

# The Serials Librarian

From the Printed Page to the Digital Age

ISSN: 0361-526X (Print) 1541-1095 (Online) Journal homepage: <https://www.tandfonline.com/loi/wser20>

## Can Richer Metadata Rescue Research?

Jennifer Kemp, Clare Dean & John Chodacki

To cite this article: Jennifer Kemp, Clare Dean & John Chodacki (2018) Can Richer Metadata Rescue Research?, The Serials Librarian, 74:1-4, 207-211, DOI: [10.1080/0361526X.2018.1428483](https://doi.org/10.1080/0361526X.2018.1428483)

To link to this article: <https://doi.org/10.1080/0361526X.2018.1428483>



Published with license by Taylor & Francis Group, LLC© 2018 Jennifer Kemp, Clare Dean and John Chodacki



Published online: 13 Jul 2018.



Submit your article to this journal [↗](#)






Article views: 417



View Crossmark data [↗](#)

## Can Richer Metadata Rescue Research?

Jennifer Kemp , Clare Dean , and John Chodacki 

Presenters

### ABSTRACT

This article reflects the discussion of metadata challenges faced by the scholarly communications community at this year's NASIG conference. Exploring the potential opportunities for the enrichment of metadata, we consider the evolving role of the metadata librarian and review opportunities for all librarians to contribute to an important new collaboration for the wider benefit of scholarly communications. The Metadata 2020 collaboration was the context for the session, which focused on barriers to progress and the role of nonbibliographic metadata in the evolution of library services.

### KEYWORDS

Metadata; collaboration;  
discoverability; advocacy

### Introduction

All of those who are involved with the publication of scholarly works have the same end goal: to conduct, facilitate, and/or communicate research. In the NASIG 32nd Annual Conference session “Can Richer Metadata Rescue Research?” we explored the research behind the insights guiding the Metadata 2020 collaboration at present. Metadata 2020 is a collaboration that advocates richer, connected, and reusable, open metadata for all research outputs, aiming to create awareness and resources for all who have a stake in creating and using scholarly metadata.

In particular, we looked at how research is affected by incomplete metadata.

Interviews with community stakeholders highlighted a few areas of focus:

- (1) Authors want increased visibility
- (2) Researchers need easier reproducibility
- (3) Funders and institutions are looking for better performance data
- (4) Publishers and service-providers need to demonstrate value with increased usage

In this article, we report on the discussion of the current role of librarians' use of metadata, the challenges faced in making use of enriched metadata, and the opportunities for broader influence in the scholarly communications community.

### Metadata for libraries: State of play

In the 2013 *Insights* article “How Libraries Use Publisher Metadata,” Steve Shadle indicates the following commonly seen areas of metadata applicability:

- (1) ARC-based library catalogs
- (2) OpenURL link resolvers
- (3) Library discovery systems

These three systems are not mutually exclusive since link resolution and machine-readable cataloging (MARC) catalog records often support library discovery systems. Library service has never exclusively been about the library catalog but it is even less so in the current service environment.<sup>1</sup>

These types of bibliographic metadata have long been a focus of metadata librarians. There are now a variety of programs and services designed to assist with the enrichment of bibliographic metadata, and how it can be made more discoverable:

Such projects include BIBRAME, the LODLAM community and the Linked Data for Libraries project. Through community engagement, advocacy, training and testing, they are exploring the role of standards in future-proofing bibliographic metadata improvements.<sup>2</sup>

But to what extent should nonbibliographic metadata also be considered by librarians? What other types of metadata (e.g., unique resource identifiers) should be considered both in the library and in institutional repositories? And how do we build better business cases for better metadata?

Discussion of nonbibliographic metadata at the NASIG session among the approximately 40 attendees revealed a preference for identity data, including organizational IDs, highlighting concerns regarding privacy issues and trusted third parties. Acquisitions metadata and related standards were also considered. Additionally, supporting data and data for nontraditional formats, (i.e., non-text) were discussed, resulting in further interest in digital assets management and preservation.

