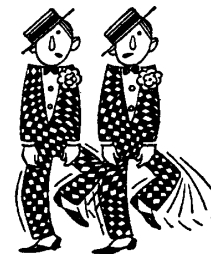


Standards Column — Transforming Metadata



by **Todd Carpenter** (Managing Director, NISO, 1 North Charles Street, Suite 1905, Baltimore, MD 21201; Phone: 301-654-2512; Fax: 410-685-5278) <tcarpenter@niso.org> www.niso.org

Metadata is among the most critical requirements of our community. It is the one thing that ties producer to purchaser, acquisition through management and curation, searcher to content, and reader to reference. Each supplier and user of metadata, though, has different needs, different formats, and different priorities for the metadata created and used. It is these subtleties that over time have led to a variety of approaches, a number of community-specific standards, and problems in quality within the chain of information from creator to library and end users.

Today the need to share metadata from different suppliers and creators is greater than ever, if for no other reason than because the creation, distribution, and useful integration of metadata are costly processes. Last year, in part in reaction to the significant costs of catalog record creation, the **Library of Congress** convened a **Working Group on the Future of Bibliographic Control** (<http://www.loc.gov/bibliographic-future/>). That group's report (<http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>) and the **LC** response (<http://www.loc.gov/bibliographic-future/news/LCWGResponse-Marcum-Final-061008.pdf>) both highlighted the need of the library community to rely more heavily on publisher-supplied metadata to reduce the tremendous costs within the library community of creating catalog records.

There are certainly challenges to this approach of building cataloging. Consider the differences between the **ONIX** data format and the **MARC** cataloging record format, partly due to the dissimilar purposes and uses of **ONIX** and **MARC**. For example, publishers use **ONIX** data to provide forthcoming information to booksellers that could significantly change by the final release of a text, while libraries want their **MARC** data to reflect the final publication. These issues, among many others, make the use of publisher supplied metadata in cataloging fraught with potential problems. Earlier this year, the **Library of Congress** announced a follow-up study to research and describe the marketplace for cataloging records in the **MARC** format to explore the economics of current practices and the incentives and barriers to sharing information.

Publishers, too, are focusing on the exchange of metadata and the costs within the publishing supply chain. The library community is only one recipient of their metadata. During the **Charleston Conference** last year, **Andreas Biedenbach** ([\[linkedin.com/pub/dir/andreas/biedenbach\]\(http://www.linkedin.com/pub/dir/andreas/biedenbach\)\), eProduct Manager Data Systems & Quality at **Springer Science + Business Media** \(<http://www.springer-sbm.de/>\) described the variety of organizations, to whom his departments distribute metadata — and the many formats that those organizations require. The list was long and the challenges many. It is not surprising that **Springer** has a large team focused on this issue. Likely, many publishers have similar teams invested in addressing the problems of distributing metadata to their community.](http://www.</p></div><div data-bbox=)

In an environment when controlling costs is a high priority for all organizations, the management and sharing of metadata can be an area of significant outlay. In several organizations who are intermediaries between suppliers and end users, there are large teams of people whose sole job is to clean and append information to publisher-supplied metadata. Obviously, there are significant perceived benefits and a return on the investments for improving the supplied metadata before it is passed on or made available to the broader community. Otherwise these organizations would not invest such significant resources in improving the data.

Improving the interchange of metadata was one of the main recommendations of the **Digital Libraries and Digital Collections Thought Leader** meeting that **NISO** sponsored in 2008. The **Thought Leader** meetings — funded in 2008 by a grant from The **Andrew W. Mellon Foundation** — were held with the goal of identifying and prioritizing new initiatives of importance to the information community. The group discussing digital collections suggested that **NISO** sponsor the creation of a suite of tools that publishers could use to assess the quality of the output they are supplying to the community. However, determining the costs and potential savings for publishers of both doing such assessments and improving quality is critical for justifying the investments that likely will be needed. If a compelling case is not made for a return on investment for publishers, it is unlikely that the publishing community would use any compliance tools and even more unlikely that they would invest in any improvements necessary to improve conformance with the various metadata standards.

To address these issues, **NISO** is co-sponsoring, along with **OCLC**, some research into the supply chain exchanges including the different needs of the various metadata supply chain stake-

holders and the inherent costs. This research will build a map of the supply chain, identifying the hand-offs of metadata between suppliers and recipient, the transformations that are done with the metadata before further hand-offs, and the costs to the community for transforming metadata. A key component of this project will be the exploration of potential solutions.

OCLC is organizing a by-invitation symposium in March to be hosted at the **OCLC** offices, that will bring together many of the key participants in the supply chain of metadata in the community. The initial research will be discussed along with the various needs of the organizations exchanging information. We hope that the discussions will identify potential solutions. Among these potential solutions might be an application of **OCLC's Next Generation Cataloging** (<http://www.oclc.org/partnerships/material/nexgen/nextgencataloging.htm>) **pilot project**. The goal of this project is “to explore upstream metadata capture and enhancement using publisher and vendor **ONIX** metadata”. Centralized federations of metadata are but one of many potential solutions to improving metadata. Another is the **Book Industry Study Group** (<http://www.bisg.org/>) and their **ONIX Data Certification Project** (http://www.bisg.org/documents/certification_productdata.html).

NISO's goal is to build understanding among the variety of players in this process of transforming metadata to fulfill the needs of the many different users and uses in the chain. The subtleties of differences in needs and the significant infrastructure investments made by different constituencies make it unlikely that the community can settle on one single data structure or transport mechanism. What is potentially more likely is creating standardized crosswalks and application profiles for different standards used in the community. Obviously, standards or best practices will play a role in the eventual solutions or improvements to the exchange of metadata. However, just as important will be a deeper understanding of the investments and the strengths that each participant in the exchange process brings to the table. Each constituency will have something to learn from the others in the chain, which might help reduce costs and improve functionality for everyone. 🐘