IMPLEMENTING BIBFRAME IN A PILOT PROJECT AT THE YALE UNIVERSITY LIBRARY

Timothy Thompson
Librarian for Applied Metadata Research
Yale University Library
New Haven, Connecticut, USA
timothy.thompson@yale.edu

Daniel Mugaburu
Metadata Technical Assistant
Yale University Library
New Haven, Connecticut, USA
daniel.mugaburu@yale.edu

ABSTRACT

In the context of academic libraries in the United States, efforts to move from the MARC 21 formats to standards based on linked open data are currently underway. From 2018 to 2020, librarians at the Yale University Library took part in a larger initiative called Linked Data for Production. They carried out a pilot project to implement the Library of Congress BIBFRAME standard for original resource description. BIBFRAME is an extensible standard that can accommodate specialized description via domain-specific vocabularies such as the Art and Rare Materials (ARM) ontology extension.

At Yale, a team of 20 staff members was trained to use the Sinopia Linked Data Editor, a new cataloging tool being developed at Stanford University. A total of 200 items was cataloged before the project was cut short by the COVID-19 pandemic. During the relatively short period spent cataloging, it took nearly twice as long to catalog resources in Sinopia as it typically might in a traditional MARC cataloging environment. Academic libraries are still in the early stages of this transition, and additional investments in training and technology are needed for it to proceed successfully.

This paper can be outlined in three parts: the context of the project, the BIBFRAME model and the way in which it was used in our project, and finally, the Sinopia Editor, the tool we used to implement the BIBFRAME vocabulary and create new bibliographic descriptions.

Our project was carried out in the context of an initiative called Linked Data for Production (LD4P). The word "Production," in this regard, refers to the transformation of cataloging production, which is currently based on the MARC 21 formats, and the adoption of new practices based on the standards of the Semantic Web.

The LD4P initiative is a collective effort among research libraries in the United States. The leaders in planning its activities have been Stanford University, Cornell University,

Harvard University, and the University of Iowa, although Stanford has been primarily responsible for developing proposals and managing projects.

From the beginning, the projects of LD4P have been funded by grants from the Andrew W. Mellon Foundation. In its second phase, in which Yale was a participant, the initiative received a grant to foster the development of a community of practice among a greater number of libraries. In 2018, a call for proposals was published, and interested institutions were able to propose their own subprojects with a budget of up to US$50,000. In the end, proposals from a total of 17 U.S. libraries were accepted.

The main focus of the LD4P2 program was the development of a new cataloging tool, the Sinopia Editor, which is based on the principles of linked open data and was designed according to the standards of the Semantic Web. Each library in the group of 17 committed to using and testing the Sinopia Editor in its project and to developing workflows that would embody a new approach to library cataloging.

In our LD4P2 subproject, we took a very practical approach, with three main goals:

- First, we wanted to explore, in an iterative way, the implications of transitioning our cataloging practices and workflows and adapting to an approach based on linked open data. We also wanted to provide training to our cataloging staff and give them the opportunity, for the first time, to gain practical experience with BIBFRAME and other linked data standards.
- Second, we chose a set of materials for cataloging that would allow us to implement not only BIBFRAME, but other semantic vocabularies, especially the Art and Rare Materials Ontology (ARM), which is an emerging standard that had been developed during the first phase of LD4P.
- We worked primarily with a selection of pamphlets and ephemera from the 19th century and beginning of the 20th century. Some of these items have been moldering on the backlog shelves in our technical services department for some time. Because they were not high-priority items for cataloging, they were a good fit for our project. We could take as long as necessary to catalog them and, at some time in the future, convert our BIBFRAME data to MARC 21 records for display in our discovery environments. This conversion continues to be necessary as our current OPACs still depend on the MARC formats. As of the present date, this data conversion component is still a pending task.
- Finally, and precisely because we were not able to expose our work in the OPAC, we wanted to explore the possibilities of search, retrieval, and visualization that working with BIBFRAME might provide. To that end, we used the majority of our project budget to purchase a software license for a platform called metaphactory, which is a specialized search, visualization, and data management solution for linked data.

Our project was a collaboration between the technical services departments of the Yale Central Library and the Beinecke Rare Book & Manuscript Library. We divided the project into three teams, with a group of 20 people in the Metadata Creation Team, responsible for cataloging in the Sinopia Editor. Each person on the Metadata Team was expected to devote 20 percent of their time to cataloging in Sinopia, equivalent to one day per week.

Over the course of the project, we faced several challenges, and it was necessary to adjust our original timeline. Because LD4P2 was a complex project, involving more than 20 institutions in total, there were administrative delays in managing individual subgrants. Nevertheless, the greatest impact was a delay in delivering the Sinopia Editor. The software was originally expected to be released in April 2019, but version 1.0 was not ready for use until August.

