USING METADATA TO BUILD AN ENRICHED LIBRARY CATALOG

How do libraries build an enriched library catalog using metadata? Libraries have been providing enrichment for centuries through their catalogs. The bibliographic or catalog record contained in the library online public access catalog (OPAC), the finding aids provided by archives and special collections, the item descriptions provided by museums—all these elements are enrichments of library items. Enrichments provide the user with surrogates for the item itself and assist in determining whether the user should retrieve the item from the shelf, ask to view the archival collection, or view the museum object.

These surrogate descriptions are metadata. So, if librarians have been providing enrichment to their users already through catalog metadata descriptions, how can they enrich their library catalog even further in the Web environment? Commercial ventures such as Amazon.com and Barnes & Noble have shown what metadata can do to enrich the access and description of print and digital items. Tables of contents, book cover images, reader reviews of the items, and access to an ordering and delivery system have enabled the success of these commercial ventures. Information organizations would do well to imitate these commercial information centers that use metadata to enrich the bibliographic descriptions of their collections.

OPAC vendors are just beginning to launch enriched metadata software that allows libraries to integrate their collections and services, print and digital, via a single search mechanism and related digital modules that link up to a library's digitized and digital collections as well as other local and remote resources.

Features desired in an enriched catalog

A library can include many features in a metadata-enriched catalog. Many libraries now offer enriched OPACs, similar to Amazon.com. They include images of book jackets, allow users to both read and add reviews of books to the bibliographic information, provide table of contents information, or merge other institution's metadata records with their own MARC records. OCLC has plans for an Extended WorldCat, a catalog that provides both enriched bibliographic records and expanded access to multilingual and other metadata records.

The Library of Congress has also set up an internal committee called the Bibliographic Enrichment Advisory Team (BEAT). BEAT is experimenting with Electronic Cataloging in Publication (ECIP), Business and Economics Online (BEOnline+), and digital tables of contents. The Library of Congress is working with many metadata standards but is specifically focusing on the ONIX standard for the digital tables of contents project and a NewBooks initiative in progress with participating publishers to provide information on soon-to-be-published and just-published books

NewBooks initiative, http://lcweb2.loc.gov:8081/ecip/celli/welc01.html

Many libraries are working on digitization and digital projects that are constructed both inside and outside the OPAC, with enriched metadata

records either embedded in a digital collection, or produced outside the collection in the OPAC, or both.

The University of Nevada, Las Vegas, launched its first digital project, called Early Las Vegas. The collection is set up like an art exhibit, with various galleries through which the user scrolls horizontally. Photographs, pictures, maps, documents, and blueprints are some of the digitized collections available in the digital galleries (Now & Then, The Depot, The Cottages, and Early Las Vegas). Each digitized item in the collection has been provided with embedded MARC metadata, which also has been incorporated into the library's OPAC so users searching for information on early Las Vegas also can access this source from there.

Early Las Vegas, www.library.unlv.edu/digproj

Some libraries have included bullets specifically related to a metadata-enriched catalog in their strategic planning. Cornell University, for instance, included the following in its Library Digital Futures Plan for July 2000 through July 2002:

Library Digital Futures Plan for July 2000 through July 2002, www.library.cornell.edu/staffweb/CULDigitalFuturesPlan.html

II. Services: Offering Enriched Access via a State-of-the-art Infrastructure and Assistance from Expert Staff

Provide the human and technical support needed to enable users to discover, navigate, critically evaluate, and effectively use digital collections and electronic resources.

A1. Convert finding aids to digital form.

A2. Add tables of contents to records for monographs in the online public catalog.

A3. Eliminate the remaining nonworking cataloging backlogs so that complete records are available to the public in the online catalog.

A4. Reduce significantly the number of catalog records that remain to be converted to machine-readable form so that users have increasingly only one stop to make—the online catalog—when searching for library holdings.

A5. Develop methods, including metadata (i.e, systems for identifying the structure of digitized documents to facilitate finding information within those documents), to improve the visibility and accessibility of library materials in all formats.

These ideas are just a sampling of ways that a library can initiate and implement the use of metadata, either imported by its own staff or exported from participating vendors and other libraries to provide enrichment to the library's collections and services through its OPAC or Web site.

