Information Services & Use 40 (2020) 177–179
DOI 10.3233/ISU-200087
IOS Press

A conversation about linked data in the library and publishing ecosystem

John Chapman*
Senior Product Manager, OCLC, Inc., Dublin, OH, USA

Abstract. During the inaugural 2020 NISO+ conference, the "Ask the Experts about… Linked Data" panel included discussion of the transition of library metadata from legacy, record-based models to linked data structures. Panelists John Chapman (OCLC, Inc.) and Philip Schreur (Stanford University) were the speakers; NISO Board of Directors member Mary Sauer-Games (OCLC, Inc.) was the facilitator. The event was an open-ended conversation, with topics driven by questions and comments from the audience.

Keywords: Linked data, structured data, MARC, BIBFRAME

1. Introduction

On 24 February 2020, Philip Schreur (Stanford University) and John Chapman (OCLC, Inc.) were the featured panelists for the "Ask the Experts About… Linked Data" session, facilitated by Mary Sauer-Games, held as part of the inaugural "NISO+" Conference in Baltimore, Maryland, USA.

In brief introductory remarks, Chapman and Schreur described grant-funded projects that address different parts of the linked-data environment. Chapman is the grant manager for the Entity Management Infrastructure project at OCLC [1], funded through a January 2020 grant from the Andrew W Mellon Foundation and matching internal funds. The session facilitator, Sauer-Games, is the Principal Investigator for the grant. Schreur is Principal Investigator for a series of grants also from the Mellon Foundation, most recently awarded in April 2020, under the Linked Data For Libraries or 'LD4L' [2] banner.

The OCLC project is primarily designed to provide infrastructure in the form of persistent URIs (Uniform Resource Identifiers) for creative works and persons and providing descriptive information and links to related resources. The infrastructure, which is scheduled for a public release in December 2021, is designed to support the linked -creation, interface building, and data modeling work being pursued across the global library ecosystem.

One of the premier linked-data efforts in the library world, the LD4L grant projects have investigated a number of topics, including Application Programming Interface (API) and web service development, ontology creation, and end user interface development. Incorporating within their cohort a number of libraries who are members of the "Program for Cooperative Cataloging" (PCC) [3], the project's recent "Linked Data for Production" grants have sought to progress new linked-data approaches for cataloging

*Tel.: +1 614 764 6000; E-mail: chapmanj@oclc.org.

and technical services workflows. This experimentation has been largely focused on 'BIBFRAME' [4], an entity-relationship model that is being adopted for bibliographic data.

2. Moving linked-data efforts forward

Chapman and Schreur described the desire to advance the field beyond its initial phase, which has been dominated by conversion of legacy data, usually in the relatively 'flat' Machine Readable Cataloging (MARC) format [5], into models such as BIBFRAME. This initial phase has brought to light the complexity and long, evolving history of not only the MARC format, but also of its underlying semantics. Schreur mentioned the work under LD4L to better model recorded music resources [6], and the challenges that this complex area presented. Both speakers mentioned the leadership of several European libraries, including the Bibliothèque nationale de France (BnF), Deutsche Nationalbibliothek (DNB), and the National Library of Sweden, in managing their library data "natively" in new data structures not limited by the MARC data model. The challenges of moving beyond this to a shared effort among many participating libraries are shared by both grant projects.

Asked to provide a brief description of linked-data, Chapman stressed that it is describing materials and data using terms that everyone can define and put in context using the Web. It privileges identifiers over free text and uses structured data to convey relationships. Schreur commented that it enables machine interoperability in a drastically more powerful way than text-based descriptions.

Given the promise of linked-data, the leadership of European libraries, and the new models such as BIBFRAME, what needs to shift for a more widespread adoption? The panelists agreed that common reference points are important, and that BIBFRAME represented a center of gravity that could help align disparate projects. Chapman cautioned that there are significant challenges to pulling together metadata for published material, realia, and objects, and collections of documents and images. Schreur mentioned the challenges around balancing collective description, at the network level, while allowing for local customization. Some of the benefits of centralized metadata, including serendipitous discovery across multiple collections, efficient indexing, and traffic analysis, are harder to engineer across local, heterogeneous collections.

