Enhancing Library Catalog Records with ONIX Metadata

Alba L. Scott

Library 281

Professor Karpuk

May 17, 2007
Abstract

Although greatly improved since the days of card catalogs, today's OPACs lag behind when compared to web search engines and the bibliographic information available through publishers and online book vendors' websites. This paper presents the Library of Congress’ Bibliographic Enrichment Advancement Team efforts to offer richer bibliographic information through its catalog by using and adapting metadata produced by the publishing industry and delivered in ONIX standard format.
Libraries have relied on their catalogs as the main source of finding holdings in their collections for many years. There is no doubt that catalogs have improved, and technology has provided the opportunity for much advance in this area. However, as Karen Markey points out, libraries failed to add "value to the online catalog but the only thing we changed was the catalog's medium." (¶ 4) In other words, the first OPACs offered the same information the old card catalogs offered; both librarians and systems developers missed the chance to offer more information beyond the simple bibliographic description already available.

Online searching now allows for keyword searches, and a user may not be so restrict when looking for a particular subject because of lack of knowledge of subject headings. But when compared to search engines such as Google or the bibliographic information available through Amazon.com, even the most sophisticated OPAC may seem a little antiquated.

Cataloging rules provide stability and harmonization to records, exactly what they are supposed to do. Descriptive cataloging offers the possibility to describe a record with detail that users can become as close as possible of materials without having them in hands. But library users have expected more from library catalogs from the very birth of the online catalog in the early 1980's, such as "subject searching improve[ment] … and table of contents". (Markey, ¶ 6)

In 1999 Steve Coffman wrote an article titled "Building Earth's Largest Library: driving into the future". In the article, Coffman uses the success of Amazon.com to draft a library system mirrored in the online wholesaler operations.

One of the issues pointed out in the article is of the library catalog and how Amazon has managed to bring the "book experience" closer to the user, even though there is no holding of the book, by "duplicating much of the experience of examining a book by fleshing out those bare-
bones catalog records with all kinds of information and content.” (Coffman, ¶ 15)

Although the world's largest library as conceptualized by Coffman may not be coming to existence soon, there have been many discussions concerning how the library catalog can be enhanced so library catalogs offer richer information about their holdings. Many larger libraries today offer table of contents, cover pictures, and links to related websites through their OPACs. And many libraries are also going a step further to offer a catalog that will fully integrate bibliographic information and the availability of full sources through their catalogs, just as the web does.

This paper will present the Library of Congress' Bibliographic Enhancement Advisory Team (BEAT) ONIX initiatives to enhance bibliographical records available through the Library of Congress catalog. The ONIX initiatives at the Library of Congress is one of many programs that addressing the possibilities of bringing more information attached to each record in the library catalog.

ONIX Metadata and Libraries

ONIX is a metadata standard used by publishers, book wholesalers, bookstores, and others involved in the industry to exchange bibliographical information of print and electronic publications. ONIX is a result of the combination of the Association of American Publishers' Guidelines for Online Information Exchange (ONIX) and EDItEUR's EPICS (EDItEUR Product Information Communication Standard) Data Dictionary. Both works were independent metadata specifications concerning the book trade industry (Cwiok, p.121), and a joint effort of EDItEUR, the Book Industry Communication, and the Book Industry Study Group oversee ONIX International. (Medeiros, p.114)

ONIX 's main objective is to "standardize the transmission of product information so that
wholesalers, retailers and other in the supply chain will all be able to accept information electronically transferred in ONIX International format.

(\url{http://www.bisg.org/onix/onix_faq.html}).

As the standard to communicate bibliographic information in the publishing industry, libraries are well aware of the richness of the information that can be acquired through ONIX records. An ONIX to MARC 21 mapping shows the relation between the fields in the two schemas. The full ONIX to MARC 21 mapping can be accessed at \url{http://www.loc.gov/marc/onix2marc.html}.

In 2001, the Cataloging and Classification Section of the Association for Library Collections and Technical Services (ALCTS) published its final report based on a task force charged to explore ONIX and its relation to library metadata, i.e., AACR2 and MARC21. Among its findings, ALCTS task force found that "all of the elements are present in ONIX that are needed in a MARC record to make a core (or full) record except for type of material and indicators." (ALCTS, p. 3). It is important to note that some aspects of ONIX differ greatly from traditional library cataloging. For example, ONIX has no requirement regarding authority control for such fields as author and subject. Because entry control is a major concern and one of the foundations of cataloging rules, using ONIX records may require lots of adaptation once a record is downloaded into a library system.

