Integrating a Digital Library and a Traditional Library: Librarians and Scientists Collaborating for Sustainability

By

Susannah Benedetti
Sue Ann Cody
Arlene Hanerfeld

ABSTRACT. With the proliferation of digitized materials and digital libraries, academic librarians have an opportunity to use their metadata and database design expertise to cultivate relationships with faculty and digital-object users. This article describes how librarians at the University of North Carolina Wilmington (UNCW) have made an effort to expand access to resources in a digital library both in their library catalog and in WorldCat.

KEYWORDS. Digital libraries, metadata harvesting, cataloging, electronic resources, learning objects, collaboration, sustainability, academic libraries.

Susannah Benedetti, BFA (Communication Arts), MLIS, is Coordinator of Cataloging Services, University of North Carolina Wilmington, 601 S. College Road, Wilmington, NC 28403-5616 (E-mail: benedettis@uncw.edu).

Sue Ann Cody, BS (Social Sciences), MLS, MA (U.S. History), is Associate University Librarian for Public Services, University of North Carolina Wilmington, 601 S. College Road, Wilmington, NC 28403-5616 (E-mail: codys@uncw.edu).

Arlene Hanerfeld, BS (Economics), MLS, is Associate University Librarian for Technical and Collection Services, University of North Carolina Wilmington, 601 S. College Road, Wilmington, NC 28403-5616 (E-mail: hanerfelda@uncw.edu).

NOTE: This article was originally published in *Technical Services Quarterly*, v. 24, no. 3, with different page formatting and is cited as follows: Benedetti, S., Cody, S. A., & Hanerfeld, A. (2007). Integrating a digital library and a traditional library: Librarians and scientists collaborating for sustainability. *Technical Services Quarterly, 24* (3), 15-27.

Advances in digital technologies have made it increasingly easy to digitize, scan, and electronically distribute information such as image, audio, and video files; ebooks; ejournals; electronic documents; Web sites; and databases. Responding to the proliferation of such materials and recognizing them as valid library materials, librarians have developed rules and standards for cataloging electronic resources, expanding and updating core cataloging resources such as *Anglo-American Cataloguing Rules* (AACR2) and the OCLC guide *Bibliographic Formats and Standards*. The result is that libraries are beginning to provide patrons with access to digital objects through library Web pages and online library catalogs. However, according to OCLC researchers, only approximately 1.5 percent of the resources contributed to WorldCat are electronic materials.[1] Librarians at the University of North Carolina Wilmington (UNCW) have made an effort to expand access to resources in a digital library both in their library catalog[2] and in WorldCat.

In the mid-1990s, public discovery of the Internet led independent creators to gather digital collections and provide access in Web-based digital libraries. According to the National Information Standards Organization (NISO), a digital collection consists of digital objects that are selected and organized to facilitate their access and use, optimally with metadata to describe and manage them, and with an interface that provides access for searching, browsing, retrieving, and using the objects.[3] The National Science Foundation (NSF) recognized the value of digital libraries to improve mathematics and science instruction as early as 1994.[4] The National Science Digital Library (NSDL)

program began in January 2000 with the first requests for proposals resulting in funding for 42 projects.[5] The iLumina project, based at UNCW, was one.

Lynch reports that "the lion's share of the NSF funding went to computer science groups, with libraries often being only peripherally involved, if at all."[6] While few librarians were involved in the early creation of digital libraries, they recognized the intellectual value of such resources and cataloged them as collections. Collection level cataloging, while providing access to digital libraries' portals, does not provide access to individual resources in the digital libraries. Librarians at UNCW have created a model to harvest the individual records from a digital library and integrate them into the traditional library catalog. This approach vastly increases access because the individual digital resources will be retrieved along with other library materials during library catalog searches. Contributing these records to WorldCat makes the individual resources accessible to OCLC's global network of more than 50,000 libraries. The project was grounded on in-depth collaboration between librarians and the faculty scientists who created the iLumina digital library. This article will examine the background of the iLumina digital library, the genesis of the faculty/librarian collaboration, and the details of developing crosswalks and metadata harvesting.

