The Evolution and Premise of the Library of Congress's BIBFRAME Model

Eryn N. Duffee

University of Southern Mississippi

Table of Contents

Abstract

The Library of Congress began an ambitious project in the 2000's, resulting in the launch of BIBFRAME 1.0 in 2011. BIBFRAME was tested and improved upon by a group of beta-users from various American universities and BIBFRAME 2.0 was published in 2016. BIBFRAME seeks to replace MARC and make bibliographic cataloging interconnected and searchable via the world wide web.

Keywords: BIBFRAME, Library of Congress, Resource Description Access, Uniform Resource Identifier, LOC, RDA, URI.

The Evolution and Premise of the Library of Congress's BIBFRAME Model

Introduction

For fifty years, MARC (machine readable cataloging) reigned supreme as the standard for library cataloging. When it first appeared with the advent of the "computational machine" it was revolutionary. Catalogs could be searched using keywords and a list result would be returned. MARC was created with three primary functions in mind: First, recording data related to the intellectual properties of a work. Second, recording data relating to the physical instance of a work. Third, recording metadata such as control numbers and ISBNs (Miller, 2012).

As computers evolved and with the invention of the world wide web, our lives and our information sources became more connected. Information was no longer contained by location and controlled vocabulary searches like the ones used in MARC were no longer adequate. Thus, in 2011, the Library of Congress launched BIBFRAME 1.0 (McCallum, 2017). After testing by beta groups, BIBFRAME 2.0 was launched with a new controlled vocabulary set (Taniguchi, 2017).

According to the Library of Congress FAQ page about BIBFRAME, there are three main goals for the initiative. First, differentiate between concept and manifestation. Second, clearly identify entities. The third, perhaps most important and ambitious, objective is to show and utilize relationships among entities to make the metadata more searchable and more widely available across the internet (Library of Congress, BIBFRAME FAQ, n.d.).

BIBFRAME uses RDA (Resource Description Access) as its theoretical foundation and is designed to integrate cataloging data using a variety of frameworks, such as RDA and AARC2, in order to create a connected source of metadata that can be integrated into the world wide web (Taniguchi, 2017). However, BIBFRAME is designed in a way that is meant to be independent of any one particular cataloging rule set (Library of Congress, BIBFRAME FAQ, n.d.).

History

Linked data became a goal for libraries in the 1990's as the internet began to proliferate. This led to the creation of several formats for linked data cataloging, eventually landing on RDA (Resource Description Access) and URIs (Universal Resource Identifiers) to create machine readable entry points that can be linked together (McCallum, 2017). In 2008, the decision to test and implement RDA was reached. RDA continues to be modified and built upon as it seeks to be a language that can effectively leverage catalog holdings within the environment of the world wide web.

In 2007, the Library of Congress created a community inquiry that resulted in the formation of a committee and a large report called the "On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control." One of the conclusions reached with the ability to have a large impact was that MARC should no longer be used, but replaced by a system that made library catalog information available and searchable on the world wide web (McCallum, 2017). This would mean that when a person searches using a mode such as Google, library catalog holdings would show up in the results list along with web entities.

In 2015, an important revision to RDA was written that included changes to how data was linked and what entities and attributes were to be included in a record. The first of these changes set out the dual objectives that subject relationships should be recorded in a way that all information on a subject can be found in a search and secondly, “the data should reflect all significant subject relationships,” (RDA 23.2, 2015).

To quote El-Sherbini’s 2018 article, “Another important section in Chapter 23 deals with the methods of presenting subject relationships by either an identifier, an authorized access point, or a description that indicates what the work is about. This allows the cataloger to link to outside subject thesauri or even classification systems, or any other identifier.”

RDA has had a significant, positive effect on the cataloging of foreign language entities. The framework sets three ways that agencies can record names, including the recording of names as variant names. This structure allows for non-Roman script to be contained within the authority record, thus making foreign language names and titles searchable (El-Sherbini, 2018).

