

The Latest Developments on BIBFRAME Initiatives

Elias Royal

School of Information, University of South Florida

LIS 6711: Organization of Knowledge

Professor Jinfang Niu

June 22, 2022

In library communities, the benefits of linked data are abundantly clear by how libraries are currently sequestered from a broader information community on the Web. While the implementation of the Bibliographic Framework (BIBFRAME) initiative is a “disruptive innovation” (Schreur & Carlson, 2020), it will positively alter the processes for controlling bibliographic and authority data (Misu Kim et al., 2021). While some librarians are skeptical about the practical utility of linked data on libraries (Jones, 2021), this skepticism may emerge from certain characteristics of the BIBFRAME implementation. For example, the BIBFRAME implementation is a large-scale and an ongoing project with high level of complexity (Bigelow & Sparling, 2022). Because a small collection of national and research libraries from around the world are developing and implementing semantic or linked data technology, librarians may be justified in their skepticism about these eleven yearlong projects (Bigelow & Sparling, 2022). As a result, this report will present the latest developments on the BIBFRAME implementation as it pertains to bibliographic and authority controls.

A central variable converting legacy data into linked data is the enrichment of Uniform Resource Identifiers (URIs) within a linked data environment (Samples & Bigelow, 2020). The existing web is document-centric with links directing users to other webpages, but the semantic web is data-centric with data being linked automatically to URIs with the same identifier (Misu Kim et al., 2021). According to Casalini (2022), a URI is the ideal code or identifier because it circumvents natural language patterns, while also adhering to the data structure of the semantic web model, Resource Description Framework or RDF (Casalini, 2022). As a linked data structure that facilitates access and cultivates web data, RDF data consists of triples (and sometimes quads) of data that consist of a subject, predicate, and object (Jones, 2021). According to Saha and Mandal (2021), linked data technology assists with the standardization and

intersection of data formats, which ensure a higher level of interoperability and interchange (p. 21). Because of the challenging undertaking of converting all legacy data in libraries, the successful implementation of the Bibliographic Framework (BIBFRAME) will depend on all conversion efforts.

The integration of standards or ontologies accelerates conversion efforts. During the 2017-2020 period, national libraries and research libraries have made significant progress on implementing BIBFRAME initiatives. This progress is assisted by significant modifications to the resource description and access (RDA) encoding standard that originates almost concurrently with BIBFRAME. While RDA is not compatible with MARC (MACHine Readable Cataloging), a new version of RDA has been released in December 2020 (Bigelow & Sparling, 2022) to improve compatibility with MARC and anticipate BIBFRAME implementation. The current BIBFRAME developments are driven by iterative endeavors, which implement linked data by building on the legacy data of MARC records. Most of the implementation efforts that have been deployed by community-based initiatives would have been impossible without cross-functional collaboration.

A benefit of international collaboration (Bigelow & Sparling, 2022) is the development of BIBFRAME hubs, which reconciled the various incompatibilities of MARC title authority records. While converting MARC title authorities to BIBFRAME bibliographic work description (BWD), the Library of Congress (LC) designates these BWD as “hubs” (McCallum, 2022). With their goal of converting 1.5 million MARC title authorities, LC utilizes work/expression classes in the functional requirements for bibliographic records (FRBR) and the RDA models to reconcile the various sets of cataloging guidelines in MARC21 records (McCallum, 2022, p. 48). According to McCallum (2022), this conversion process is exceedingly difficult, but subject

headings also present various difficulties because subject headings are part of the work description and not an instance property as it is defined by the FRBR and RDA models (p. 49).

