March 2009

Metadata Plus: How Libraries Assure Discovery of Locally Created Content

Melanie Feltner-Reichert
University of Tennessee Knoxville, mfeltner@utk.edu

Marie Garrett
University of Tennessee Knoxville, mgarrett@utk.edu

Linda L. Phillips
University of Tennessee Knoxville, llphillips@utk.edu

Recommended Citation
Melanie Feltner-Reichert, Marie Garrett, and Linda L. Phillips. "Metadata Plus: How Libraries Assure Discovery of Locally Created Content," in *Pushing the Edge: Explore, Engage, Extend; Proceedings of the Fourteenth National Conference of the Association of College and Research Libraries, March 12-15, 2009, Seattle, Washington*. Chicago: American Library Association, 2009, pp. 75-80. Available at: http://works.bepress.com/linda_phillips/4

Metadata Plus: How Libraries Assure Discovery of Locally Created Content

Melanie Feltner-Reichert, Marie Garrett, and Linda L. Phillips

A growing number of academic libraries create local content. Digitization programs, scholarly publishing, institutional repositories, blogs, and web-based finding aids result in new information products. Unique special collections such as letters, photographs, and diaries are increasingly available in digital form. Library digital imprints bring to light specialized books, journals, and multimedia that might never be published through commercial channels. Emerging institutional repositories serve both as archives and promotional tools for a university's intellectual capital. Blogs and web-based finding aids contain a wealth of commentary and information representative of a rapidly evolving scholarly communications culture.

New information items, available in full text with instantaneous access via the internet, can be invaluable for scholars and other information seekers from kindergarten to life-long learners, provided they find the content. Locally created content often eludes traditional bibliographic channels for discovery. Thus, making the content discoverable to its intended audience is an essential complement to creation. If librarians demonstrate the numerous strategies available to assure content discovery, scholars are more likely to put their work into library digital collections and become partners in enabling discovery.

From the perspective of a scholar in the academic community, several options exist to discover digital resources. Faculty and students are likely to use web search engines such as Google and Yahoo as a first approach.[1] For more in-depth research, traditional scholars are accustomed to using bibliographic access conventions (such as cataloging, indexing/abstracting, and reviewing tools) developed for locating print resources and identifying distinctive characteristics of individual items. This paper demonstrates how metadata tagging complements the bibliographic apparatus developed by publishers of printed content to promote its access. Metadata enables discovery through web search engines. The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) permits harvesting in many ways; Digital Object Identifiers promote discovery of intellectual property by registration of individual articles. The following non-

Melanie Feltner–Reichert is Director of Digital Library Initiatives, e–mail: mfeltner@utk.edu; Marie Garrett is English Studies and Theatre Librarian, e–mail: mgarrett@utk.edu; Linda L. Phillips is Head, Scholarly Communication, e–mail: llphillips@utk.edu. All are from the University of Tennessee, Knoxville.

technical explanation of metadata mechanics, incorporating examples that can be understood by anyone familiar with a bibliographic record, offers a simple foundation for librarians who wish to understand and explain to their academic communities the context and benefits of metadata's critical role in discovering locally created content.

Metadata as a concept has become increasingly familiar to scholars. Social networking tools, from blogs to Second Life, and new forms of publishing such as institutional and disciplinary repositories provide models for content creators to enable discovery. Librarians, publishers, faculty members, and information technology specialists participating in a 2008 Council on Library Resources discussion about libraries in the 21st century predict that libraries will expand opportunities for users to take advantage of embedded information such as digital links, metadata harvesting, and connections to data sets.[2] Urging librarians to get closer to their communities as a strategy to thrive in a world where content produced on a massive scale is measured in exponential bytes, R. David Lankes advocates incorporating patron knowledge into digital data management.[3]

In the following discussion we illustrate, in a way that can be understood by librarians and our clientele, the way metadata works. Visualizing metadata in action empowers its potential creators. If seeing is believing, librarians can demonstrate that digital libraries are the most reliable place for creative work to be preserved and discovered. Metadata transparency also creates incentives for content creators to contribute digital content and descriptive information. Beyond traditional cataloging, metadata offers standardized contextual elements that enable harvesting and discovery. Templates for collecting OAI-PMH data are easy to use and explain. Registering to participate in OAI-PMH is simple. Metadata harvesters, such as OAIster, are no more complex than the latest full-text database purchased by the library and yield powerful search results. If librarians grasp the basics for combining digital discovery tools with traditional bibliographic and publishing conventions, we are well-prepared to inspire confidence in users who entrust their precious scholarship to libraries where the content is privileged as a part of the collection, easily discovered via internet search engines, and preserved for future generations. We can watch it work!

Metadata

Priscilla Caplan's definition of metadata as "structured information about an information resource of any media type or format" is a helpful starting point for considering the role of metadata in facilitating discovery.[4] The definition is broad enough to encompass all forms of metadata, including descriptive, technical, administrative, and structural. For the purposes of end-user discovery, however, descriptive metadata is the key. Like traditional records in MARC format, descriptive metadata is comprised of a set of access points which act as a surrogate for an intellectual object. As such, metadata enables discovery of content. It allows users to mine the "aboutness" of an intellectual object as would an abstract, identifying relevant materials. Therefore, metadata creation is crucial to dissemination of locally created content.