It must be observed that the Library of Congress is only one group working on the BIBFRAME model. This is an international undertaking with global implications. The LOC’s own paper on the subject discusses the many contributions by other organizations such as the British Library and Deutsche Nationalbibliothek, as well as WorldCAT and Schema.org (Miller, 2012).

Schema.org is a collaborative entity created by Google, Bing, and other web crawlers designed to allow website creators to express searchable metadata. While the Schema.org objective is designed to meet the needs of its commercial creators and not

exactly in-line with the library's objectives, the ability to leverage this new tool has wide implications within the library world (Miller, 2012).

In 2016, the Library of Congress launched BIBFRAME 2.0. The LOC first mapped the current MARC catalog to BIBFRAME ontology and created a software program to convert the data with minimal information loss. The second part of this project was to create an input description tool. This description tool does not use the numbers and codes as the MARC system does, but instead it uses labels for the elements using the RDA cataloging rules (McCallum, 2017).

BIBFRAME is meant to be a starting point, not a static end-point, in the evolution of library cataloging. The Library of Congress recognizes this and encourages collaboration, adaptation, and evolution of the framework to meet the needs of libraries for many years to come (Miller, 2012).

Objective

The importance of linked data and the goals that the library community hopes to achieve are succinctly stated by Nancy Fallgren's 2014 quote where she says:

> "Although library metadata made the transition from card catalogs to online catalogs over 40 years ago, and although a primary source of information in today's world is the Web, metadata in our OPACs are no more free to interact on the Web today than when they were confined on 3" by 5" catalog cards in wooden drawers. What if we could set free the bound elements? That is, what if we could let serial titles, subjects, creators, dates, places, and other

> elements, interact independently with data on the Web to which they are related? What might be the possibilities of statement-based, Linked Data environment?" (Fallergren, 2014).

The main goal of BIBFRAME is to create a framework that is useful in today's digital world. Creators seek to replace MARC21 with BIBFRAME to this end. Explaining the difference in information arrangement between the two systems can illuminate the need for this change.

In MARC21 format, the goal is to have all of the bibliographic information about a work collected in one record. BIBFRAME seeks to deconstruct that record and create linked data that shows relationships between bibliographic information and then points to an authority file. For instance, in the case of a book's title, the many different iterations of the title will all be searchable and then link to an authority file on the work. The relationship between the data and the authority file is presented as an RDA Triple-subject/property/object (El-Sherbini, 2018). RDA triples will be defined later in the paper. BIBFRAME also seeks to be an inclusive format that "[accommodates] different content models and cataloging rules, exploring new methods of data entry, and evaluating current exchange protocols." (LOC FAQ, n.d.).

Framework

The BIBFRAME model shifts cataloging into a framework and methodology that focuses on linked data. This change is designed to make bibliographic data up to date with modern technology by making library holdings searchable and usable within the internet.

The RDA registry model has the entities work, expression, manifestation, and item as well as person, family, and corporate body.

Within the RDA model, the resource instance is an HTTP attached to a RDA entity. For instance, "http://rdaregistry.info/Elements/c/C10001", It then has an English language lexical alias of "http://rdaregistry.info/Elements/c/Work.en" (Taniguchi, 2017). This model focuses on the connection between entities but still points towards the RDA entity for authority (Taniguchi, 2017).

In the BIBFRAME model, the URI (uniform resource identifier) is parallel to the RDA resource instance. Within BIBFRAME, the primary entities are work, instance, and item. Each of these items to subcategories of contribution, agent, and provision activity. Work is also connected to event. Work, instance, and item then all point to the rdfs:resource. This is demonstrated below in the graph created by Taniguchi in 2017.