History and Background

The iLumina digital library[7] was created by faculty scientists to collectively improve the quantity and quality of digital teaching resources.[8] A group of UNCW science faculty in biology, chemistry, computer science, mathematics, and physics entered into discussions with science faculty at Georgia State University, Grand Valley State University (Michigan), Virginia Polytechnic Institute and State University, and

researchers at the higher education service provider Eduprise. Together they recognized that science educators around the country were creating valuable digital resources for teaching, and that many of the educators were eager to share these learning objects. Yet there was no digital repository for such an exchange of resources. In order to create such a repository, in 2000 the faculty scientists applied for and received funding from the NSF's Digital Library Initiative – Phase II (NSF DLI-II), and iLumina ultimately became one of over 450 collections in the NSDL.[9] William Mischo and Lee Zia provide background information on the NSF's role in developing this national digital library supporting science, mathematics, engineering, and technology education at all levels.[10] At this point it did not occur to the scientists at UNCW that librarians on their campus could provide valuable assistance with this project.

Upon receipt of the NSF funding, the iLumina team developed the digital library using IMS, a metadata standard created specifically for educational resources by the IMS Global Learning Consortium.[11] IMS was derived from the Institute of Electrical and Electronics Engineers-Learning Object Metadata (IEEE-LOM) standards and was developed for tagging metadata in Extensible Markup Language (XML) format, in hopes of facilitating the exchange of metadata between digital libraries. IMS, with 19 core elements and 67 sub-elements, allowed for a richer set of metadata than the simpler Dublin Core metadata with only 15 elements.[12] IMS was thus an attractive choice for describing iLumina's extremely granular resources, such as individual images, video clips, interactive Java Applets, documents, presentations, and entire courses. Figure 1 is an example of one of the iLumina image resources, "Boiling Water." [place Figure 1

here] The iLumina team developed protocols for metadata records in IMS, not knowing that the NSDL would ultimately adopt Dublin Core as its standard metadata scheme.

As soon as the announcement of the initial funding to develop the digital library appeared in the UNCW campus newsletter, the Associate University Librarian for Public Services issued an offer to help with the project. She already had an excellent and longstanding relationship with the scientists on the team through a variety of campus projects. Following a flurry of individual questions concerning taxonomies and user-centered interfaces, the scientists recognized the need for librarians' expertise in database development and maintenance, and they invited her to join the team. By the end of the two-year funding period, iLumina was functional and contained over 1,400 shareable digital learning objects for undergraduate education in science, technology, engineering, and mathematics.

Collaboration for Sustainability

The UNCW scientists now faced the issue of iLumina's future: How could the digital library sustain itself and grow? As with all grant-funded projects, when the original funding is exhausted, alternative funding schemes are necessary to cover operating costs, maintenance and growth of the collection, and expansion of the user base. Realizing that many of the NSDL projects would face such sustainability issues, the NSF initiated the Targeted Research Track of the NSF 03-530: National Science, Technology, Engineering, and Mathematics Education Digital Library. The purpose of this funding category was to explore various business models for ensuring the continuation of the NSDL projects.

During the grant application process, the Associate University Librarian for Public Services became aware of the XML Harvester, an automated cataloging tool to create library catalog records from metadata stored on external servers.[13] The XML Harvester was being developed by Innovative Interfaces, Inc. (III), the integrated library system used at UNCW. The librarian proposed the XML Harvester as a possible tool to migrate the iLumina metadata into the library's catalog, thus sustaining and expanding access to the individual learning objects. At this point the librarian recognized the need for the cataloging and systems expertise of the Associate University Librarian for Technical and Collection Services and Associate University Librarian for Computing Systems. They joined the discussion and eventually the iLumina team.

The scientists were unaware of the depth of the traditional library community's expertise and vast experience with metadata and cataloging principles and bibliographic standards. As library materials have evolved from monographs and serials to sound, visual, and electronic formats, librarians have established their expertise in applying time-tested and relevant standards to Web-based and digital formats. A review of the literature demonstrates that there is a developing symbiotic relationship on university campuses between faculty and librarians. Faculty are developing digital objects for classroom and research activities, and librarians are recognizing, as Lynch notes, "an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution."[14] Zia observes changing relationships between faculty and librarians as faculty become interested in metadata and exposure for their digital objects, and librarians recognize the need to convey resources developed on their campus to the world.[15] Borgman identifies

this dynamic as a possible tension between researcher-librarian communities,[16] but the newly constituted iLumina team of scientists and librarians quickly saw an opportunity for collaboration and experimentation.