Figure 1. Metadata Record: Simple Dublin Core

Title: To advance their opportunities : federal policies toward African American workers from World War I to the Civil Rights Act of 1964
Creator: MacLaury, Judson
Date: 2008
Publisher: Newfound Press, University of Tennessee.
Rights: The author has licensed the work under the Creative Commons Attribution Noncommercial 3.0 United States License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/3.0/us/>.
Subject: United States. Committee on Fair Employee Practice – History.
Subject: African Americans – Employment – Government policy – History – 20th century.
Identifier: TU:DLC:Filename:0012_000053_000200_0000
Identifier: http://idserver.utk.edu/?id=200700000001691
Format: image/pdf
Type: text

Adherence to metadata standards and best practices ensures that records make sense beyond their local context, widening the audience for locally created content and maximizing the potential of the metadata. Quality metadata records communicate context and display coherence, consistency, and conformance to standards.[5] Records that are understandable and meaningful beyond their native collection environment convey context about the collection, exhibit coherence and consistency within the metadata set, and conform to standards such as uniform date formatting.

Figure 1 is a metadata record for an electronic book published by the University of Tennessee Newfound Press. The record contains elements like a MARC record, such as author and title; and further describes the features of the digital work, such as file format and persistent identifiers. Persistent identifiers are unique digital tracking mechanisms which facilitate access, discovery and preservation.

Benefits of Participation in OAI

Launched in 1999, the Open Archives Initiative (OAI) identified means by which content creators could expose digital content across repositories. A partnership among the Coalition of Networked Information (CNI), Digital Library Federation (DLF) and the National Science Foundation (NSF), the Initiative explored two impediments to interoperability: (1) the absence of machine-based methods of sharing metadata and (2) end-users' difficulty in navigating manifold search interfaces to discover content. To address these obstacles, the community developed a procedure for sharing metadata, the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). The community further identified the need for: (1) a protocol for transferring metadata, (2) a simple metadata format (Simple Dublin Core) as the baseline for sharing, (3) required metadata elements to ensure a common level of quality, and (4) intellectual property attributes outlining rights and usage limitations. The OAI-PMH addresses these needs, providing content creators with both a standard and a means for disseminating records and facilitating access to content.[6]

Figure 2 describes the same content as Figure 1, but displays the XML encoding required for OAI transfer and harvesting.

Creating OAI-compliant records and sharing them via the protocol increases exposure to locally created digital content. When a library registers as a

---

Figure 2. Metadata Record in OAI-PMH Format

```
<metadata>
<oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:dc="http://purl.org/dc/ele-
     ments/1.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://
     www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
<dc:title>To advance their opportunities : federal policies toward African American workers from World War
     I to the Civil Rights Act of 1964</dc:title>
<dc:creator> MacLaury, Judson</dc:creator>
<dc:date>2008</dc:date>
<dc:publisher> Newfound Press, University of Tennessee.</dc:publisher>
<dc:date>2008</dc:date>
<dc:rights>The author has licensed the work under the Creative Commons Attribution-Noncommercial
     3.0 United States License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-
     nc/3.0/us/>.
</dc:rights>
<dc:subject>United States. Committee on Fair Employee Practice – History.</dc:subject>
<dc:subject>African Americans – Employment – Government policy – History – 20th century. </dc:subject>
<dc:identifier>TU:DLC:Filename:0012_000053_000200_0000</dc:identifier>
<dc:identifier>http://idserver.utk.edu/?id=200700000001691</dc:identifier>
<dc:format>image/pdf</dc:format>
<dc:type>text</dc:type>
</oai_dc:dc>
</metadata>
```

---

data provider, its records become public and available to all harvesters and service providers affiliated with the Initiative. This means of sharing records increases the chances that content will be found and used by the academic community at large. For faculty who contribute content, visibility of their research is heightened, and content becomes more easily discoverable by peers. Libraries participating in the Initiative secure these benefits for their community, widening the audience for their institution's intellectual output.

Specifications

The OAI-PMH is comprised of protocols for both exposing and harvesting metadata records. The mandatory format for participation is Dublin Core, a descriptive metadata format comprised of 15 core elements. This metadata format ensures "low overhead" for participation, enabling smaller institutions (with fewer resources to create metadata) to share records and increase the discoverability of their digital content.

Those who offer records for harvesting, the data providers, register with the Initiative and share records through a static or dynamic repository. A repository is a server that hosts OAI-compliant records for harvesting. Static repositories are best suited for smaller, established collections ranging from one to five thousand items, and for institutions with minimal technical capabilities.[7] Conversely, dynamic repositories are best suited for larger, growing collections. Dynamic repositories require significant technical expertise to generate live responses to harvester requests. An institution may acquire and implement a software system that incorporates OAI-PMH features. Systems such as ContentDM, Digitool, and Digital Commons are among the many proprietary products that generate and deliver OAI-compliant records. Thus, institutions must either pay high costs for software or employ technical expertise to create and manage a dynamic repository. This makes the static repository an attractive, economically sound option for small institutions to become an OAI provider.