We spent three months preparing to use the editor, designing templates and workflows. Then, we spent an additional three months, from November 2019 to January 2020, providing training for members of the Metadata Team. We finally began creating metadata in February 2020, but with the arrival of the COVID-19 pandemic, our time available for cataloging was cut drastically short.

Notwithstanding, we believe the project was a success. We were able to catalog a total of 200 items in the Sinopia Editor, which was enough to fulfill our original goals and provide our cataloging staff with substantial practical experience using BIBFRAME and linked data in a hands-on environment.

Overview of the BIBFRAME Model and its Use our project at Yale University.

BIBFRAME is a semantic data model and vocabulary developed by the Library of Congress in the United States. It is worth observing that BIBFRAME is not being developed by an international standards organization; rather, it belongs to the Library of Congress, which has its own institutional priorities, preferences, and even political concerns. It is also interesting to note that, while the MARC 21 formats continue to be modified year after year with new proposals for fields and subfields, the same process does not yet exist for BIBFRAME, which has been frozen since the release of version 2.0 in 2016.

When the creation of BIBFRAME was first announced, in 2011, it was emphasized that the project was intended to transition MARC 21 into a new Iinternet-friendly format. Indeed, MARC 21 has not adapted well to technological change over the years. Today, as we know, online search engines have largely replaced libraries and their catalogs and have come to serve as the primary method of information retrieval for most people. BIBFRAME, for its part, was designed using specifications intended to promote interoperability between bibliographic metadata and the World Wide Web.

When speaking of BIBFRAME, we also need to keep in mind its relationship to RDA: Resource Description and Access, which is our international cataloging code. RDA has a core set of four entities: Work, Expression, Manifestation, and Item. In BIBFRAME, the Work entity combines the characteristics of Work and Expression in RDA. Manifestation and Instance entities are basically equivalent, as is the Item entity.,

In our project, we chose to follow the RDA model for recording the attributes of Work and Expression entities. For technical reasons, we chose to adopt a new entity, which we called Superwork, to represent the RDA Work entity.

During the first phase of LD4P, various data modeling groups were formed to create ontology extensions based on BIBFRAME. These extensions were developed from the perspective of individual communities of practice, specialized in the description of specific

formats and materials. In total, four BIBFRAME ontology extensions were created: Art and Rare Materials, Cartography, Moving Images, and Performed Music.

One good thing about working with BIBFRAME is precisely that its provides the freedom to introduce elements from other semantic vocabularies into a description when it is necessary or convenient to do so. With our pilot project, for example, we focused on the implementation and development of the ARM ontology, which is a vocabulary designed to extend the scope of BIBFRAME for more detailed descriptions of Item entities

Using Sinopia Editor for Cataloging

Sinopia can be accessed in a hosted environment as a sandbox instance (https://stage.sinopia.io) or as a production instance (https://sinopia.io). Users must register and create an account to use the hosted version of the tool. The software is open source and also available for local installation (https://github.com/LD4P/sinopia_editor).

From the Sinopia homepage, click on "Linked Data Editor" to open the cataloging interface.

In Sinopia, each data entry form requires a corresponding template. The hosted platform includes default templates for different formats (for example, monographs, serials, or cartographic resources). There is a list of existing templates in Sinopia, and each template has a corresponding identifier. Users can develop their own templates, which is what we did at Yale for our pilot project. The primary templates used by Yale were designed for monographs, with separate forms for Superwork, Work, Instance, and Item.

The workflow that we developed began with creating a description of the Superwork entity and then continuing down the line until reaching the Item entity. Each description was linked back to its "preceding" entity along the way.

On the Superwork level, we record information about the creation or origin of the resource, such as primary contribution, and its intellectual content, such as associated subjects.

From the Superwork form, information can be entered for contributions to the creation of the entity. For example, we can record the primary contribution and specify that the contribution role is that of an author.

If a representation of the corresponding person entity exists in an external linked data source that has been configured as a lookup, the IRI (International Resource Identifier) for that entity can be inserted into the local resource description form.

If a person entity is not found in an external linked data source, the name can be entered as a string (using standard rules for formatting names) by clicking on the button for "Authorized Access Point."

A view of the Expression-level data entry form is as follows: in the first field on the form, we would add a hyperlink back to the corresponding Superwork entity after it has been created.

In conclusion, BIBFRAME represents a more flexible and extensible approach to cataloging. However, adapting to new tools, such as Sinopia, requires significant effort and an investment in training and technology. In the beginning, workflow efficiency is likely

to suffer until tools and workflows become more mature. For example, we observed that it took almost twice as long to catalog a resource in Sinopia than in MARC 21. To advance in the implementation of better, more useful metadata, everyone needs to work together, across institutions and across international borders.