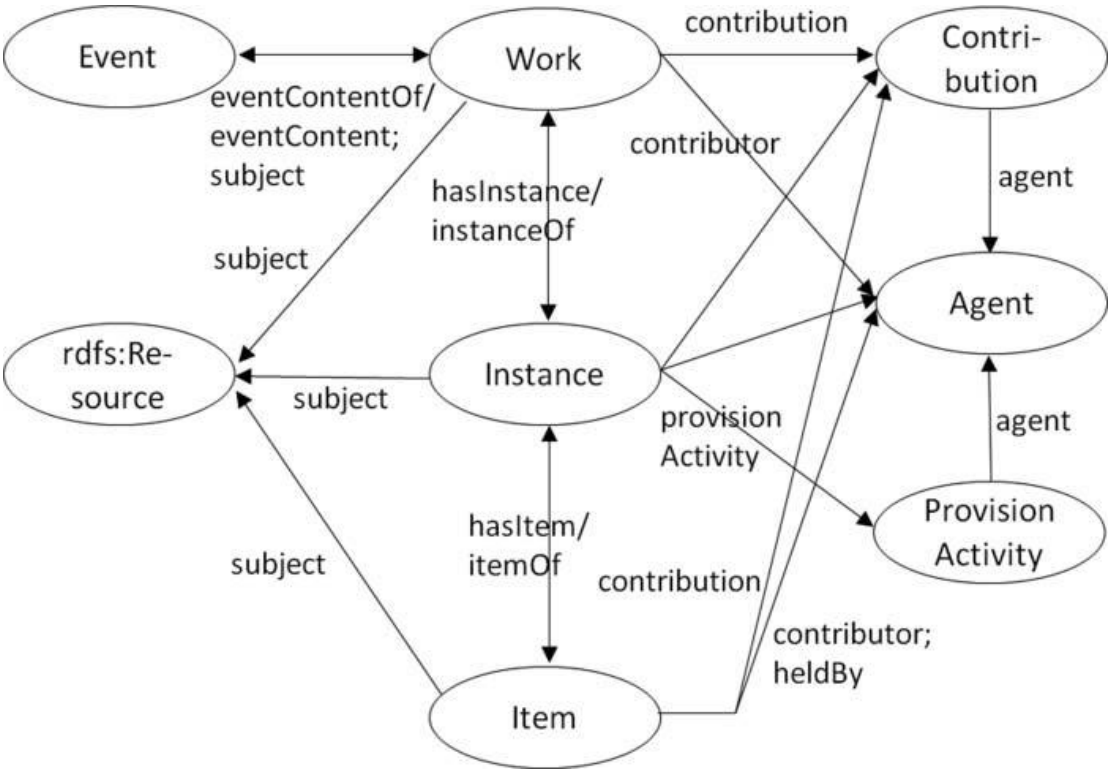

Figure 1. BIBFRAME model: Core Classes and properties relating to the classes (Taniguchi, 2017).

Within the BIBFRAME 2.0 model, the class work looks like "http://id.loc.gov/ontologies/bibframe/Work and the compact URI is bf:work with bf standing in for BIBFRAME (Taniguchi, 2017).

This use of the HTTP model puts catalog data online and makes it searchable by web crawlers, thus enabling the catalog information to be presented to end-users as part of the search results. An example of what this would look like would be as follows: a person searches Google using the term "George Washington" and in addition to returning web results, the results list would include http://id.loc.gov… results showing library catalog holdings about, authored by, or otherwise related to George Washington.

Entities

In order to understand BIBFRAME, one must understand how each of the major entities within the framework is defined. For this, we will go to the Library of Congress webpage dedicated to giving an overview of BIBFRAME.