The librarians presented the scientists with the possibility of using the MARC format and bibliographic standards that provide the basis of library catalogs. The scientists were unaware that libraries adhered to such standards and were amazed that sophisticated metadata such as the MARC format had been in place since the 1960s. This discussion was a harbinger of many instances of the two communities exchanging information about their respective disciplines. The need for collaboration between librarians and scientists is noted in the literature. Lynch recalls early support of major federal research funding agencies, and notes, "what they didn't do was systematically create a funding program dealing with digital libraries research in a general way or that strongly encouraged cross-disciplinary collaboration, including collaboration among computer and information scientists, engineers, librarians, and social scientists."[17] In "Dewey Meets Turing: Librarians, Computer Scientists, and the Digital Libraries Initiative," Paepcke et al. review the NSF funding initiatives of the 1990s that offered an opportunity for groups from each discipline to benefit from each other's expertise. Computer scientists could provide information technology skills to develop digital libraries, and libraries could benefit from an infusion of funds to develop online library catalogs capable of providing enhanced access to new formats of scholarly information. The authors describe how the advent of the World Wide Web threatened this partnership as it created a "culture of laissez-faire retrieval."[18] This attitude undermined faith in library tenets of data collection, reliability, and integrity. Nevertheless, the authors

emphasize, "While information access now rests on a highly technical infrastructure, the core function of librarianship remains. The information must be organized, collated, and presented."[19] The UNCW librarians and scientists based their collaboration on this principle.

As the iLumina team prepared to apply for funding from the NSF's Targeted Research Track, they focused on the fact that over 80 percent of the NSDL projects are university-based initiatives. The team recognized the university library as a potentially sustainable environment for iLumina and a possible model for connecting other digital libraries to their own existing traditional library infrastructures.

The librarians embraced this opportunity in part because the project would fulfill the library's mission to provide access to information in all formats.[20] It would also provide a level of funding not often available to libraries in support of new experimental endeavors. In addition, it would provide the library faculty with valuable knowledge about the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), an interoperability framework created to facilitate the dissemination of digital content.[21] Knowledge of this protocol could be applied to future digital initiatives at the UNCW library. Finally, it became apparent that the iLumina digital library might serve as a case study for the development of an institutional repository at UNCW. The library faculty were already investigating the institutional repository model, well aware that the management of digital resources created by university faculty is certain to be a part of the future of university libraries. Indeed, in 2003 Lynch defined an institutional repository as "a recognition that the intellectual life and scholarship of our universities will

increasingly be represented, documented, and shared in digital form, and that a primary responsibility of our universities is to exercise stewardship over these riches."[22]

Goals and Objectives

In 2003, the iLumina team received the NSF grant (award number 0333628) to integrate the iLumina digital library with the UNCW library catalog, to contribute the iLumina records to WorldCat, and to provide a sustainability model for other digital libraries. The iLumina digital library would also continue to be a component of the NSDL. Specific objectives included integrating iLumina records into library catalog searches with an option of limiting searches to iLumina resources and developing new procedures for reviewing and accepting future submissions to the digital library. These objectives first required the team to investigate crosswalking IMS metadata to the MARC format as well as to Qualified Dublin Core (QDC),[23] adopted by the NSDL as its official metadata standard. The team's objectives were grounded on the principles of adapting library traditions to the digital environment. At this point the Coordinator of Cataloging Services and the Special Formats Catalog Librarian joined the team to offer expertise in bibliographic standards and cataloging protocols.

Integration

The cataloging librarians created MARC records for two iLumina resources to provide an opportunity for the scientists to view the records in the library catalog. The entire team then participated in discussions for crosswalking IMS metadata to MARC and QDC. The team initially hoped to create one crosswalk, but after examining the specifics of each metadata scheme, determined that any crosswalk taking data from the relative simplicity of QDC to the richness of MARC would inevitably result in significant data

loss. The team ultimately agreed that two separate crosswalks would be necessary, one for converting IMS metadata to QDC, and one to MARC. The final crosswalks, iLumina to MARC21 and iLumina to Qualified Dublin Core, are available on the iLumina Technical Documents Web page.[24]

The next step used the skills of undergraduate computer science programmers to develop a metadata converter to automate the crosswalks. The program was intended to convert the original IMS metadata to QDC and MARC XML. It was necessary for the MARC data to be in XML format in order to conform to the functionality of the Innovative XML Harvester and the OAI-compliant harvester used by the NSDL. The scientists on the team had originally foreseen the crosswalks as a means of creating a simple batch process for converting records that would not require individual editing. The librarians, however, insisted on following MARC and AACR2 standards. They would be contributing the records to WorldCat, and although they did not plan to submit them as full-level cataloging, they intended the records to follow standards whenever possible.