With relative ease, an institution can set up a static repository. Participation in OAI as a data provider requires elementary use of XML[8] and an openly accessible web server.[9] The steps are as follows: (1) Combine simple Dublin Core records for a collection into a single XML file; (2) Validate the file according to instructions at the OAI site; and (3) Save the valid file, according to OAI naming conventions (http://host:port/path/file.xml), to the web server.[10] For example, the URL for a static collection of sketches by Tennessee artist Catherine Wiley hosted on the University of Tennessee's Digital Library server might look like this: http://diglib:080/static/wiley.xml. Once the XML file is mounted on the web server, the host institution registers as a data provider with the Initiative at http://www.openarchives.org/Register/ValidateSite. To register, submit the URL for validation and indicate (1) the institution name, (2) contact information, (3) the version of OAI-PMH employed in the static repository. The reward for this investment is global visibility of locally created content via numerous gateways and search engines.

Metadata Plus

In our quest to make scholarship more widely available, librarians must learn from our clientele and other allies in the information world. People who use libraries have turned to Google as a primary tool for discovering information. Library partnerships with Google and similar ventures are making local content more discoverable, as are sophisticated search engines that mine embedded tags. Longstanding bibliographic devices developed by publishers and librarians to promote discovery remain useful. Book publishers purchase ISBN numbers, solicit book reviews, and register content with the Library of Congress, Amazon, and R. R. Bowker (*Books in Print*). Journal publishers secure ISSN numbers, register content with abstracting/indexing services and *Ulrichsweb*, and obtain agreements with journal aggregators who disseminate full text.

Production of library catalog records for print and electronic content continues. As libraries collect, generate, and preserve more digital content, discovery hinges on creating and sharing metadata. The formidable issue of accommodating massive amounts of data requires collaborative discovery solutions from both print and digital worlds.

In today's rapidly changing digital library landscape, librarians must appeal to faculty whose scholarship represents both commodity and achievement for the university. Faculty who understand the multitude of discovery options now available will be more likely to entrust their scholarship to the library. Illustrating the ways metadata and more traditional bibliographic devices converge, librarians advance the scholarly lit-

eracy advocated by Courant and reinforce the library's role in access, discovery and preservation. Via presentations at faculty meetings, at university functions, in newsletters, and in campus publications, librarians might also convince faculty to contribute metadata that makes content discoverable.

As more robust means for discovery emerge, libraries are providing greater visibility for faculty publications and other locally created content. By demonstrating the ways discovery mechanisms work, librarians promote an understanding of the potential for unique material to reach the widest audience possible through university publishing and archiving. Recognition of metadata basics and complementary traditional bibliographic access tools prepares librarians to uphold the library's image as a safe and sustainable archive for faculty work in the digital age. The simple application of metadata protocols and standards ensures the best possible chance for the exposure of locally created content. Publishing content with the library is more likely to lead to its discovery than any personal web space.

Notes

1. Paul Courant. "Scholarship, the Wave of the Future in the Digital Age." In: The Tower and the Cloud. Richard N Katz, ed. http://net.educause.edu/ir/library/pdf/PUB7202t.pdf

2. *No Brief Candle: Reconceiving Research Libraries for the 21st Century*. Washington, D.C.: Council on Library and Information Resources, August 2008. www.clir.org/pubs/abstract/pub142abst.html.

3. R. David Lankes, "Collecting Conversations in a Massive-Scale World," *Library Resources & Technical Services* 52 no. 2 (April 2008): 12-18.

4. Caplan, p. 3.

5. Sarah L. Shreeves, Jenn Riley, and Liz Milewicz. "Moving towards shareable metadata." *First Monday* 11, no. 8 (7 August 2006). http://firstmonday.org/htbin/cgi-wrap/bin/ojs/index.php/fm/article/view/1386/1304

6. Leona Carpenter (coordinating author), "History and Development of OAI-PMH," *OAI for Beginners, the Open Archives Forum online tutorial*. University of Bath, 2003. http://www.oaforum.org/tutorial/english/page2.htm.

7. See http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm.

8. XML (Extensible Markup Language) is the machine-readable format for storing metadata records and commonly held knowledge of programmers and metadata specialists. In the absence of this type of support, numerous online tutorials exist.

9. Specifications drawn from OAI website documentation available at http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm.

10. See http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm.

Bibliography

Caplan, Priscilla. *Metadata Fundamentals for All Librarians*. Chicago: American Library Association, 2003.

Lynch, Clifford.A. "Metadata Harvesting and the Open Archives Initiative." *ARL Bimonthly Report 217*, August 2001. http://www.arl.org/resources/pubs/br/br217/br217mhp.shtml

Smiragalia, Richard P. Metadata: A Cataloger's Primer. Binghamton, NY: Haworth Information Press, 2005.

Van de Sompel, Herbert., Lagoze, Carl. The Santa Fe Convention of the Open Archives Initiative. *D-Lib Magazine*, vol.6, no.2. February 2000. http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html