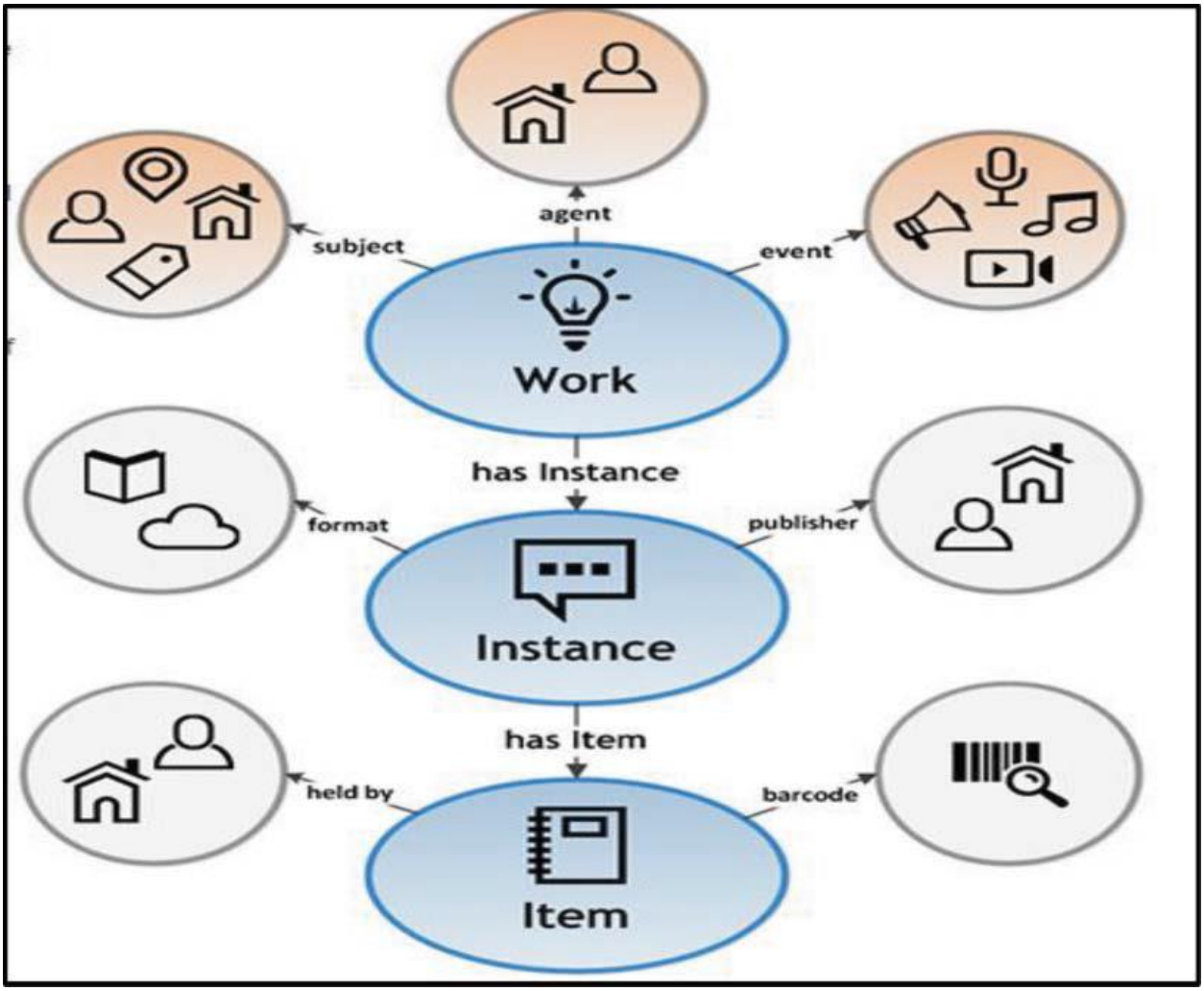

Figure 2. Overview of the BIBFRAME 2.0 Model (LOC, BIBFRAME Overview, n.d.).

Work, instance, and item are the three primary levels within BIBFRAME. Work is an instance of subject, agent, or event, as can be seen in the figure above. A work “reflects the conceptual essence of a cataloged resource,” (LOC, BIBFRAME Overview, n.d.). Instance gives information about the various formats and publishers that a work may have. The item refers to a single entity that may be held by a library and gives information such as the barcode, ISBN number, or who holds the item (LOC, BIBFRAME Overview, n.d.).

