ON THE DUBLIN CORE FRONT

by Norm Medeiros
Coordinator for Bibliographic and Digital Services
Haverford College
Haverford, PA

Repurposed metadata: ONIX and the Library of Congress' BEAT Program
_____

"*When I was a boy of 14, my father was so ignorant I could hardly stand to have the old man around. But when I got to be 21, I was astonished at how much the old man had learned in seven years.*" – Mark Twain

ABSTRACT
This article reviews the ONIX-based efforts of the Library of Congress' Bibliographic Enrichment Advisory Team (BEAT). The article describes BEAT's table of contents, publisher description, and sample text initiatives, and the ways libraries and their patrons can benefit from these efforts.

KEYWORDS
Library of Congress; LC; Bibliographic Enrichment Advisory Team; BEAT; Online Information Exchange; ONIX; bibliographic records

Recently I took part in a study to investigate the kinds of URLs that appear in book records in my consortium's online catalog. I didn't know what to expect, since by rule our catalog librarians remove all 856 fields from records for resources other than journals and government publications. Occasionally, a stealthy URL makes its way to a public display, usually being discovered when a patron clicks the link expecting the full-text of the book, but instead receives much less. My fellow investigators and I wanted to know what the "much less" was, since lately our cataloging units had been deleting more and more links to it.

If not for the observation of a particularly attentive Haverfordian, I probably wouldn't care a whit about these book-based 856 fields, nor would I recognize how prolific these links had become, appearing in about every eleventh catalog record. Being a serialist at heart, my notion of an 856 field is strict and long-standing. At the conclusion of our study, however, my appreciation for links to data other than full-text had grown enormously. The aforementioned Haverfordian questioned the consortium's practice of blindly deleting linking fields, since he noticed that sometimes these links delivered useful content, such as publisher descriptions and sample text. Clearly the time was right to revisit the practicality of deleting these useful URLs.

THE EXPERIMENT

Our study looked at the number and types of 856 fields in bibliographic records for books. The original study reviewed three months worth of records near the end of 2003. For the purposes of this

article, I reran the report, looking at records entered in the system from 1 December 2003 through 29 February 2004. Of the 6,185 book records added during these three months, 541, or approximately 9%, contained 856 fields. A total of 737 links existed in these 541 records, about 1.36 links per record. The link breakdown follows:

Link type	Number of links	% of records	% of overall links
Tables of contents	457	84%	62%
Publisher descriptions	236	44%	32%
Sample text	28	5%	4%
Full-text content	10	2%	1%
Book reviews	3	1%	<1%
Misc.	3	1%	<1%
Total	**737**		

EXAMPLES

Most records contained either a single table of contents link; two links, one for a table of contents and one for a publisher description; or three links, one for a table of contents, one for a publisher description, and one for sample text. A record that contains all three of these links is *Extreme Ultraviolet Astronomy* by Martin Barstow and Jay Holberg (Cambridge University Press, 2003).
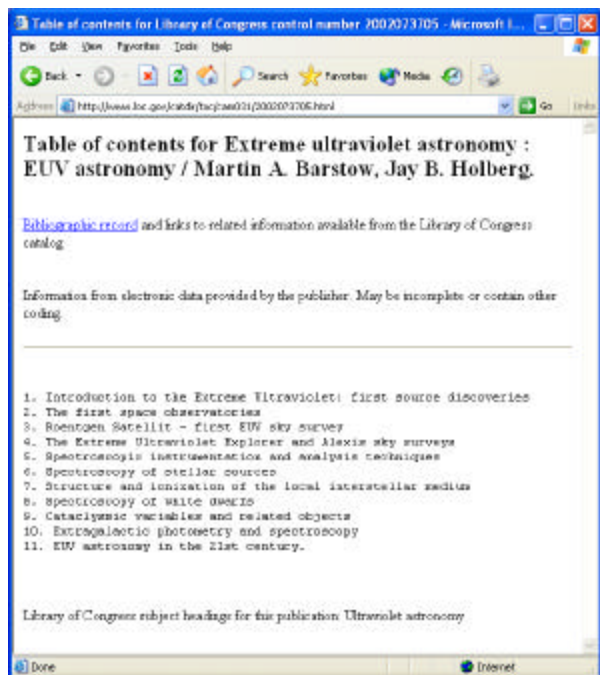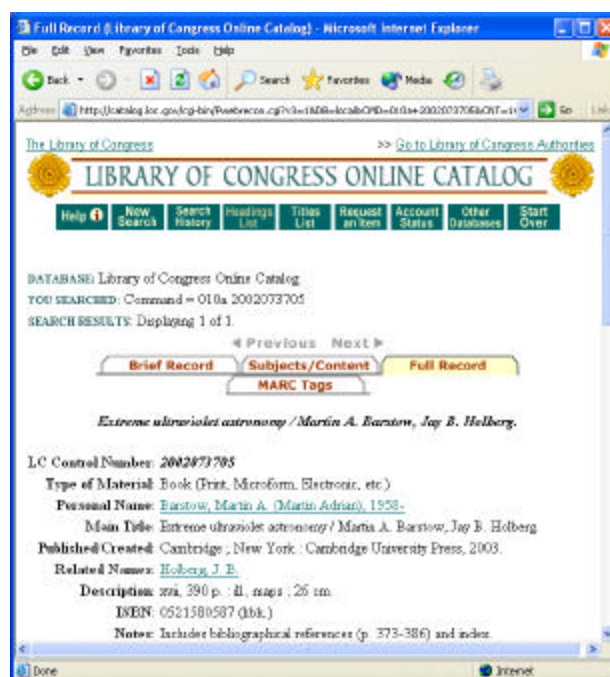

