

Assessing BIBFRAME 2.0: Exploratory Implementation in Metadata Maker

Brinna Michael
Emory University,
United States
bamicha@emory.edu

Myung-Ja K. Han
University of Illinois,
United States
mhan3@illinois.edu

Abstract

As interest in linked data grows throughout the cultural heritage community, it is necessary to critically assess current tools for conversion and creation of linked data “records” and to explore new avenues for creating and encoding data using existing frameworks. This paper discusses the BIBFRAME 2.0 model and Library of Congress conversion specifications from MARC21 through the process of designing and implementing an adapted, minimal-level conversion framework into the cataloging web application, Metadata Maker. In the process of assessment, we identified and addressed local solutions for three key structural issues resulting from the Library of Congress conversion specifications: duplicated data, pervasiveness of blank nodes in RDF/XML, and prevalence of literal data values over URIs stemmed from the current MARC records environment. Additionally, we address concerns with how the BIBFRAME 2.0 model currently conceptualizes *Work* and linked data as a static “record.”

Keywords: MARC21; linked data; BIBFRAME 2.0; linked data conversion; web applications

1. Introduction

In 2008, the Library of Congress (LC) recognized the need to begin developing a new cataloging framework to replace MARC 21 as technology continued to advance and data management, exchange, and curation techniques changed (LC, 2008). Investigation resulted in the announcement of the Bibliographic Framework Initiative, or BIBFRAME, in 2012 which sought to “translate the MARC 21 format to a Linked Data... model while retaining as much as possible the robust and beneficial aspects of the historical format.” (LC, 2012) While linked data and semantic web concepts continue to be of great interest to both the general public and information professionals as a means of more efficiently and effectively disseminating, linking, and accessing information across a variety of sources, however, the creation of catalog data as linked data has not been discussed widely.

Currently, libraries mainly rely on LC’s BIBFRAME editor (LC, Bibframe Editor) for testing creation of linked data or using conversion tool (LC, 2017) for MARC21 records to BIBFRAME as it is freely available and easy to use. To explore an alternative option to the LC tools, we decided to expand the functionality of a metadata creation tool previously produced by the University of Illinois at Urbana-Champaign (U of I) Library, Metadata Maker (U of I, 2017). The goal was to design a BIBFRAME export feature to complement Metadata Maker’s current MARC21, MARCXML, MODS, and Schema.org (encoded as HTML) export features. This paper shares detailed procedures of how we reviewed LC’s conversion tool by analyzing BIBFRAME outputs for monographs. The paper also discusses the challenges of navigating current conversion specifications (LC, 2019^a) to develop a concise RDF/XML template using a different application of the BIBFRAME vocabularies for Metadata Maker

2. Methods

To prepare for creating an export function for BIBFRAME 2.0 RDF/XML in Metadata Maker, we selected two example monographs, one literature and one non-literature, and created minimum-level bibliographic records for each using Metadata Maker as literature monographs require a specific set of information. The Metadata Maker for monographs module was created based on LC's Minimal Level Record Examples (2003) and considered as the most straightforward in its structure among available modules. By creating these records with Metadata Maker, we could ensure their focus remained on those MARC fields which were relevant to the tool and the type of resource it describes, thereby naturally limiting the selection of BIBFRAME 2.0 classes and properties being assessed. Once these records were exported in MARCXML format, we used the LC MARC to BIBFRAME conversion tool (2017) to create an initial set of BIBFRAME 2.0 RDF/XML outputs.

Using the BIBFRAME 2.0 outputs, we assessed the resulting linked data output: first by extracting a list of all triples encoded within the RDF/XML, and then by identifying the value types of individual nodes and visualizing those relationships using a network visualization. Through this process, nodes were categorized based on the data type of the value they held: Literal, URI, or Blank. From the results, we identified several core issues, including an abundance of blank nodes and duplicated data. Blank nodes here refer to elements within the RDF/XML which, when serialized within an RDF graph, contains no literal or URI value. This prompted us to create a modified, or "condensed," version of the BIBFRAME 2.0 outputs as shown in table 1 with the intent of producing a concise template for use in designing a BIBFRAME 2.0 export feature into Metadata Maker.