The Library of Congress also defines subjects, agents, and events. Subjects define the concept contained within a work. Agents are any people involved with a work, from the author, illustrator, composer, publisher, and others. Events are the occurrences of a work (LOC, BIBFRAME Overview, n.d.).

Conclusion

BIBFRAME will continue to evolve as it is tested and refined by various universities and coalitions. The effort is a great example of collaboration between a government entity- the Library of Congress, universities, international libraries, American public libraries, corporations, and the w3c coalition, who sets the standards for the world wide web. Libraries, publishers, and other entities are slowly making the transition from MARC to BIBFRAME and as more groups adopt this new format, new areas for improvement and advancement will be discovered and built. This is a moment of great transition within the library world, and as such, there will be triumphs and frustrations but, in the end,, the results of having a globally connected catalog system that is searchable using the world wide web will be worth the work and effort put into the project.

References

McCallum, S. H. (2017). BIBFRAME Development. JLIS.It, Italian Journal of Library, Archives & Information Science, 8(3), 71–85. https://doi-org.lynx.lib.usm.edu/10.4403/jlis.it-12415

Tharani, K. (2015). Linked Data in Libraries: A Case Study of Harvesting and Sharing Bibliographic Metadata with BIBFRAME. Information Technology & Libraries, 34(1), 5–19. https://doi-org.lynx.lib.usm.edu/10.6017/ital.v34i1.5664

El-Sherbini, M. (2018). RDA implementation and the emergence of BIBFRAME. JLIS.It, Italian Journal of Library, Archives & Information Science, 9(1), 66–82. https://doi-org.lynx.lib.usm.edu/10.4403/jlis.it-12443

Xu, A., Hess, K., & Akerman, L. (2018). From MARC to BIBFRAME 2.0: Crosswalks. Cataloging & Classification Quarterly, 56(2/3), 224–250. https://doi-org.lynx.lib.usm.edu/10.1080/01639374.2017.1388326

Taniguchi, S. (2018). Is BIBFRAME 2.0 a Suitable Schema for Exchanging and Sharing Diverse Descriptive Metadata about Bibliographic Resources? Cataloging & Classification Quarterly, 56(1), 40–61. https://doi-org.lynx.lib.usm.edu/10.1080/01639374.2017.1382643

Fallgren, N. et al., (2014). “The Missing Link: The Evolving Current State Of Linked Data For Serials”. Serials Librarian 66, No. 1–4 (2014): 123–38, http://dx.doi.org/10.1080/0361526X.2014.879690.

Miller, E., et al. (2012). Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services. Library of Congress. Washington, DC.

No author, n.d. Overview of the BIBFRAME 2.0 Model. Library of Congress. Retrieved from http://www.loc.gov/bibframe/docs/bibframe2-model.html.

No author, n.d. BIBFRAME Frequently Asked Questions. Library of Congress. Retrieved from http://www.loc.gov/bibframe/faqs/.

Tables

Table 1

[Table Title]

Column Head	Column Head	Column Head	Column Head	Column Head
Row Head	123	123	123	123
Row Head	456	456	456	456
Row Head	789	789	789	789
Row Head	123	123	123	123
Row Head	456	456	456	456
Row Head	789	789	789	789

Note: [Place all tables for your paper in a tables section, following references (and, if applicable, footnotes). Start a new page for each table, include a table number and table title for each, as shown on this page. All explanatory text appears in a table note that follows the table, such as this one. Use the Table/Figure style to get the spacing between table and note. Tables in APA format can use single or 1.5 line spacing. Include a heading for every row and column, even if the content seems obvious. To insert a table, on the Insert tab, tap Table. New tables that you create in this document use APA format by default.]

Figures

Figure 1. [Include all figures in their own section, following references (and footnotes and tables, if applicable). Include a numbered caption for each figure. Use the Table/Figure style for easy spacing between figure and caption.]

For more information about all elements of APA formatting, please consult the *APA Style Manual, 6th Edition*.