However, adapting library systems and using ONIX bibliographic information still benefit and speed the cataloging process, making items available more promptly to library users. As Roy Tennant explains, it is essential that a metadata infrastructure "should be capable of ingesting, merging, indexing, enhancing, and presenting to the user, metadata from a variety of sources…” And the ONIX record could be enhanced "with information form an OCLC record
when it becomes available." (Tennant, ¶ 8) The fact that some fields need to be changed once downloaded into the library system does not diminish the added value to the library catalog, nor affects the speed of the cataloging process.

While using bibliographic information available in an ONIX record can speed the cataloging process, other information contained in an ONIX file can offer information that is not traditionally available in the library OPAC. ONIX metadata comes with data such as table of contents, author information, and cover jacket image. It is this kind of information that can enrich users' experience while browsing a library catalog. The library of Congress Bibliographic Enrichment Advisory Team has taken the task of making some of this information available through the Library of Congress catalog by adapting its system to display richer information to its catalog users.

The Library of Congress BEAT Program

Created in 1992 by the Library of Congress Director for Cataloging, the Bibliographic Enrichment Advisory Team's main objective is to "conduct research and undertake initiatives to enhance the utility of bibliographic records." (Byrum and Williamson, p. 4)

Since its creation, BEAT has developed several initiatives that encompass the group's working realm. BEAT's initiatives are grouped in four main categories:

(http://www.loc.gov/catdir/beat/)

1. Extending a record: Linking Researchers, Catalog Records & LC Web data

Under this category are the following projects: dTOC Project, ONIX TOC, ONIX Reading Group Guides Project, ONIX Descriptions, ONIX Sample Texts, Links to Book Jackets, Contributor Biographical Information, and H-Net Reviews.

2. Enriching the content: Adding more information to catalog records
Under this category are the following projects: Adding Abstracts & Annotations, MARS Best Annotations & Reviews, Reviews of Reference Sources, HLAS Reviews, TOC for E-CIP, Machine Generated Tables of Contents

3. Enhancing Access: Linking the catalog to electronic resources

Under this category are the following projects: Web Access To Publications in Series, Web Access To Works in the Public Domain, and BeCites+.

4. Tools and Cooperative Projects

This category includes the following projects: Web Cataloging Assistant, INFOMINE iVia software project, Electronic CIP (E-CIP) support, BEOnline (Business and Economics Online) DCC/LCC correlations, support for the development and rollout of the NewBooks program, support for Portals to the World, and Pre-1970 Congressional hearings.

The table below shows the progress of the BEAT team for some of the above mentioned projects:

<table>
<thead>
<tr>
<th>Project</th>
<th>FY04</th>
<th>Total to date</th>
</tr>
</thead>
<tbody>
<tr>
<td>dTOC</td>
<td>8,370</td>
<td>25,444</td>
</tr>
<tr>
<td>ONIX TOC</td>
<td>5,830</td>
<td>48,783</td>
</tr>
<tr>
<td>ONIX Contributor Information</td>
<td>N/A</td>
<td>12,675</td>
</tr>
<tr>
<td>ONIX Publishers' Descriptions</td>
<td>51,791</td>
<td>109,279</td>
</tr>
<tr>
<td>ONIX Sample Texts</td>
<td>9,223</td>
<td>16,741</td>
</tr>
<tr>
<td>Web Access to Works in Public Domain</td>
<td>593</td>
<td>2,626</td>
</tr>
<tr>
<td>E-CIP TOC (TOC Web-linked)</td>
<td>35,819</td>
<td>35,819</td>
</tr>
<tr>
<td>E-CIP TOC (TOC in record)</td>
<td>10,490</td>
<td>28,639</td>
</tr>
<tr>
<td>H-NET Reviews</td>
<td>1,022</td>
<td>7,195</td>
</tr>
<tr>
<td>Web Access to Publications in Series</td>
<td>N/A</td>
<td>24,000 (links to monographs)</td>
</tr>
<tr>
<td>WebCat Assist</td>
<td>303</td>
<td>303</td>
</tr>
<tr>
<td>Web page hits on all TOC projects</td>
<td>N/A</td>
<td>3,600,000</td>
</tr>
</tbody>
</table>

BEAT's ONIX Initiatives

Because ONIX is the standard for the publishing industry, its data structure does not necessarily relate to the library systems based on the MARC format. As Michael A. Chopey points out, "the burden of making ONIX metadata records usable in or through the library catalog therefore fall naturally to libraries." (p. 29) So it is no surprise that the first efforts to use and import ONIX metadata to a library catalog has come from the Library of Congress through the BEAT program.

Currently, the BEAT team has five ONIX related ongoing projects: ONIX TOC, ONIX Reading Group Guides Project, ONIX Descriptions, ONIX Sample Texts, Links to Book Jackets, and Contributor Biographical Information.

