February 2009

Standards Column: Transforming Metadata

Todd Carpenter

NISO, tcarpenter@niso.org

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>

Part of the [Library and Information Science Commons](#)

Recommended Citation

Carpenter, Todd (2009) "Standards Column: Transforming Metadata," *Against the Grain*: Vol. 21: Iss. 1, Article 52.

DOI: <https://doi.org/10.7771/2380-176X.2535>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

eBook Intelligence: The 8th Annual Health Sciences Lively Lunch — Presented by **Sandra Wenner** (Assistant Director for Content Management, Rush University Medical Center Library); **Pam Harley** (ePublishing Strategy & Product Development, American Psychiatric Publishing, Inc.); **Deborah Ruck** (Information Resources Librarian, Medical College of Wisconsin Libraries); **Meg White, Moderator** (Director Technology Services, Rittenhouse)

Reported by: **Ramune K. Kubilius** (Northwestern University, Galter Health Sciences Library)
<r-kubilius@northwestern.edu>

After the introductions of moderator **White** and brief “highlights” of health publishing/library industry trends from the past year (by **ATG** session reporter **Kubilius**), panelist **Harley** revealed some eBook “secrets” from an association publisher’s perspective. Although her organization is “not-for-profit,” the publishing division is expected to turn a profit, maximizing the association’s brand. **APPI** products (their own platform) are designed for individual users, since 92% of the marketing of **APPI** products is to members. Associations also often have to deal with VIP author demands. **Ruck** provided examples of challenges in collection development/management of eBooks: e.g., a publisher requirement that faculty needs to license ancillary materials; bandwidth problems; difficulties with usage statistics and tracking eBook collection usage; different licensing start dates in the “brick by brick” model. She threw out the challenge to advocate for changes and communicate with publishers on what is wanted, needed. Lawyer and librarian **Wenner** provided some cautionary notes about licensing and reminded librarians of their responsibilities — leave yourself time, ask questions, make revisions, read carefully, watch for clause traps (copyright, statute of limitations, etc.). She shared some common misconceptions — you cannot ask for changes/deletions, everything is written in stone, “the vendor won’t like me if I ask for this.” Her conclusion? Contracts don’t kill anyone.

Just What the Doctor Ordered: A Remedy for Breaches — Presented by **Mary Ann Mahoney** (Head, Chemistry & Chemical Engineering Library, UC Berkeley); **Margaret Phillips** (Electronic Resources Librarian, University of California, Berkeley)

Reported by: **Miranda Schenkel** (SLIS Student, University of South Carolina, Columbia) <schenkem@mailbox.sc.edu>

Mahoney and Phillip’s presentation focused on licensing breaches and suggested best practices for publishers and librarians in the midst of dealing with violations. Because users may not be aware of restrictions on their use of databases, it is important to educate users on access limitations. Data and text mining are becoming more commonplace, as these methods are being used more as the nature of research changes. But how does one compel publishers, vendors, and access providers to view data and text mining as legitimate research? Future contracts may reflect these changes in the “academic use” of information, and perhaps allow a higher threshold for downloading information, as “excessive use” is the most common type of breach.

Just What the Doctor Ordered: A Remedy for Breaches

Second Report by: **Ann Marie Miller** (SLIS Student, University of South Carolina, Columbia) <annmarie.miller@gmail.com>

The speakers discussed scenarios where security was breached, usually accidentally, by searchers looking through online records. They suggested dealing with security breaches by giving users the benefit of the doubt, establishing a high threshold, limiting suspension to the single IP address, not asking for certification of deletion of data, not contacting multiple enforcement sources simultaneously, understanding the changing nature of research, and to not be restrictive out of fear, not to be a policeman, and don’t assume that patrons understand appropriate use.

The speakers took questions, and gave a number of real life examples where users breached the licensing terms by conducting searches that touched a large number of records without knowing that what they were doing was not a proper use of the system. 🐼

*That’s all the reports we have room for in this issue, but we do have more reports from the 2008 Charleston Conference. Watch for them in upcoming issues of **Against the Grain**. You may also visit the **Charleston Conference Website** at www.katina.info/conference for additional details.*

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** (<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.

