb database.

Coleman, A. S. (2005). From cataloging to metadata: Dublin Core records for the library catalog. In R. P. Smiraglia (Ed.), *Metadata: A cataloger's primer* (pp. 153-181). Binghamton, NY: Haworth Information Press.

Fleming, A., Mering, M., & Wolfe, J. A. (2008). Library personnel's role in the creation of metadata: A survey of academic libraries. *Technical Services Quarterly, 25*(4), 1-15. doi:10.1080/07317130802127983

Government Information Locator Service (GILS): About (2007). Retrieved February 21, 2009, from http://www.gpoaccess.gov/about/index.html

Greenberg, J. (2005). Understanding metadata and metadata schemes. In R. P. Smiraglia (Ed.), *Metadata: A cataloger's primer* (pp. 17-36). Binghamton, NY: Haworth Information Press.

Hillmann, D. (2005). Using Dublin Core. Retrieved February 21, 2009, from Dublin Core Metadata Initiative Web site: http://dublincore.org/documents/usageguide/

Hsieh-Yee, I. (2004). Cataloging and metadata education. In G. E. Gorman & D. G. Dorner (Eds.), *Metadata applications and management* (pp. 204-234). Lanham, MD: Scarecrow Press.

Hudgins, J., Agnew, G., & Brown, E. (1999). *Getting mileage out of metadata: Applications for the library*. Chicago: American Library Association.

Humboldt State University Library (n.d.). Humboldt Digital Scholar. Retrieved February 27, 2009, from http://dscholar.humboldt.edu:8080/dspace/

Institute of Museum and Library Services (n.d.). Status of technology and digitization in the nation's museums and libraries. Retrieved February 23, 2009, from http://www.imls.gov/resources/TechDig05/Technology%2BDigitization.pdf

International Federation of Library Associations and Institutions (1998). Part 3. In *Functional requirements for bibliographic records: Final report.* Retrieved February 9, 2009, from http://www.ifla.org/VII/s13/frbr/frbr3.htm

Khurshid, Z. (2003). The impact of information technology on job requirements and qualifications for catalogers. *Information Technology and Libraries, 22*(1), 18-21. Retrieved February 22, 2009, from WilsonWeb database.

McCrory, A., & Russell, B. M. (2005). Crosswalking EAD: Collaboration in archival description. *Information Technology and Libraries, 24*(3), 99-106. Retrieved February 18, 2009, from EBSCOhost database.

Miller, P. (2004). *Metadata: What it means for memory institutions*. In G. E. Gorman & D. G. Dorner (Eds.), Metadata applications and management (pp. 4-16). Lanham, MD: Scarecrow Press.

Networked Digital Library of Theses and Dissertations (2008). ETD-MS: An interoperability metadata standard for electronic theses and dissertations. Retrieved February 20, 2009, from http://www.ndltd.org/standards/metadata/ etd-ms-v1.00-rev2.html

Smiraglia, R. P. (2005). Introducing metadata. In R. P. Smiraglia (Ed.), *Metadata: A cataloger's primer* (pp. 1-5) [Introduction]. Binghampton, NY: Haworth Information Press.

Sutton, S. (2004). Navigating the point of no return: Organizational implications of digitization in special collections. *portal: Libraries and the Academy, 4*(2), 233-243. Retrieved February 22, 2009, from Project MUSE database.

Texas A&M University (n.d.). Digital initiatives: Research & technology. Retrieved February 24, 2009, from http://di.tamu.edu/

Understanding metadata (2004). [Pamphlet]. Bethesda, MD: NISO Press. Retrieved February 1, 2009, from National Information Standards Organization Web site: http://www.niso.org/publications/press/UnderstandingMetadata.pdf

University Library, California State University, Sacramento (n.d.). CSUS Japanese American Archival Collection ImageBase. Retrieved February 22, 2009, from http://digital.lib.csus.edu/jaac/

Wrenn, G., Mueller, C. J., & Shellhase, J. (2009). Institutional repository on a shoestring. *D-Lib Magazine, 15*(1/2). doi:10.1045/january2009-wrenn

Zhu, L. (2008). Head of cataloging positions in academic libraries: An analysis of job advertisements. *Technical Services Quarterly, 25*(4), 49-70. doi:10.1080/07317130802128072