There were several lively discussions between the librarians and scientists on the team about the value of adhering to bibliographic standards. The scientists, as Paepcke et al. describe, "could not understand why librarians are so annoyingly deliberate about metadata,"[25] but the librarians persisted, and the student programmers wrote code that addressed the librarians' concerns. For example, the creators of the IMS metadata had formatted author names as first name, middle initial, last name, often preceded by the professional title "Dr." For the MARC personal author field, the librarians insisted that the metadata converter program switch the format to last name, first name, middle initial, and also remove any professional title. The librarians then insisted that the program insert

the main entry personal author into the statement of responsibility field, reverting to the first name, middle initial, last name format. The student programmers also inserted the note fields "Title from title screen (viewed November 5, 2004)" and "Mode of access: World Wide Web" into each MARC record. In addition, they inserted the correct AACR2 punctuation into appropriate MARC fields. This programming ensured that records would adhere to standards, and no extensive editing was needed.

Figure 2 demonstrates the harvesting procedure. [place Figure 2 here] The metadata converter writes the IMS metadata to QDC and MARC XML. The OAI data provider/harvester software, developed by the Digital Library for Earth System Education (DLESE), harvests both formats.[26] The NSDL harvests the QDC metadata file. The XML Harvester harvests the MARC XML file from the OAI provider and exports it to library catalog.

Once the records were harvested for the library catalog, the III functions Create Lists and Global Update were used to provide additional batch editing of records. Create Lists was used to gather all iLumina records into a file, and Global Update was used to insert the LCSH "Internet in education" into each record. Since iLumina taxonomies were not LCSHs, they were mapped into local subject fields. Each record's taxonomy field included the discipline represented by the resource, such as biology, chemistry, computer science, mathematics, or physics. Librarians once again used Create Lists to create separate files according to discipline from the taxonomy field, and using Global Update added discipline-appropriate LCSHs to each record. For example, for each record that had "Chemistry" as the first word in the taxonomy field, Global Update was used to add the LCSH "Chemistry -- Computer-assisted instruction." Although the scientists on

the team had originally envisioned the harvesting process as being fully automated, they were impressed with the technology available in the library's catalog to perform these batch editing functions, and in the end appreciated the value added to the records. Figure 3 shows the final version of the MARC record for the iLumina resource "Boiling Water."

[place Figure 3 here]

The librarians were now ready to contribute the enhanced records to WorldCat. Once again they used Create Lists to gather all the iLumina records into a file and to export them, in MARC format, to a file in OCLC's Cataloging MicroEnhancer (CatME) software. From CatME they were batch submitted to the WorldCat database. The iLumina learning objects are now available as shareable information resources to libraries throughout the world.

iLumina's Future

The iLumina digital library now exists within the UNCW library catalog, and the librarians manage and maintain it. The original iLumina gateway was retained in order to house the interface for new submissions. The librarians are the managers of iLumina, and a librarian has transitioned into the role of digital library editor, reviewing new submissions and overseeing future harvests. The original science team members remain involved as consultants and discipline editors.

The iLumina collaboration led to other unforeseen opportunities as scientists and librarians gained improved awareness of shared interests and special skills. For instance, librarians were invited to develop and teach a course for a new minor in information technology offered by the Computer Science Department, and the library was able to

offer internships to Computer Science students. The iLumina gateway is now being considered as a way to develop an institutional repository for the university.

Conclusion

Although the digital library funding boom beginning in the 1990s did not initially include librarians in leadership roles, there has been a gradual acknowledgement that foundations of librarianship (organization, collocation, evaluation, collection, reliability) remain valuable, even critical, for digital library creation, preservation, and growth. The iLumina digital library project illustrates that there are opportunities not only for librarians' involvement, but also for their vision and leadership. Huwe commends librarians for their advances in incorporating metadata principles and applying standards to new formats, and for the increasing visual sophistication of library Web sites that link to collections of digitized materials and other relevant resources. However, he identifies "an overall shortfall in the 'vision' department" and stresses the need for "far more integration of born-digital, or 'built' content with traditional library collections and services."[27]

This attitude reflects the development of a new dynamic. Librarians must cultivate ongoing relationships with faculty who are creating digital objects and digital libraries. The iLumina project might not have received continued funding if the Associate University Librarian for Public Services had not reached out to the faculty scientists, offering expertise in sustainability, consistency, and accessibility. Huwe points out that librarians need "to get serious about creating and using repositories as key elements in our overall information ecologies."[28] By working with faculty scientists to harvest a

digital library and assume its management, the librarians on the iLumina project feel that they have answered the call to integrate and expand access to digital resources.

