

MARC to BIBFRAME: An Exploration of the Future of Cataloging MLA Annual Meeting 2015, Denver, Colorado

Summary written by Elizabeth Cribbs, Northern Illinois University

Session presenters: Kimmy Szeto, Baruch College CUNY; Casey Mullin and Nancy Lorimer, Stanford University; Michael Colby, UC-Davis

This session, sponsored by the MLA Bibliographic Control Committee, explored the current status of Bibliographic Framework Initiative otherwise known as BIBFRAME and what certain groups are learning about how it might work. The full session can be viewed here:

http://www.musiclibraryassoc.org/page/mla_2015_media

Kimmy Szeto, chair of the BIBFRAME Task Force, first provided an overview of what BIBFRAME hopes to accomplish and where the BIBFRAME Task Force is in that process. He began with an explanation of the four building blocks required for cataloging systems to work: content standards, schema, serialization, and exchange systems. BIBFRAME is designed to replace MARC (our current schema) by breaking the information that currently comprises our bibliographic records apart into atomic units that can then be reconfigured and reassembled by our library systems into various presentations. BIBFRAME currently allows us to either convert current MARC records into BIBFRAME format by using the MARC to BIBFRAME Transformation Service within MARCEdit or create a new BIBFRAME “record” by using the BIBFRAME Scribe workform that uses a born-digital platform and many fields that already tie into controlled vocabulary lists. So far BIBFRAME’s development has been led by the Library of Congress, Zepheira, and other implementers and developers. Szeto then went through BIBFRAME’s timeline, the various discussion papers and reports that have been issued so far, and the current and future plans for the task force.

Casey Mullin then presented the work performed so far with linked data projects designed to explore how libraries can use linked data and the Semantic Web to improve discovery and access to scholarly information. In 2014 Stanford University, the Harvard Library Innovation Lab, and Cornell University were awarded a Mellon Foundation grant to create a Scholarly Resource Semantic Information Store model. This project, named Linked Data for Libraries (LD4L, website <https://wiki.duraspace.org/pages/viewpage.action?pageId=41354028>), seeks to link three large sources of data about scholarly resources: bibliographic data, person data, and usage data; and to connect library resources with institutional and other data on the web. LD4L also intends to provide a transparent mapping from the MARC records that currently house much of this information to Solr via BIBFRAME. The project eventually hopes to produce an open source LD4L ontology compatible with BIBFRAME and other LOD efforts, an open source LD4L semantic editing, display, and discovery system, and a Project Hydra compatible interface to LD4L.

Meanwhile, 15 staff members at Stanford tested the LC MARC to BIBFRAME converter with many different formats and permutations and read and discussed many BIBFRAME white papers. While the converter was not quite as far along as the staff members had hoped and the MARC “mindset” provided numerous possible conceptual traps, the benefits produced included highlighting previously inadequately represented data and improving the efficacy of the converter. Once this was done, 24 Stanford staff members underwent Zepheira training and were

thus able to compare and contrast the Zepheira and LC converters. Since LD4L was already developing an ontology and linked data framework including BIBFRAME, the Stanford Technical Services department decided that the next step should include exploring how linked data would work in a technical services workflow environment. Three workflows are being considered: copy cataloging from vendor records, original cataloging, and music sound recordings. However, creating usable BIBFRAME work and instance data from MARC records will have many challenges and probably require considerable editing. Many decisions and questions still need to be considered: How will the cataloging be entered into BIBFRAME? How do we expand the vocabulary? How will new authorities link to the work and instance data? How much editing and cleanup will be required to make the information usable? To help with beginning to answer these questions, a group of six libraries—Stanford, Cornell, Columbia, Harvard, Princeton, and the Library of Congress—met at ALA to begin planning how to develop the production, workflows, tools, and tests necessary to bring BIBFRAME and other linked data projects to a usable place for themselves and other libraries.

Finally, Michael Colby discussed the project BIBFLOW, in which his institution has attempted to answer the question “What might adoption of BIBFRAME mean to technical services workflows in an academic library?” To try to answer this question, the University of California-Davis obtained a grant from the Institute of Museum and Library Services and partnered with Zepheira and Kuali to begin to develop a roadmap that would try to focus on academic library technical services processes and explore the impact of new standards on related library operations such as circulation, ILL, and discovering, selecting, and obtaining resources. They identified and collected test data, mapped it, and explored the conversion and ingestion of test data while they also developed and tested a prototype of a discovery display system and a BIBFRAME-based transfer and exchange system. The project has four phases over twenty-four months and will finish in April of 2016. The deliverables will include sample test data sets, prototype discovery and display system code, links to related projects, and project reports, some of which will be available soon on the project’s website (<http://www.lib.ucdavis.edu/bibflow>).

BIBFLOW’s focus is on developing a roadmap for migrating essential library work efforts to a BIBFRAME/LOD ecosystem, and the complexity of the workflows involved has led to the conclusion that linked data requires an evolutionary leap and not a simple migration. Moving forward, UC-Davis and Zepheira will enhance the BIBFRAME Scribe by adding external services and developing BIBFRAME profiles, program the Kuali-OLE product so that UC-Davis users can use BIBFRAME-Scribe to describe various materials and store data in the BIBFRAME-RDF triplestore, develop and test data transformations services/tools, and identify and connect an open-source OPAC to the triplestore.