More recently, other community-based initiatives have used the LC's concept of BIBFRAME hubs to apply it to various tools and processes. Firstly, BIBFRAME hubs reconcile the incompatibles of MARC21 cataloging rules by converting distinct authority data into bibliographic work descriptions (McCallum, 2022). As it has been emphasized by McCallum, BIBFRAME hubs facilitated the preservation of collocation, which clusters title descriptions by author names and titles (p. 46). Because of the successes of LC's pilot programs, the Share Virtual Discovery Environmental (Share-VDE) initiative adopts the concept of hubs and creates opuses as another form of bibliographic work descriptions (Casalini, 2022). Share-VDE uses an opus to generate an authoritative knowledge base (Sapientia) with the following entity model classes: Opus, Work, Instance, and Item (Casalini, 2022). Casalini (2022) also mentions the efforts of other research libraries in the Linked Data for Production (LD4P) cohorts to employ opuses in their linked data initiatives (p. 51). Within the context of linked data, authority control resources strive to establish standards and controls of bibliographic data (Casalini, 2022).

Collaboration in the library and developer community has a significant impact on the implementation of BIBFRAME. During the experimentation phase, the early adopters of BIBFRAME were decentralized and uncoordinated in their implementation efforts, but these early adopters now appreciate the value of coordinated collaboration. For example, OCLC collaborates with PCC, LC, and LD4P2 to create an "Entity Management Infrastructure" to support linked data initiative in the library community (Misu Kim et al, 2021). According to Misu Kim et al. (2021), these forms of collaborations promote data exchange between LD4P2 and WorldCat (p. 139). With a focus on staff training, improving conversion processes,

application profiles, and data modelling, the LD4P2 Cohort, Share-VDE, and the Program for Cooperative Cataloging (PCC) cultivate collaboration efforts to establish effective workflows and skills (Bigelow & Sparling, 2022). These collaborative efforts are instrumental in pursuing the daunting task of converting legacy data to linked data.

Through the efforts of the Share-VDE, BIBFRAME is tested at scale, which facilitates data conversion for authority controls (Samples & Bigelow, 2020). Regarding data conversion, Casalini (2022) states that Share-VDE converts MARC data to RDF with BIBFRAME vocabulary and other relevant ontologies (p. 109). As a community-driven initiative, which is comprised of an international assembly of national and research libraries, Share-VDE applies linked data to descriptive metadata (Samples & Bigelow, 2020). According to Samples and Bigelow (2020), this is an important development because the scale of the BIBFRAME implementation allows a vast number of major research libraries to analyze processes and metadata. It is important to note that Share-VDE builds on the actions of the LD4P projects. The scale of the collaboration between research libraries throughout the world accelerates the implementation of a linked data environment with descriptive metadata. As evident by the research, this type of international collaboration is novel within the eleven-year history of the BIBFRAME initiative because of the decentralization of BIBFRAME experimentation and implementation.

In addition to the data conversion that is viable with international collaboration of major research libraries, the Share-VDE also supports the enrichment and reconciliation of MARC data to linked data (Samples & Bigelow, 2020). According to Casalini (2022), enrichment of MARC records occurs with identifiers from external sources. Furthermore, reconciliation and collection of entities occurs with the issuing of the Cluster Knowledge Base (CKB) that aims to develop a

data tool (Casalini, 2022). As Casalini (2022) states, the CKB or Sapiientia is envisioned as an authoritative knowledge base with tools to improve authority control with algorithms and an editor (J. Cricket) to manage the collaboration of individual users and data producers (i.e.: librarians) who improve the quality of the authority data (p. 112). In addition to tools for authority controls, Share-VDE implements Wikidata to search for sources and enrich Share-VDE data (Casalini, 2022). Furthermore, Share-VDE utilizes Application Programming Interfaces (APIs) to enable scalability, functionality, reusability, connectivity (Casalini, 2022).

In recent years, some researchers have focused on how metadata within library records (i.e.: MARC records) have negatively impact social justice and social equity because the data is not integrated with the Web (Schreur & Carlson, 2020, p. 54). Because library records describing cultural heritage collections on minorities and marginalized groups are insulated within library catalogs, this information remains inaccessible to a broader internet-based information community (Boczar et al. 2021). As Boczar et al (2021) highlight, semantic web or linked data technologies, such as BIBFRAME, are uniquely suited to the task of integrating library generated metadata records onto the Web.