TABLE 1: BIBFRAME 2.0 Output Comparison

	XML Segments	Classes	Blank Nodes	URIs	Literals
LC Conversion Tool	269	125	33	62	30
Condensed BIBFRAME	125	47	5	34	8

3. Linked Data Structure

3.1. Duplicated Data

The first prominent issue we identified when assessing BIBFRAME outputs from the LC conversion tool was the prevalence of duplicated data. MARC21 requires the duplication of information in several different fields for the purposes of human- and machine-readability, meaning that information such as publication and copyright date are recorded in three to four different places. With the current MARC to BIBFRAME conversion specifications laid out by LC, each of these instances are being mapped to duplicated BIBFRAME elements, resulting in repeated BIBFRAME classes, mostly containing near-identical literal values.

To remedy this, we identified values which were duplicated and pared the occurrences in the resulting BIBFRAME output down to one (or two if necessary). For example, instead of including two separate instances of the <bf:copyrightDate> property, one to hold the value "©2013" and another with the Extended Date Time Format standard "2013," we decided to include only the second instance, with the reasoning that the copyright symbol may be added as necessary when rendering the metadata in a user interface.

3.2. Blank Nodes

The second issue we identified was the frequency of blank nodes. In its current form, BIBFRAME 2.0 RDF/XML output is densely hierarchical. While this is logical within the RDF/XML syntax, when translated to an RDF graph, the result is complicated and shows an

unnecessarily expanded system of relationships which add a number of unnecessary placeholders between related values, as shown in Fig. 1.

In contrast, the adjustments made in Fig. 2 remove duplication of information and integrate the information previously stored in literals into URIs which serve as the core of linked data. We recognized that the Fig. 2 reduces a number of triples that would work as a place holder for human readable labels. However, like copyright symbols, true linked data will fetch the preferred labels or the source information from the URIs used for the value, such as in the case of `<bf:code>` in Fig. 1, which can be derived from the URI itself.

```

<bf:subject>
  <bf:Topic rdf:about="http://example.org/d0e1#Topic650-14">
    <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#Topic"/>
    <rdfs:label>XML (Document markup language)</rdfs:label>
    <madsrdf:authoritativeLabel>XML (Document markup language)</madsrdf:authoritativeLabel>
    <bf:source>
      <bf:Source>
        <bf:code>fast</bf:code>
      </bf:Source>
    </bf:source>
    <bf:identifiedBy>
      <bf:Identifier>
        <rdf:value>fst01181899</rdf:value>
        <bf:source>
          <bf:Source>
            <rdfs:label>OCoLC</rdfs:label>
          </bf:Source>
        </bf:source>
      </bf:Identifier>
    </bf:identifiedBy>
  </bf:Topic>
</bf:subject>

```

FIG. 1. LC Conversion Tool `<bf:Topic>` Structure

```

<bf:subject>
  <bf:Topic rdf:about="http://id.worldcat.org/fast/1181899">
    <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#Topic"/>
    <bf:source>
      <bf:Source rdf:about="http://id.loc.gov/vocabulary/identifiers/fast"/>
    </bf:source>
  </bf:Topic>
</bf:subject>

```

FIG. 2. Condensed BIBFRAME `<bf:Topic>` Structure

3.3. Data Values: Reconciliation Work

The final issue we identified was the lack of the reconciliation processes to replace literal values with URIs. This arises from the functionality of the LC conversion tool itself, which, in its current form, pulls values directly from the provided MARC record, meaning that values are almost exclusively literals. While the Program for Cooperative Cataloging (PCC)’s Task Group on URIs in MARC has been working to identify MARC fields that can contain URIs as a value and how those URIs can be added, most libraries do not add URIs as values for the MARC fields (LC, 2019^b). We understand that literal string values are necessary for display and human readability of linked data, however, lack of URIs ultimately undermines the purpose of encoding metadata within a linked data model: to link related concepts and provide broad, dynamic context for those concepts.