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

continued on page 69

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 stakeholders 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. 🐘

From the University Presses — The Google Settlement: Boon, Boondoggle, or Mixed Blessing?

Column Editor: **Sanford G. Thatcher** (Director, Penn State Press, USB 1, Suite C, 820 N. University Drive, University Park, PA 16802-1003; Phone: 814-865-1327; Fax: 814-863-1408) <sgt3@psu.edu> www.psupress.org

Everyone seems to agree that the **Google** settlement announced in October 2008 represents a milestone of some kind in the development of access to information, but there is a wide spectrum of views about whether, overall, this is a good thing or a bad thing as far as the general public interest is concerned. Publishers appear to be as mixed in their opinions as librarians.

A lively debate is ongoing over the liblicense listserv on the merits of the settlement. **Rick Anderson**, in a posting on January 23, prefers to accent the positive: “Look at what the **Google** settlement has done: the general public now has far better (though still imperfect) access to vastly more literary and scientific writing than it ever has had before. This access is, by any sane definition of the term, free. (More comprehensive access is available at a price, but what's available at no charge is still amazing.) Even better, the content to which we now have access is, for the first time ever, fully searchable, and we can get it from our homes and around the clock. Better still, the public has paid virtually nothing in return for what it now gets.” To the skeptics, he says: “Sometimes I think we've actually made an art out of letting the perfect be the enemy of the good.” **Ann Okerson**, in her posting on December 17, also finds “commendable aspects” in the settlement and points out: “What I hear from readers is that they are waiting for the day when a click on a library catalog entry will take them directly to the full text of the item and speed up their ability to get information and do research. The **Google** partnerships and projects bring us closer to a version of that day, much sooner than

we could have imagined even five years ago. Is this good? Yes.”

Bernie Sloan, replying to **Okerson** on December 20, observes: “Sure, people are better off than they were five years ago as far as getting online access to book-based info. And that's a good thing. I don't think the critics are necessarily opposed to **Google Book Search** per se. I think the critics are wondering whether the ‘settlement’ is a step forward or a step back in the journey towards reaching **Ann's** goal.” **Bonnie Klein** worries, in her December 18 message, about the further corrosive effect of the settlement on rights that libraries have traditionally relied upon: “What is at stake are the current exceptions in copyright law — Sections 108, 109, and to a lesser extent 110 — that are key to library operations, whether brick or click. We are moving to accept as common general practice that every instance of online access may be controlled by the copyright owner [or authorized agent] and subject to toll or metered use. Over time this may undermine and erode the relevance and need for Title 17 exceptions.” And **Bernie Sloan**, on January 14, reminds us of the qualms **Siva Vaidhyanathan** had initially expressed about the settlement: “My major criticisms of **Google Book Search** have always concerned the actions of the university libraries that have participated in this program rather than **Google** itself.... Libraries at public

universities all over this country...have spent many billions of dollars collecting these books. Now they are just giving away access to one company that is cornering the market on on-line access. They did this without concern for user confidentiality, preservation, image quality, search prowess, metadata standards, or long-term sustainability. They chose the expedient way rather than the best way to build and extend their collections.... I am sympathetic to the claim that something is better than nothing and sooner is better than later. But sympathy remains mere sympathy...we must reflect on how complicit some universities have been in centralizing and commercializing knowledge under a single corporate umbrella.”

Others have more explicitly developed **Vaidhyanathan's** critique in terms of an alleged monopoly or quasi-monopoly that the settlement has effectively created for **Google**. **Robert Darnton**, writing about “**Google & the Future of Books**” in the *New York Review of Books* (February 12, 2009), concedes that “**Google** can make the Enlightenment dream come true,” but reminds us that “the eighteenth-century philosophers saw monopoly as a main obstacle to the diffusion of knowledge — not merely monopolies in general, which stifled trade according to **Adam Smith** and the Physiocrats, but specific monopolies such as the **Stationers' Company** in London and the booksellers' guild in

continued on page 70