ONIX TOC

Perhaps the most important of all the ONIX projects in progress in the Library of Congress, ONIX TOC is one of three tables of content initiatives underway through the BEAT team. The focus on the table of contents may come from the importance of TOCs in searching and retrieving information. Citing research done by Pappas and Herendeen, Byrum and Williamson explain that a table of contents enhances a bibliographic record by "[helping] users to determine the relevancy of particular titles; … greatly [improving] search effectiveness; [and] … providing content-indicative information." (p. 4)

ONIX TOC was made possible by the design of a Visual Basic program by David Williamson, a Library of Congress cataloging automation specialist. (LC table of contents project update, ¶ 3) Also, ONIX TOC is the most cost-effective of the three TOC projects, comparing to $40 a record for manually entering a table of contents into a record, to $2 per record for automatically processing table of contents provided by publishers through ONIX
metadata. (Byrum and Williamson, p. 9).

ONIX Reading Group Guide Project

The ONIX Reading Group Guide Project was developed in 2002 and it started by linking forty-four catalog records published by Houghton Mifflin to their corresponding reading group guide on the Internet. The idea of the project is to offer users links to reading groups and guides directly from the bibliographic record. As these are outside links provided by the publisher, the project will take the burden of watching the links to "determine the reliability of the URLs involved." (Beat's Reading Group Guide, ¶ 1) Below is a sample record with an external link to a reading guide for the book The Walking Tour.
ONIX Descriptions

The ONIX Descriptions project adds publisher-supplied descriptions of books to the bibliographic record. In the same fashion of the TOC project, the BEAT team uses an in-house created program to extract the data and make it available permanently in the catalog record.

(Chopey, p. 27) Below is a record sample with its corresponding publisher-provided description.
Enhancing Library Catalog Records with ONIX Metadata

ONIX Sample Texts

Through the ONIX Sample Texts project, the Library of Congress catalog offers a sample text supplied by the publisher to the bibliographic record in the catalog. As with the other ONIX related projects, this link is provided by the publisher in its ONIX record. The information provided is similar to that available through many publisher and e-book sellers, where an image type page that may contain the first pages of a chapter or parts of a book index. The BEAT team recognizes that the project "significantly enhances the information about a book that the Library of Congress makes available to a researcher."

(http://www.loc.gov/catdir/beat/onix.sampl.texts.html)
The example below is a sample text for the book The Walking Tour.

Sample text for The walking tour / Kathryn Davis.

Contributor Biographical Information

The Contributor Biographical Information project makes publisher-supplied information about a book's author, contributors, and other involved in the creation of the work. In 2005 there were 12,675 files processed into the library catalog where the contributor information was available. (http://www.loc.gov/catdir/beat/beat.html#tools)

Here is a sample of biographical information for a book titled The New Options Market.

Publisher-supplied biographical information about contributor(s) for The new options market / Max Ansbacher.
Conclusion

Libraries face a great dilemma with their catalogs. Today, it is not enough to offer the same bibliographic information made available through regular OPACs. The Internet and its possibilities have raised the standard and offer users not only access to full text of information but also more information about books and electronic publications than libraries are able.

When compared to the World Wide Web and publishers and retailers' online catalogs, library catalogs seem lacking content and resources. Library systems are not designed to support the same kind of information available elsewhere, and cataloging rules and format requires lots of modifications in order to make catalogs more flexible and useful.

The whole library is affected and changing in response to today's information structure. Use of library catalogs, as first point of access to library collections, must be in the forefront of concern and more initiatives must be developed. This issue is even more urgent as more information is available online and virtual collections grow. As Markey concludes, interested parties should be engaged so "serious dialog, system prototyping, decision making, and action so the online catalog of the future hits the ground just as mass digitization projects end." (¶ 35)

It will not be enough to change cataloging rules to incorporate the online environment and different means of information production. The library catalog interface must be reviewed and rethought too. As more people search and access information through the World Wide Web, the gap between the library catalog and the online environment becomes even larger.

The Library of Congress Bibliographic Enhancement Advancement Team has made great progress since its creation in 1992. And pioneer work in the library field has made information about books available through their catalogs that was only imaginable years ago. The use of publisher-provided metadata has enriched cataloging records, and surveys have shown that users
are looking for that kind of information.

Perhaps this project is a small step when compared to other initiatives and the big task of redesigning the library catalog. But the project shows that libraries can use metadata to enrich their records and provide more and better information to users. A library catalog where bibliographic and other information is available through a user-friendly library system, integrating both print and online resources is not only conceptually desirable but also possible. The BEAT project shows that libraries can use metadata available in the "market", offering services at a better cost. With the expertise of librarians and the technology available, the library catalog can become a powerful source of information.
References


