NASIG
TRANSFORMING THE INFORMATION COMMUNITY

BIBFRAME at the Doorstep

Lisa McColl and Greg Edwards

Abstract

As several large institutions in the United States, most notably the Library of Congress, move forward with efforts to migrate from MARC21 to BIBFRAME, the staff of Lehigh University Libraries is engaged in early stages of migration as well. Lehigh University is a medium-sized institution with an academic library that has two full-time professional metadata librarians on its staff. Considerations of most aspects of BIBFRAME fall into the hands of these few professional librarians prior to migration. This presentation is the third in a series of presentations intended to capture current considerations, questions, and recommendations in planning for BIBFRAME at Lehigh University, with the hope that peer institutions will find the discussion relatable and informative.

Keywords: cataloging, MARC21, BIBFRAME migration

Lehigh University Library Environment

A library's environment in terms of size, technology, and mission will have a direct impact on its staff's approach to BIBFRAME. In 2024 the undergraduate student full-time equivalent of Lehigh University was 5,515 and the graduate was 1,281.[1] The library staff is made up of nineteen professional librarians, eleven staff members, and four developers who are dedicated to the library. Of these nineteen professional

librarians, two are full-time metadata librarians. Because migrating to BIBFRAME is such a large endeavor in terms of time and cost, the value of that change needs to be communicated to administrators. For institutions with a similar staff size, this puts a large responsibility on few people. As Han, Heng, and Lampron pointed out in their 2023 presentation, "Considerations of Designing a User Friendly BIBFRAME Editor," catalogers are now assuming new roles as "coordinator, facilitator, and more."[2] Furthermore, they note that new skills are required to lead this change. The smaller a library staff, the more of these new roles a cataloger may need to assume. The effect of this at Lehigh University is apparent in the range of very specific to broad issues that the cataloging staff has already begun to address.

Lehigh University has been using the open-source platform FOLIO for its Library Services Platform (LSP) since 2020. The Library of Congress adopted FOLIO in 2022.[3] Since the Library of Congress is one of the leaders of the shift from MARC to BIBFRAME in the United States, their move to FOLIO accelerated the need and therefore development of a BIBFRAME editor for FOLIO. At Lehigh, this changed BIBFRAME from a long-time theoretical model to something that was growing closer to practical application. The cataloging staff and library developers began discussions about implications that BIBFRAME has for discovery. Lehigh Libraries uses the open-source discovery system, VuFind, for its online catalog and is fortunate to have a developer on staff who contributes code back to the VuFind project. With support from Index Data as the library's FOLIO host and in-house support from the library's developers, as well as administrative support for experimentation, the cataloging environment at Lehigh is highly flexible. In addition, Lehigh University Libraries became a member of Share-VDE in January 2024. Share-VDE is a "community of libraries in a shared discovery environment based on linked data."[4] One of the first steps after becoming a member is to send test MARC records to be converted to BIBFRAME. It was this process that led to some concrete plans to clean Lehigh's MARC data to prepare for BIBFRAME. These environmental factors drive the issues and outcomes described in this presentation.

Cleaning MARC for a Better BIBFRAME Conversion

Developers who work on the Share-VDE ecosystem reviewed Lehigh's test sample of MARC records. Their questions led to two definitive decisions for how Lehigh University could clean MARC records for a more successful conversion to BIBRAME.