Furthermore, the integration of archival metadata onto the semantic web augments the relevance of the library community because they will have a greater presence on the Web. However, the Web and its users also benefit because they will have access to quality-controlled metadata descriptions, which have been organized and maintained by a vast network of distinct libraries with expert knowledge. By ensuring greater access to resource, the library community reveals resource of cultural and societal significance that may elevate the level of public discourse on social justice and social equity.

From the current research on BIBFRAME implementation, it is evident that librarians appreciate the immense benefits of transiting to a linked data environment. Even if this technology will ultimately disrupt the legacy procedures and processes of the library community, this innovation strengthens the relevance of libraries within a broader information community (Schreur & Carlson, 2020; Misu Kim et al., 2021). Librarians may be skeptical about the actual value of linked data technology on the library community because BIBFRAME is an evolving technology with few experts (Christman & King, 2020). According to Christman and King (2020), librarians should experiment and learn the functions of this evolving technology in preparation for the eventual implementation of BIBFRAME. Overall, this report has aimed to present the latest developments on the BIBFRAME implementation within the scope of bibliographic and authority controls.

References

- Bigelow, I., & Sparling, A. (2022). Control or Chaos: Embracing Change and Harnessing Innovation in an Ecosystem of Shared Bibliographic Data. *JLIS.It, Italian Journal of Library, Archives & Information Science*, 13(1), 67–85. <https://doi-org.ezproxy.lib.usf.edu/10.4403/jlis.it-12735>
- Boczar, J., Pollock, B., Mi, X., & Yeslibas, A. (2021). Bridging the Gap. *Information Technology and Libraries (Online)*, 40(4), 1-15. <https://doi.org/10.6017/ital.v40i4.13063>
- Casalini, M. (2022). The future of bibliographic services in light of new concepts of authority control. *JLIS.It, Italian Journal of Library, Archives & Information Science*, 13(1), 107–115. <https://doi-org.ezproxy.lib.usf.edu/10.4403/jlis.it-12766>
- Christman, D., & King, E. (2020). BIBFRAME Basics: A Crash Course. *Serials Librarian*, 78(1–4), 17–21. <https://doi-org.ezproxy.lib.usf.edu/10.1080/0361526X.2020.1730375>
- Jones, E. (2021). Reality Check. *Technicalities*, 41(1), 1-9. <http://ezproxy.lib.usf.edu/login>
- McCallum, S. H. (2022). Collocation and Hubs. Fundamental and New Version. *JLIS.It, Italian Journal of Library, Archives & Information Science*, 13(1), 45–52. <https://doi-org.ezproxy.lib.usf.edu/10.4403/jlis.it-12760>
- Misu Kim, M. C., & Montgomery, D. (2021). Moving toward bibframe and a linked data environment (Ser. *Advances in library administration and organization*, volume 42). In S. Schmeh Hines (Ed.) *Technical Services in the 21st century*. (pp. 131-154). Emerald Publishing Limited. Retrieved July 25, 2022.
- Saha, S., & Mandal, S. (2021). Application of tools to support Linked Open Data. *Library Hi Tech News*, 38(6), 21–24. <https://doi-org.ezproxy.lib.usf.edu/10.1108/LHTN-09-2021-0060>

- Samples, J., & Bigelow, I. (2020). MARC to BIBFRAME: Converting the PCC to Linked Data. *Cataloging & Classification Quarterly*, 58(3/4), 403–417. <https://doi-org.ezproxy.lib.usf.edu/10.1080/01639374.2020.1751764>
- Schreur, P. E., & Carlson, A. J. (2020). Bridging the Worlds of MARC and Linked Data: Transition, Transformation, Accountability. *Serials Librarian*, 78(1–4), 48–56. <https://doi-org.ezproxy.lib.usf.edu/10.1080/0361526X.2020.1716584>
- Niu, J. (2020). Comparing the diffusion and adoption of linked data and research data management services among libraries. *Information Research*, 25(2), N.PAG.