References

1. Brian Lavoie, Lynn Silipigni Connaway, and Ed O'Neill. "Mining for Digital Resources: Identifying and Characterizing Digital Materials in WorldCat." Paper presented at the 12th Association of College and Research Libraries Conference (ACRL) Minneapolis, MI, April 7-10, 2005.

2. University of North Carolina Wilmington, "William Madison Randall Library," <http://library.uncwil.edu/>.

3. NISO Framework Advisory Group. *A Framework of Guidance for Building Good Digital Collections* (Bethesda: NISO Press, 2004), 3.

4. William H. Mischo, "United States Federal Support for Digital Library Research and its Implications for Digital Library Development," in *Digital Libraries*, ed. Judith Andrews and Derek Law, 5-17 (Aldershot, Hants, England: Ashgate, 2004).

5. National Science Digital Library, "Progress Report Fall 2000-Summer 2003," <http://nsdloutreach.comm.nsdl.org/annual_report/2003/NSDL.Rpt.final.Pgs_1-10.pdf>.

6. Clifford Lynch, "Where Do We Go from Here?" *D-Lib Magazine* 11, no. 7/8 (July/August 2005), <http://www.dlib.org/dlib/july05/lynch/07lynch.html>.

7. iLumina home page, <http://www.ilumina-dlib.org/>.

8. Marilyn McClelland et al., "Challenges for Service Providers When Importing Metadata in Digital Libraries," *D-Lib Magazine* 8, no. 4 (April 2002), <http://www.dlib.org/dlib/april02/mcclelland/04mcclelland.html>.

9. National Science Digital Library home page <http://nsdl.org/>.

10. Mischo, "United States Federal Support for Digital Library Research and its Implications for Digital Library Development," and Lee Zia, "Growing a National Learning Environments and Resources Network for Science, Mathematics, Engineering, and Technology Education," *D-Lib Magazine* 7, no. 3 (March 2001), <http://www.dlib.org/dlib/march01/zia/03zia.html>.

11. IMS Global Learning Consortium, Inc. "IMS Learning Resource Meta-data Best Practices and Implementation Guide v1.0," <http://www.imsproject.org/metadata/mdbest01.html>.

12. Ibid.

13. Innovative Interfaces, “XML Harvester,” <http://www.iii.com/mill/digital.shtml#xmlharvester>.

14. Clifford A. Lynch, “Institutional Repositories: Essential Infrastructures for Scholarship in the Digital Age,” *ARL Bimonthly Report* 226 (February 2003), <http://www.arl.org/newsltr/226/ir.html>.

15. Zia, “Growing a National Learning Environments and Resources Network.”

16. Christine L. Borgman, “What are Digital Libraries? Competing Visions,” *Information Processing and Management* 35 (1999): 231.

17. Lynch, “Where Do We Go from Here?”

18. Andreas Paepcke, Hector Garcia-Molina, Rebecca Wesley, “Dewey Meets Turing: Librarians, Computer Scientists, and the Digital Libraries Initiative,” *D-Lib Magazine* 11, no.7/8 (July/August 2005), <http://www.dlib.org/dlib/july05/paepcke/07paepcke.html>.

19. Ibid.

20. UNCW Randall Library, “Randall Library Mission Statement and Goals,” <http://library.uncwil.edu/goals.html>.

21. Open Archives, “The Open Archives Initiative Protocol for Metadata Harvesting,” <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

22. Lynch, “Institutional Repositories.”

23. Dublin Core Metadata Initiative, “What is the Difference between "Simple" ("unqualified") and "Qualified" Dublin Core?” <http://dublincore.org/resources/faq/index.shtml#whatisthedifference>.

24. iLumina, “Technical Documents for iLumina,” <http://www.ilumina-dlib.org/documents/>.

25. Paepcke et al., “Dewey Meets Turing.”

26. Open Archives, “DLESE OAI Software,” <http://www.dlese.org/oai/index.jsp>.

27. Terence Huwe, “My Three Wishes for Digital Repositories,” *Computers in Libraries* 25, issue 4 (April 2005).

28. Ibid.