Another area of exploration related to linked-data is machine learning. In response to an inquiry from the audience, Chapman said that the focus for OCLC's current work was to improve metadata for the use of end user discovery, and that formalizing it in order to make it capable of being processed by inference engines was out of scope. Schreur said that the first goal needed to be more complex and meaningful relationships between resources. Using the example of recorded music, he described the rich set of relationships that could be modeled - the composer, the musical composition, the concert at which it was recorded, the performers, and the genres or topical connections that could also be connected. Communicating all of this in MARC is difficult or impossible - the promise of linked-data is that we do not have to transcribe this into records; instead we can link to sources that offer us identifiers and context for these other resources.

Chapman agreed, saying that the MARC standard was originally created for the purposes of inventory control, and then became the basis of user searching via electronic indexes. However, when these catalogs were networked together, the systems of access points and authority control could not keep up with user expectations.

Meeting these expectations will take some significant workflow and cultural changes among libraries. Many libraries, including those participating in or advising the projects that LD4 and OCLC are working on, are very eager to move ahead. But the switching costs of new approaches are real. Chapman mentioned

this as a key dynamic in the library cataloging world, particularly among smaller libraries, or those with less staffing flexibility.

Given this issue, both panelists mentioned the participation of vendors, especially aggregators and publishers, in adopting new models. From the library point of view, receiving data as linked data - or at the very least, with consistent linking to linked data sources - means less time spent in lossy [7] conversion routines. From their perspective, libraries see this as metadata work "moving upstream", and it is a critical factor in consistency, avoiding multiple and possibly divergent conversion routines.

MARC will likely remain for some time in its core function of inventory management, as it works well for acquisition workflows. Schreur said that at the start of the series of grants he has managed at Stanford, he was expecting to completely move beyond MARC for nearly all data in the library. Now he expects a hybrid approach, leveraging linked-data for discovery, and retaining MARC for some time for those workflows relating to inventory.

3. Conclusion

In closing remarks, Chapman and Schreur described much of the value of participating in linked-data initiatives as coming from the discipline and rigor of structuring data in ways that machines can easily process. Thinking this way leads to a new perspective on the importance of consistency, collaboration with partners, and participating on the web. By reusing the work of others, using the vocabularies that they use, and cooperating on data sharing and modeling, libraries and their vendor partners can enable new efficiencies and partnerships. These will be critical in ensuring that library collections continue to be discovered and used.

References

[1] *OCLC Awarded Mellon Foundation Grant to Develop Infrastructure to Support Linked-Data Management Initiatives* [Press release], OCLC, Inc., Dublin, Ohio, [updated January 9, 2020], Available from: http://oc.lc/mellon-grant, last accessed June 6, 2020.

[2] Home - LD4L [homepage on the Internet]. LD4L, Available from: https://www.ld4l.org, last accessed June 6, 2020.

[3] Library of Congress. *Program for Cooperative Cataloging*. Library of Congress, Washington, D.C., 2020, Available from: https://www.loc.gov/aba/pcc, last accessed June 6, 2020.

[4] BIBFRAME – Bibliographic Framework Initiative. Library of Congress, Washington, D.C., Available from https://www.loc.gov/bibframe/, last accessed June 6, 2020.

[5] MARC standards (Network Development and MARC Standards Office). Library of Congress, Washington, D.C. [updated 2020 March 13], Available from https://www.loc.gov/marc/, last accessed June 6, 2020.

[6] Performed Music Ontology – LD4P public website – LYRASIS Wiki. Atlanta: LYRASIS [updated 2019 January 13], Available from https://wiki.lyrasis.org/display/LD4P/Performed+Music+Ontology, last accessed June 2, 2020.

[7] "lossy", *Merriam Webster Dictionary,* https://www.merriam-webster.com/dictionary/lossy, last accessed June 6, 2020.