Collaborative partnering

In the digital library and OPAC environments, development is underway for direct input of metadata—a concept that allows users to contribute metadata information directly and dynamically into information resources and services. This concept allows not only for a more direct interface among users, information providers, and the information itself, but it also allows users and information providers to partner as collaborators in creating information spaces and forums. Three useful examples of this type of collaborative information environment follow.

iLumina is a digital library of undergraduate teaching materials for mathematics, engineering, science, and technology being developed by a consortium of higher education groups and companies on the East Coast. In the article, "Challenges for Service Providers When Importing Metadata in Digital Libraries," the authors discuss the metadata standards they are using, and how iLumina imports Instructional Management System (IMS) metadata both from people and in batches. iLumina assists authors and creators wishing to include a resource in iLumina by providing a metadata template to fill out. In this way both creators and users of iLumina provide and access the metadata.

New concepts such as collaboratories and shariums are being incorporated into some digital library designs. In the article, "Collaboration Services in a Participatory Digital Library: An Emerging Design," the authors describe the construction of shariums—workspaces with powerful tools and rich content where people can collaborate with others or work independently to solve information problems, learn, and explore information resources—as part of the new design of digital libraries.

These shariums are learning environments that combine a local library with a collaboratory atmosphere, where users collectively engage in research by sharing information, similar to the visual image of a front porch on a house. One major project is the "American Front Porch," (AFP), which incorporates different levels of metadata at different qualities and quantities, from different servers, for many types of resources in different media. Besides providing examples of how this sharium would work, the authors describe types of collaboration services that would be offered, including collaborative information creation where any number of people could create a resource for their own use or donation to the AFP, collaborative information exploration between users or librarians, and synchronous and asynchronous interaction among teams of people. Collaborative metadata would play an important role in the construction and architectural structure of these shariums.

The Harvard Self-Enriching Library Facilities (SELF) Project is a digital library prototype comprised of user client software, a data store server, a metadata or catalog server, and a contribution acceptance server. This digital library allows users to contribute to the archives and add useful, organized information as a natural part of their research. In particular, users are able to add links to related works, to contribute their own data and results to the library, and to enter evaluative comments. SELF attempts to invest in tools to assist users in the capture of their knowledge, rather than devoting resources to one-time digitization of printed materials.

The article discusses the shortcomings of traditional library services and how this new model attempts to solve them. The status of this project is unknown (the paper was presented in 1994), but what is striking is that this initiative was developing before the explosion of the Internet in public life.

See also "Meta-design of a community digital library," www.dlib.org/dlib/may02/wright/05wright.html

"Challenges for Service Providers When Importing Metadata in Digital Libraries," www.dlib.org/dlib/april02/mcclelland/04mcclelland.html

"Collaboration Services in a Participatory Digital Library: An Emerging Design," www.ils.unc.edu/ils/research/reports/TR-2001-03.pdf

Harvard Self-Enriching Library Facilities (SELF) Project, www.csdl.tamu.edu/DL94/paper/harvard.html

Natural language recognition technology

The ability of the Web to present information to those who are challenged by audio and visual disabilities has become a near-term reality incorporating XML transport standards. The ability of computers to recognize written and spoken speech patterns, typically referred to as natural language recognition/generation, and translate these expressions into multilingual, understandable textual and audible presentations, is quickly being accomplished in the commercial and technology arenas.

To provide the mechanisms necessary to work in these new challenging environments that the Web has fostered, metadata is needed to facilitate and assist with unilateral translation, coding, transport, and presentation of information.

Unicode

Unicode standard, www.unicode.org

The Unicode standard is key to providing unfettered creation and access to multilingual documentation on the Web. With the proliferation of information and the Web has come the need for one encoding system to cover all the needs of all the languages of the world, as well as provide the interoperability and the standardization necessary to ensure usability.

The Unicode standard is a character coding system designed to support the worldwide interchange, processing, and display of the written texts of the diverse languages of the modern world. In addition, Unicode supports classical and historical texts of many written languages. While multiple encoding systems cover many of the world's languages, many of which did not translate or speak well among themselves, Unicode has become the lingua franca of the future digital world.

The idea for Unicode first began in the late 1980s at both Xerox and Apple Computer companies. By 1991 Version 1.0 had been developed, and Unicode is in Version 3.2.

Unicode is being implemented in many information organizations. The Library of Congress has been working to align MARC 21 to Unicode, and OCLC is moving the WorldCat database toward a global, multilingual metadata database with Unicode as the native data representation standard. OPAC vendors are incorporating Unicode into their strategic planning as well.

Endeavor Information Systems, for instance, has incorporated the Unicode UTF8 standard into its ENCompass 1.1 product, the only fully integrated system for accessing the local library catalog, local and remote electronic databases, and digital collections with combined search results sets and relevance feedback.