The discussion at NASIG reflected the chicken-and-egg nature of this push: Without systems that can handle additional metadata, how useful is additional metadata within the limits of MARC? And the larger questions for librarians remain: Is bibliographic metadata still a good enough minimum standard for our community? How might we broaden our perspectives on metadata within the scholarly communications community as a whole, collaborating to potentially re-purpose metadata expertise? Further, how should the library community re-think its position in the wider scholarly communications community to help facilitate richer, connected and reusable metadata for all research outputs within scholarly communications both in and beyond the library?

## **Obstacles to providing and integrating complete metadata**

The audience raised two longstanding issues as key obstacles: legacy data and indexing. First, the recognition that legacy data is difficult for the community to address. There was an understanding that adding more and better metadata to new content going forward is possible, but that enriching existing metadata records can be difficult and time-consuming. By the same token, the more content producers provide, the more useful it is likely to be. Second, indexing is key to the discussion of metadata quality and librarianship. Indexing is the other side of the metadata coin and must be included in discussions of how to make improvements. If such a longstanding, fundamental use of metadata such as indexing doesn't take full advantage of available elements, then discoverability is compromised and the push to enhance metadata may be called into question. Although it may have a particular meaning in scholarly communications "indexing" was one of a few topics that was raised in the session to be considered more broadly; in this case it could be shorthand for "re/use of metadata."

In addition to these two obstacles, audience members indicated that libraries also face multiple external obstacles such as insufficient metadata provision upon the delivery of content from publishers (e.g., data about article license information), prohibitive costs of vendor systems and services; and the absence of single-service options that provide multiple solutions (and that are not easily synchronized). Schneider echoes this view of interoperability:

Another issue is metadata vs. technology. As ontologies and standards are updated, systems that use those things are often built without thoughts of upgrades or new releases. Therefore, for one ontology, there may be many different sites using any number of different releases, rendering interoperability moot. (e.g., VIVO ISF)<sup>3</sup>

## **Fixing metadata is a top priority for publishers**

Slowly but surely, stakeholders across the scholarly communications chain are understanding the importance and responsibility of obtaining and integrating richer metadata. Funders, for example, are beginning to mandate that proper attribution is given to the funding grant when published research derives from their funding sources. For publishers, metadata ranks as the highest priority for publishers across all verticals but also represents the largest gap in current organizational ability. Determined to overcome key challenges and make strategic investments to accelerate their progress, 90% of all publishers are planning to invest in metadata over the next three years.<sup>4</sup> Publishers are not only incentivized by enhanced discoverability, but also by book sales. In the 2016 “Nielsen Book US Study: The Importance of Metadata for Discoverability and Sales,” a positive correlation between enhanced metadata and book sales is demonstrated.<sup>5</sup>

Regardless of individual publishers’ activities, cohesion, and integration with remaining systems and processes and new initiatives from service providers, funders, aggregators, data publishers, and libraries will surely be vital.

## **Librarians as well-positioned metadata facilitators**

### ***Librarians are experts***

The scholarly communications community is beginning to explore how to repurpose the skills and knowledge of metadata librarians. Some of the larger publishers—and in particular the large University Presses—are utilizing the knowledge of metadata librarians by employing them to assist with the organization of metadata and structure of taxonomies. At some research institutions, researchers are starting to use librarian expertise to assist in the application of metadata to their datasets. Librarians should consider how they can effectively reach beyond the library community to best use these skills.

Librarians, at least, spend far, far too much time talking to each other, and not nearly enough in creating personal relationships with researchers. We need to be at their conferences, speaking to them in their language, in order to create a shared and effective system of metadata.<sup>6</sup>

### ***Librarians make effective ambassadors***

The most direct way that librarians are involved in discoverability is, of course, through the catalogs and discovery systems of their own institutions. Their insights are invaluable to publishers and researchers; actively contributing to meetings and publications, benefitting the wider scholarly communication community as a whole. Librarians are also instrumental in the implementation of Open Researcher and Contributor ID (ORCID); both in encouraging researchers to obtain and use them, and in the integration of their own institutional ID into the article metadata they deposit. The Institutional Connect pilot program enables ORCID users to sign into their ORCID account using their institutional username and password. Libraries are also involved in the Shared Access Research Ecosystem (SHARE) initiative, which was formed through a collaboration between the Association of Research Libraries and the Center for Open Science. SHARE provides cleaning, linking, and enhancing metadata services.<sup>7</sup>

Are there other places in a research institution, or in the wider scholarly communication community, where librarian expertise could be harnessed for wider good? The Metadata 2020 collaboration plans to further facilitate discussions pertaining to this question.