The first cleanup project the cataloging staff identified was to improve the consistency of format and material type representation in Lehigh's catalog. As is the case with many library catalogs, changes in cataloging practices are exemplified in MARC records as they came into the catalog over time. One large problem for MARC to BIBFRAME conversion occurs when a single MARC record represents multiple formats. BIBFRAME's Instance entity expresses, among other pieces of information, a single format of an item.[5] This legacy "single record, multiple format" problem in library catalogs is something the Library of Congress has dealt with and explains in the presentation, "Breaking News: Splitting MARC Records to Create Better BIBFRAME Data."[6] Staff have identified some of these "multiformat records" in Lehigh's catalog that can be reasonably split. Their first step was to separate the electronic and print journal records in the catalog. Staff will also need to surface and then consider other pockets of multiformat records that are part of the current catalog. If splitting these records negatively impacts the current user experience, then the library will leave the combined format, then document and consistently apply that local customization. A need for a more consistent representation of records representing single formats also became clear as Lehigh University Libraries upgraded and improved the catalog user interface. Cataloging staff scrutinized format facets as they appeared in the online catalog. The staff then applied more consistent practices which lead to clarity and accuracy in the format labels in Lehigh's online catalog. This is an ongoing cleanup project. This consistent application and understanding of cataloging practices benefits current users of the catalog and also will enable the staff to clearly communicate Lehigh's cataloging standards for purposes of a clean migration to BIBFRAME.

Authority control is another area of need in the library's MARC data. Lehigh Libraries' catalog has been without formal authority control for at least ten years. Without controlled headings, cleanly merging messy data with a union catalog like the Share-VDE catalog is not possible. Also, without authority control, the separation of records into BIBFRAME entities is less clean and reliable. As stated by Hahn, "Creating external authoritative links is essential to produce an accurate context for the Work entity description."[7] As with format cleanup, authority control of the catalog will benefit current users in addition to facilitating a migration to BIBFRAME. Because this is a project that will need to be outsourced, the metadata librarians will need to communicate the value of the expense to university administrators.

URI Proliferation

The Library of Congress has acknowledged that during its transition from MARC to BIBFRAME, the two would need to coexist for a period of time. Libraries are seeing the beginning of a MARC/BIBFRAME hybrid environment in MARC records with the increased presence of Uniform Resource Identifiers (URIs). Since December of 2023, OCLC has added URIs for persons, places and events, and WorldCat entities to their records.[8] Because the Lehigh library uses OCLC for the majority of its copy cataloging, within three months the catalog had over 10,000 records populated with URIs for these WorldCat entities. As a member of Share-VDE, Lehigh University can receive MARC records enriched with Share-VDE's URIs. Another source of URIs has been MarcEdit's linked data tool, often used at Lehigh to enhance records. All of these enrichments, if merged together, could lead to some lengthy MARC fields (see Figure 1). Another aspect of URIs that needs to be examined in Lehigh's records is URIs that reference a particular ontology. Whether the Library of Congress BIBFRAME ontology should live side by side with OCLC's and Share-VDE's is something the cataloging staff will be investigating and then deciding on best practices.

```
=700 1\
$aGroppo,B.$q(Bruno)$1https://id.oclc.org/world-
cat/entity/E39PBJhMgpMyjVCqkXB7wgY3wC$0http://
id.loc.gov/authorities/names/n88024842$1http://
viaf.org/viaf/92456331$0https://isni.org/
isni/0000000109249492$0https://www.wiki-
data.org/wiki/Q2926817$0https://svde.org/
agents/971654264078222
```

Figure 1. MARC 700 Add Entry—Personal Name with Six URIs

Discovery

The fact that Lehigh Libraries has full-time developers on staff and an open-source discovery service gives stakeholders extensive control and customization options with the catalog and discovery experience. Since the library has a direct hand in shaping discovery, questions as to what BIBFRAME means for Lehigh's end-user experience have naturally arisen. Is a major shift needed in the way we think about discovery at Lehigh? The current catalog-searching experience probably looks very familiar to anyone working with a university catalog: a list of results, with each result being a bibliographic record representing a resource. In contrast, the Share-VDE beta catalog showcases a different kind of discovery, enabled by the fundamentally different structure of a BIBFRAME record. Depending on the search, results are representative of entities, rather than a single, discernible resource. There are interesting and exciting possibilities enabled by linked data, but there are also questions about what discovery changes are useful, possible, and practical. Should the library try to maintain its current user experience with minimal changes, or should it take this opportunity to substantially adjust how its users interact with and discover resources? Unfortunately, at this point, most of the questions relating to discovery remain unanswered. Though linked data provides many opportunities to relate and describe resources, a significant amount

of user studies and testing would need to happen at Lehigh before any of the questions raised during this BIBFRAME exploration might be answered.