As evidenced in fig. 1 and 2, a number of literal values encoded into BIBFRAME can be translated into URIs through a reconciliation service as a part of the conversion process. For example, in fig. 1, the value held in `<rdf:value>` is mapped from 650 \$0 and represents the identifier applied to the term “XML (Document markup language)” in the FAST vocabulary (OCLC). As a

literal, this value serves little purpose on its own for linking with other resources, but through either string matching or integration of a reconciliation service, these literal values can be transformed into URIs.

4. Conceptual Model

In addition to structural concerns, we identified a number of conceptual gaps in the implementation of BIBFRAME 2.0, raising questions around the role of administrative metadata in describing the “record” itself, transformed from the MARC fixed fields, what purpose such reflexive description serves, what a “record” looks like in a linked data environment, and how the concept of a “*work*” is translated into linked data.

4.1. Administrative Metadata

In considering the BIBFRAME 2.0 outputs produced by the LC conversion tool, we debated the placement of <bf:AdminMetadata> within the container of <bf:Work>. According to the BIBFRAME 2.0 documentation (2016), <bf:AdminMetadata> holds “metadata about the metadata, especially provenance information.” If this definition is accurate, what role does it play in relation to <bf:Work>? The values contained in the <bf:AdminMetadata> class and related subclasses, such as encoding level or standard used to create of descriptive metadata, do not inform the end user, either a patron or cataloger, about any conceptual aspects of a Work, rather, they describe the contents of the BIBFRAME document as a whole.

Within the MARC environment, it is logical to encode administrative metadata within the record itself since each record is self-contained and self-referential. In a linked data environment, however, this is not the case. There is no clearly bounded “record” in linked data. Instead, there is a wider network of links and strings which can be extracted to provide a snapshot view of the “state” of a specific metadata description set from the linked data network. While the initial conversion of MARC encoded administrative metadata is logical and necessary, in the case of BIBFRAME to MACR conversion, much of this information is not useful in ‘administrative’ purpose. Therefore, it would be more useful to find a way to automatically create and update <bf:AdminMetadata> data on a case by case basis when a “record” is generated, describing a single snapshot of one section of the wider linked data network.

4.2. BIBFRAME’s *Work* and *Instance*

Analysis of the LC conversation tool BIBFRAME RDF/XML outputs offered an excellent opportunity to consider the conceptualization of *Work* and *Instance* in the BIBFRAME 2.0 model in light of the coming changes to RDA. The BIBFRAME 2.0 model defines a *Work* as “reflecting a conceptual essence of a catalogued resource” (LC, 2016) and is intended to combine the FRBR concepts of *Work* and *Expression* (McCallum, pp. 79). This conceptual shift predisposes the BIBFRAME model to clash with the implementation of IFLA-LRM (2017) conceptualizations in RDA which maintains these delineations.

This conceptual dissonance raises a number of questions about the practical implementation of BIBFRAME in the wake of the updated RDA standards. How will the structure of BIBFRAME integrate RDA recommendations which are built upon the IFLA-LRM distinction between *Expression* and *Manifestation*? Does the way BIBFRAME defines its *Work* and *Instance* entities clarify the conceptual divide between the previously used concepts of *Expression* and *Manifestation*? Or are there former MARC fields that do not translate directly from a FRBR/IFLA-LRM model to the BIBFRAME model? Most importantly, does the current BIBFRAME model improve the practice of cataloging, and through this, our end users?

5. Metadata Maker

Following the assessment of LC’s recommendations for MARC to BIBFRAME conversion and the development of a modified, condensed version of the resulting RDF/XML, we prepared a

template RDF/XML document for use in designing a new export feature in Metadata Maker. While not adhering to the conversion specifications outlined by LC (2019), the template RDF/XML, and by default all outputs exported using that template, still validate against the BIBFRAME 2.0 namespace. A programmer was hired to create the scripts necessary for implementation, and using the provided template, produced JavaScript for the existing Monograph module. Moving forward, the RDF/XML template and JavaScript export file will be reassessed and modified for implementation with other Metadata Maker modules within the application.