Figure 1: *Table of contents record*

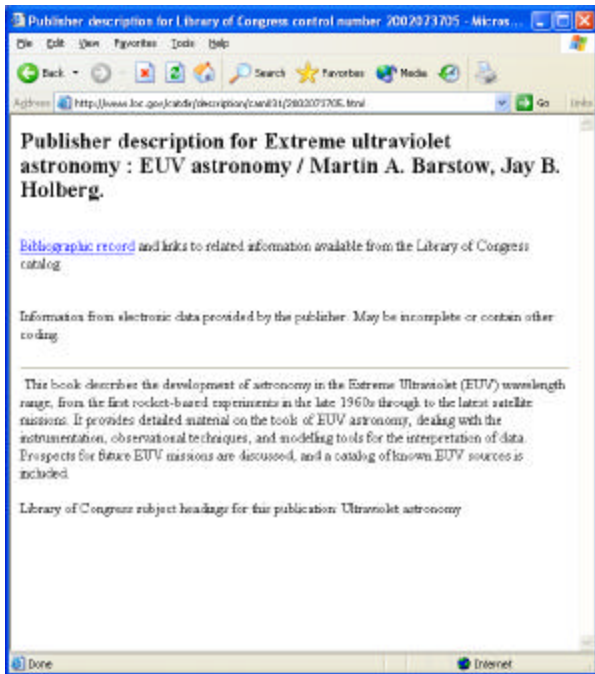Figure 2: *Bibliographic record in LC's catalog*

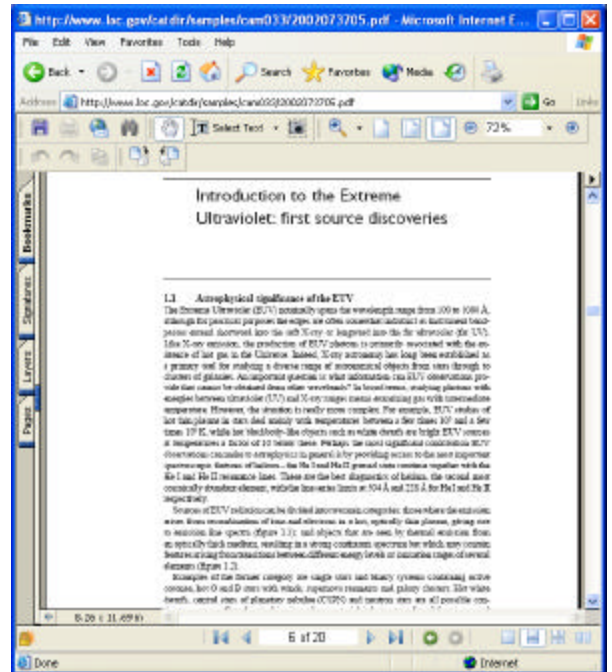Figure 3: *Publisher-provided description record*

Figure 4: *Publisher-supplied sample text*

The first link takes users to the table of contents for the book (figure 1). This record exists on a Library of Congress server, and includes a link to LC's bibliographic record (figure 2). The second link takes users to a publisher-supplied description (figure 3), which also exists on a Library of Congress server. The third link takes users to sample text supplied by the publisher (figure 4). This 20-page sample is in PDF format, and like tables of contents and publisher descriptions, is being hosted by the Library of Congress. These links point to data originally stored as ONIX files. ONIX, an acronym for ONline Information Exchange, is a publishing industry standard designed to describe books in order to facilitate their online sales. Like many metadata semantics, ONIX has been repurposed for other duties, including these Library of Congress initiatives. *Extreme Ultraviolet Astronomy* is published by Cambridge University Press, one of the publishing houses that utilizes the ONIX standard, and is therefore able to provide such data to the Library of Congress. Some other leading publishers that have adopted the ONIX standard are Houghton Mifflin, McGraw-Hill, Princeton University Press, and John Wiley & Sons. Additional information about ONIX is available in the 17:3 (2001) issue of *OCLC Systems & Services.*

LIBRARY OF CONGRESS' BIBLIOGRAPHIC ENRICHMENT ADVISORY TEAM

As noted above, the Library of Congress has actively repurposed publisher-supplied ONIX data for use with bibliographic records. LC's Bibliographic Enrichment Advisory Team (BEAT) program, a volunteer organization administered by the Cataloging Directorate, manages a number of bibliographic enrichment efforts, including three prolific ONIX-based initiatives: ONIX TOCs, ONIX descriptions, and the ONIX sample texts. As of 9 January 2004, BEAT has produced 40,000 ONIX TOC records, 60,000 ONIX descriptions, and 7,000 ONIX sample chapters. BEAT's efforts, however, are not limited to ONIX-based initiatives. In fact, ONIX TOC, ONIX Descriptions, and ONIX Sample Texts are three of nearly 20 innovative programs designed to enrich bibliographic data. BEAT's most recent effort is an extension of its Web Access to Publications in Series Project

<<http://www.loc.gov/catdir/beat/web.series.html>>, a project that increases access to working papers and technical reports in business and economics.

If you are involved in innovative projects that repurpose ONIX metadata, please send me an email <<nmedeiro@haverford.edu>> describing your project. I'll compile responses and publish them in a future *On the Dublin Core Front* column.

ACKNOWLEDGEMENTS