ENCompass 1.1 also incorporates XML and XSL for the display of metadata and related holdings for individual records. Other OPAC vendors are quickly developing similar products, moving toward a single search interface to search diverse bibliographic and metadata records, as well as incorporate the mechanisms for the multilingual database structure being developed in the globalization strategy of the WorldCat database by OCLC.

View an article on the Greenstone software at www.dlib.org/dlib/october01/witten/10witten.html.

Unicode Consortium may design its own generalized mechanism for Unicode metadata. View the results of these efforts at www.cs.fit.edu/~satkin/docs/uniescape.pdf.

Digital libraries are exploring the implementation of the Unicode standard for transport mechanisms and architectural structure. The Greenstone Digital Library, an open-source digital library software designed for the construction and presentation of information collections, is using Unicode this way.

The Unicode character set makes digital collections using the Greenstone software multilingual and international in scope and presentation. Information can be viewed in any language, as well as text-only for visually impaired users. Pictures, audio/visual clips, and music can be incorporated.

VoiceXML

The ability to translate natural language into audio format, and vice versa, has long been a dream of many people with vision or hearing disabilities. The W3C is developing and implementing VoiceXML, a suite of markup languages that control all aspects of interactive voice response applications, including dialog, speech recognition, call control, and speech synthesis.

VoiceXML is a dialog markup language for telephony applications and is designed to provide full control over the spoken dialog between the user and the application. Although other voice markup languages are available under different names (VXML and VoXML among them), the coordinated effort by the W3C to provide one widely standardized and international voice markup language becomes all the more crucial to the functionality of the Web.

VoiceXML is in Version 2.0, and a wide variety of implementers of this standard is available under Question 14 in the VoiceXML FAQ at www.w3.org/Voice.

The work on VoiceXML will eventually address the needs of blind and low-vision users and users with hearing or speaking impairments, but concern still exists about movement toward multimodal (dynamic and adaptable) interfaces—interfaces that will adjust themselves to the user's current environment and functional abilities.

Natural language recognition technology

Natural language recognition involves transforming language from its spatial form of graphical marks into its symbolic representation. For years the English language has been represented by ASCII text, but with the globalization and importance of multilingual access and dialog on the Web, the Unicode standard has become the preferred means by which text and language are represented in the digital environment.

The W3C is leading efforts to standardize natural language recognition efforts. It has many standards in the process of development. All these efforts relate to the W3C's efforts in metadata standardization as well.

A good introduction to natural language generation can be found at www.dfki.uni-sb.de/fluids/Natural_Language_Generation.html, where an extensive list of new achievements, latest commercial developments, research and market trends, and related information sources can be found. View a basic overview of natural language recognition at www.cedar.buffalo.edu/Publications/TechReps/WLR/wlr.html.

Metadata is a vital element in the implementation of natural language recognition technology into Web applications. The ability of computers to translate, synthesize, and transport various character and element sets involving numerous languages and speech protocols depends on the metadata that package and standardize that content. Information organizations will be incorporating Unicode and various software packages using natural language recognition technology into their computer systems and reference services in the near term, and the metadata that enables natural language technology.

Future use of metadata in technical services and cataloging

The explosion of information on the Internet has increased rather than decreased the need for experts in the description and organization of digital objects. The corporate and commercial world also has realized the benefits of describing and organizing internal and unique information into the digital environment. The many metadata standards indicate not only a lack of understanding concerning the expertise of information professionals but also a duplication of effort where others have already devised solutions and systems.

Information organization experts, especially those involved in technical services and cataloging operations, have a unique opportunity to actively market their skills in today's world. The business world wants people with knowledge organization and metadata experience to assist them in their digital and worldwide presence on the Internet. Although the titles of these positions reflect their origins in the corporate environment (knowledge manager, metadata specialist, information organization expert), the job descriptions are quite similar to those found advertised for librarians and archivists. The only difference is that the corporate environment pays two to three times more for these skills than do traditional information organizations.

The proliferation of interest in the development of digitization and digital projects has increased the need for those who know how to describe and organize information. Although MARC can and does fulfill the needs for most traditional information organizations in the print world, in the digital environment MARC is often a hindrance and lacks the degree of simplicity or complexity needed to describe and organize digital objects, depending on the project itself.

Technical services and cataloging personnel must begin to educate themselves in metadata and its application in the digital environment.

To stay on top of developments in metadata, librarians must:

- Actively seek and participate in digitization and digital project development and planning being initiated in their area
- Educate themselves and their colleagues about the importance of metadata and its benefits in the organization, description, access, retrieval, and preservation of digital objects
- Become active in the marketing of the skills and talents that information organizations have to assist the world as digital information increases in quantity and complexity