6. Discussion and Conclusion

In assessing the BIBFRAME 2.0 model, it is clear that while the foundations of the model are relatively sound, there are still some pressing challenges that need to be addressed before the model can be fully implemented as a linked data successor to MARC. For the purposes of this project, we maintained a focus on succinctness, keeping in line with the purpose of Metadata Maker as a tool for generating minimum-level bibliographic descriptions. As a result, it quickly became clear that concern with losing information in the transition from MARC to BIBFRAME is misplaced, resulting in “over-mapping” to the linked data framework when focus should be on identifying where information overlaps in the MARC record and consolidating it in BIBFRAME.

BIBFRAME 2.0 is still a work in progress, the model and the conversion tools built around that model are in need of several adjustments which can bring them, at least structurally, into closer alignment with linked data principles and practices. First, rather than including all properties, the BIBFRAME output could be simplified by incorporating properties which can hold values in MARC format record. This will limit the reliance on blank nodes as bridges between concepts which could be more succinctly linked in a single triple, as shown in the adjustments made between Fig. 1 and Fig. 2. Those labels and URIs can be added whenever those values are available to be generated via reconciliation services. Second, the current conversion specifications need to be updated to include methods of assessing and merging duplicate information into a more succinct and linked data complicit BIBFRAME outputs. Third, efforts should be made to integrate some sort of reconciliation service into the current LC conversion tool to generate URI values from literal values used in the MARC fields. For example, data fields like 336, 337, and 338 can be easily replaced as URIs, and reconciliation services for data fields 1XX, 7XX, and 65X can be integrated in addition to mapping the literal values.

While structural fixes are relatively simple to address, conceptual ones must be engaged with sooner rather than later. First, the BIBFRAME 2.0 linked data environment must be more consistently conceptualized. Only then can the resulting network be used to efficiently support user interfaces and cataloging software, and in doing so, begin to address the issue of when and where the <bf:AdminMetadata> class and subclasses should be used. Additionally, means of integrating self-descriptive metadata into the classes themselves should be explored in order to allow for those aspects of bibliographic description to no longer be tethered to a static “record.” Additionally, it is critical that the line between *Work* and *Instance* is concretely defined in a way that support interoperability with FRBR/IFLA-LRM concepts while improving the practice of cataloging as a whole.

References

- Library of Congress. Bibframe Editor Workspace. Retrieved from [http:// http://bibframe.org/bfe/index.html](http://http://bibframe.org/bfe/index.html).
- Library of Congress. (2003, October). MARC 21 Bibliographic. “Appendix C – Minimal Level Record Examples.” Retrieved from <http://www.loc.gov/marc/bibliographic/bdapndxc.html#book>.
- Library of Congress. (2008, January). On the record: Report of the Library of Congress working group on the future of bibliographic control. Retrieved from <https://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>.

- Library of Congress. (2012, May). The Library of Congress announces modeling initiative. Retrieved from <https://www.loc.gov/bibframe/news/bibframe-052212.html>.
- Library of Congress. (2016). BIBFRAME Ontology, version 2.0.1: BIBFRAME 2 Vocabulary. Retrieved from <http://id.loc.gov/ontologies/bibframe-2-0-1/>.
- Library of Congress. (2017). marc2bibframe2. Retrieved from <https://github.com/lcnetdev/marc2bibframe2/>.
- Library of Congress. (2019)^a. MARC 21 to BIBFRAME 2.0 Conversion Specifications. Retrieved from <http://www.loc.gov/bibframe/mtbf/>.
- Library of Congress. (2019)^b. PCC Task Group on URIs in MARC. Retrieved from <https://www.loc.gov/aba/pcc/bibframe/TaskGroups/URI-TaskGroup.html>.
- McCallum, Sally H. (2017, September). BIBFRAME Development. *JLIS.it*, 8(3). University of Florence, SAGAS Department. doi: 10.4403/jlis.it-12415.
- OCLC. OCLC Research. "FAST (Faceted Application of Subject Terminology)." Retrieved from <https://www.oclc.org/research/themes/data-science/fast.html>.
- Riva, Pat, Patrick Le Bœuf, Maja Žumer. (2017, December). IFLA Library Reference Model: A conceptual model for bibliographic information. International Federation of Library Associations and Institutions.
- University of Illinois at Urbana-Champaign Library. (2017, February). Metadata Maker. Retrieved from <http://quest.library.illinois.edu/marcmaker/>.