## Conclusion: Opportunities for improvements

As initial focus-group research from the Metadata 2020 collaboration indicated, it is important to focus on the researcher and their needs when considering metadata. Quality and completeness are considered more important than timeliness, within reason, given the need for speedy publication in some fields of research.

The opportunities to improve metadata across the scholarly communication community are significant and far-reaching. If the wider community can overcome technical, economic, and social obstacles, we will be able to reuse and remix the metadata to:

- (1) Map interconnections
- (2) Balance consistency and flexibility
- (3) Provide a common, distributed resource for all

A group of organizations (including Crossref, DataCite, ORCID, OpenAIRE, California Digital Library, Wikimedia, and OCLC Online Computer Library Center, Inc. among others) have joined forces to rally and support the community around this issue. Metadata 2020 is a collaboration that advocates richer, connected, and reusable metadata for all research outputs. The initiative is based on the beliefs that:

- (1) Richer metadata fuels discoverability and innovation;
- (2) Connected metadata will bridge the gaps between systems and communities, and
- (3) Reusable metadata eliminates duplication of effort

Interest in and appreciation of Metadata 2020 was clear at this session (the points raised were incorporated into this article), as at other librarian and publisher events held earlier in 2017. Metadata 2020 seeks active support, advice, and participation from librarians as ambassadors and educators in scholarly communication. In particular, the collaboration seeks use cases, stories, and perspectives on metadata challenges and opportunities from the librarian community, and encourages contributions be sent to [info@metadata2020.org](mailto:info@metadata2020.org). More broadly, we encourage every community within scholarly communication to give thought to how they might help provide and facilitate richer metadata while assisting other communities in doing the same.

## Notes

1. Steve Shadle, "How Libraries use Publisher Metadata," *Insights* 26, no. 3 (2013): 290–297. <http://doi.org/10.1629/2048-7754.109>.
2. Sarah Bull and Amanda Quimby, "A Renaissance in Library Metadata? The Importance of Community Collaboration in a Digital World," *Insights* 29, no. 2 (2016): 146–53. [doi:10.1629/uksg.302](https://doi.org/10.1629/uksg.302).
3. Juliane Schneider, "Thoughts on Metadata," interview by Clare Dean, 22 June 2017.
4. During the session, heartening results of a recent study of STM publishers were discussed: "Industry Leaders' Perspectives on the Digital Transformation Journey in Publishing," *IMBUE Partners* (2017): 6.
5. David Walter, "The Importance of Metadata for Discoverability and Sales," *Neilsen* (2016).
6. Schneider, interview.
7. "About SHARE," SHARE Research, <http://www.share-research.org> (accessed June 22, 2017).

## Acknowledgments

We thank the Metadata 2020 steering committee, Juliane Schneider from Harvard Catalyst, and the attendees of the NASIG 2017 "Can Richer Metadata Rescue Research?" session for their contributions.

## ORCID

Jennifer Kemp  <http://orcid.org/0000-0003-4086-3196>

Clare Dean  <http://orcid.org/0000-0003-3922-6077>

John Chodacki  <http://orcid.org/0000-0002-7378-2408>

## Notes on contributors

*Jennifer Kemp* is Outreach Manager at Crossref, Lynnfield, MA, USA.

*Clare Dean* is an Independent Consultant and Outreach Manager at Metadata 2020, Somerville, MA, USA.

*John Chodacki* is Director of the University of California Curation Center, Oakland, California.