

Linked Data for Production (LD4P): A Multi-Institutional Approach to Technical Services Transformation

Philip E. Schreur
Stanford University
557 Escondido Mall, Stanford, CA
United States
pschreur@stanford.edu

ABSTRACT

Linked Data for Production (LD4P) is a collaboration between six institutions (Columbia, Cornell, Harvard, Library of Congress, Princeton, and Stanford) to begin the transition of technical services production workflows from a series of library-centric data formats (MARC) to ones based in Linked Open Data (LOD). This first phase of the transition focuses on the development of the ability to produce metadata as LOD communally, the enhancement of the BIBFRAME ontology to encompass the multiple resource formats that academic libraries must process, and the engagement of the broader academic library community to ensure a sustainable and extensible environment. As its name implies, LD4P focuses on the immediate needs of metadata production such as ontology coverage and workflow transition. The LD4P partners' work will be based, in part, on a collection of tools that currently exist, such as those developed by the Library of Congress. The cyclical feedback of use and enhancement request to the developers of these tools will allow for their enhancement based on use in an actual production environment. The six institutions involved will focus on materials ranging from art to rare books, from cartographic materials to music, from annotations to workflows. Tool development and enhancement will also be a key aspect of the project. By the end of the first phase of this project (Spring 2018), the partners will have the minimal tooling, workflows, and standards developed to begin the transformation from MARC to LOD in Phase 2 of the project.

ACM Reference format:

Philip E. Schreur. 2018. Linked Data for Production (LD4P): A Multi-Institutional Approach to Technical Services Transformation. In *The 2018 Web Conference Companion (WWW 2018)*, April 23-27, 2018, Lyon, France, ACM, New York, NY, 2 pages. DOI: <https://doi.org/10.1145/3184558.3186201>

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.
WWW '18 Companion, April 23-27 2018, Lyon, France
© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.
ACM ISBN 978-1-4503-5640-4/18/04
<https://doi.org/10.1145/3184558.3186201>

BASIC PROJECT INFORMATION

Name of the presenter: Philip E. Schreur, Stanford University

Project duration: Phase 1: 04/2016-03/2018 ; Phase 2: 07/2018-06/2020 (recently approved)

Project Volume: Phase 1: 1.5 million dollars ; Phase 2: 3.5 million dollars (recently approved)

Funding Agency: The Andrew W. Mellon Foundation

Official Website:

<https://wiki.duraspace.org/display/LD4P/>

IMPORTANCE AND TIMELINE

Libraries' data architecture is based upon a series of communication formats developed in the 1960s called MACHine-Readable Cataloging, or MARC. It has become the de facto standard for how libraries ingest and communicate data to each other. It has also become the standard on which libraries' Integrated Library Systems (ILS) are built and with which library vendors communicate their data to libraries. It has become apparent, however, that library patrons have been turning to the Web, especially the semantic web, in search of data. It is essential that libraries transform their antiquated technology stack and move onto the Web as their means of disseminating, and gathering, data. The transition will be a highly disruptive one and must move forward as quickly as possible as the shift to the Web for most library patrons has already occurred. LD4P hopes to develop the core technologies to make this transition possible by the end of the current grant cycle and to move a core group of libraries into implementation phase by the end of the next grant cycle (2020).

CONTENT OF THE PRESENTATION

The presentation will focus on three main areas. The first will be the problems associated with the MARC formats and why it is essential for libraries to shift to a standard Web means of ingesting and disseminating data. The second will focus on the individual projects of the six institutions and how they contribute to this transformation. The last will highlight developing issues

that must be resolved as the group moves into its next phase of development.

DURABLE RESULTS

Thompson, Timothy A., Jennifer Baxmeyer, Joyce Bell, and Peter Green. 2016. From endnotes to annotations: deconstructing dedications in the library of Jacques Derrida at Princeton University. *Journal of Library Metadata* 16, issue 3-4: 146-165. doi: <http://dx.doi.org/10.1080/19386389.2016.1258908>.

McCallum, Sally. 2016. BIBFRAME and linked data for libraries. In *Linked data for cultural heritage (an ALCTS monograph)*, ed. Ed Jones and Michele Seikel. Chicago: ALA Editions.

Schreur, Philip E. Linked data in libraries' Technical Services workflows. 2017. *Communications in Computer and Information Science* 755. Heidelberg: Springer. (In press)