BIBFRAME: The Big Questions

Discovery is the next unexplored area, but larger questions involving the wider library community remain. Is BIBFRAME inevitable? How will BIBFRAME and MARC coexist, and for how long? Which linked data environment is being referenced? Lehigh is hardly the first institution confronted by these questions, but this process has forced Lehigh Libraries staff to consider them more closely. For FOLIO libraries, it seems inevitable that at some point each will have to at least consider how to handle BIBFRAME records. There has also been assurance of coexistence of MARC and BIBFRAME records, and of methods of translation and co-cataloging, which is indeed happening, but MARC has proven to be a resilient and long-lasting standard. It is understandable that many libraries with staff and budget limitations may be hesitant to invest the time and resources required to convert MARC to BIBFRAME. Perhaps the least-answerable at the moment is: whose data are we linking to? If a library works with one organization, like Share-VDE, to convert records to BIBFRAME and incorporate data into that linked data environment, and thus has Share-VDE URIs embedded, does that preclude linking to other sources of data like the Library of Congress, or OCLC? While mappings between linked data environments are often available, is a decision required for which "flavor" of BIBFRAME each library, institution, or consortium wants to employ? Unfortunately, Lehigh does not yet have answers, but is working toward a better understanding. Lehigh's BIBFRAME journey is still in its early stages, and there is an incredible amount to learn. Hopefully this presentation has highlighted some key areas of work, consideration, and questioning that any library should keep in mind when embarking on their own journey to BIBFRAME.

Contributor Notes

Lisa McColl is Metadata Services Manager, Lehigh University, Bethlehem, Pennsylvania, USA.

Greg Edwards is Metadata Librarian, Lehigh University, Bethlehem, Pennsylvania, USA.

Notes

1 "FTE & Headcount," Office of Institutional Data (OID), Lehigh University, accessed July 29, 2024, https://data.lehigh.edu/fte-headcount.
2 Myung-Ja K. Han, Tricia Lampron, and Greta Heng, "Considerations of Designing a User Friendly BIBFRAME Editor: Challenges and Future of Information Organization Professionals." (presentation, BIBFRAME Workshop in Europe 2023, KBR Royal Library of Belgium, Brussels, Belgium, September 20, 2023), accessed July 29, 2024, https://www.bfwe.eu/brussels_2023.
3 "The Recent Announcement from the Library of Congress Has EBSCO Information Services Looking for a New Word for Next-Gen," *EBSCOpost* (blog), EBSCO, October 4, 2022, accessed July 29, 2024, https://www.ebsco.com/blogs/ebscopost/library-congress-has-ebsco-information-services-looking-new-word-for-next-gen.
4 "Share-VDE: Linked Data for Libraries," Share-VDE, last modified May 28, 2024, accessed July 30, 2024, https://wiki.share-vde.org/wiki/Main_Page.
5 "Overview of the BIBFRAME 2.0 Model," BIBFRAME: Bibliographic Framework Initiative, Library of Congress, April 21, 2016, accessed July 29, 2024, https://www.loc.gov/bibframe/docs/bibframe2-model.html.
6 Kevin Ford and Jodi Williamschen, "Breaking News: Splitting MARC Records to Create Better BIBFRAME Data" (virtual presentation, 2023 LD4 Conference: Exploring Linked Data in GLAM: From Inspiration to Implementation, July 12, 2023), accessed July 27, 2024, https://www.youtube.com/watch?v=9i3y23XiNQE.
7 Jim Hahn, "Semi-Automated Methods for BIBFRAME Work Entity Description," *Cataloging & Classification Quarterly* 59, no. 8 (November 17, 2021): 853–67, https://doi.org/10.1080/01639374.2021.2014011.
8 Jeff Mixter, "OCLC's Preparation for BIBFRAME" (virtual presentation, BIBFRAME January 2024 Update Forum, Library of Congress, January 22, 2024), https://www.loc.gov/bibframe/news/source/2024jan22-bf-update-mixter.pdf.