

One Step Forward, or 2.0 Steps Back?
Representing MARC in BIBFRAME 1 and
BIBFRAME 2.0

Alex Keane

31 August 2017

INSTGo99 MA Dissertation

Antonis Bikakis

‘This dissertation is submitted in partial fulfilment of the requirements for the Master’s degree in Library and Information Studies, Department of Information Studies, UCL.’

Referencing in Chicago 2 Style

11,937 Words

Declaration

I have read and understood the College and Departmental statements and guidelines concerning plagiarism. I declare that:

- This submission is entirely my own original work.
- Wherever published, unpublished, printed, electronic or other information sources have been used as a contribution or component of this work, these are explicitly, clearly and individually acknowledged by appropriate use of quotation marks, citations, references and statements in the text. It is 11,937 words in length.

Contents

1. Introduction	3
2. Research on Bibframe: A Literature Review	9
2.1 The Demise of MARC	9
2.2 Bibframe as the New Standard	14
2.3 Bibframe in Practice	19
2.4 Conclusion.....	22
3. Comparing MARC Transformations: Bibframe 1 to Bibframe 2.0.....	24
3.1 Methodology	25
3.2 The Representation of Type in Property and Class	27
3.3 Authorities and Relationships	30
3.4 Representing the Title.....	32
3.5 The Treatment of Notes in Annotation and Item Classes	34
3.6 Conclusion.....	36
4. The Current Status of the Bibframe 2.0 Conversion Specifications.....	38
4.1 Consistency within the Conversion Specifications	39
4.2 Preservation of MARC Data.....	42
4.3 The Structure of the Specification	43
4.4 Introspection and the Library of Congress.....	47
4.5 Summary of Findings.....	48
5. Conclusion	50
Appendix.....	52
Supplementary Spreadsheet Format.....	52
Supplementary Spreadsheet Example Page.....	53
Bibliography	54

1. Introduction

This dissertation aims to evaluate the differences between the first iteration of Bibframe and Bibframe 2.0, as well as assessing how the Library of Congress (LC) has mapped Bibframe 2.0 to the corresponding MARC fields. Bibframe, short for Bibliographic Framework, is an initiative to 'evolve bibliographic description standards to a linked data model'.¹ This means creating bibliographic records from RDF (Resource Description Framework)² triples using a controlled vocabulary.³ Describing resources in such a way makes it easier for computers to understand the data and therefore means the data becomes 'machine-actionable'.⁴ Theoretically, this should allow library catalogues to be more intuitive to the needs of users and more discoverable on the web.

The LC launched Bibframe in May 2011,⁵ following a decade of growing concerns that MARC could no longer fulfil the needs of the library community in the 21st Century. The infamous article *MARC Must Die*,⁶ published in 2002, highlighted the many issues surrounding the continuing use of MARC records that were incompatible and isolated from the rest of the internet. In fact, earlier in that same year, the LC had already begun publishing their work

¹ Library of Congress (April 2016)

² RDF Working Group (2014)

³ Library of Congress (2017a)

⁴ Library of Congress (2016a) Slide 22

⁵ Library of Congress (2012) p. 3

⁶ Tennant (2002a)

towards a framework for MARC data in an XML format.⁷ However, converting MARC records to XML was not enough to unlock all the possibilities of a modern library catalogue. The tree-like structure of the XML model⁸ is still syntactic, wherein it describes the structure of information, but does not allow computers to discern the meaning of the data.⁹ For machines to be able to understand, interpret and use data more effectively, the data model needs to be semantic, focusing on providing the meaning of information as well as the raw data. Antoniou *et al.* describes a 'layered approach to the Semantic Web'¹⁰ that begins with XML as a language at the bottom, before moving to RDF as a data model on the level above. This shows that for the library catalogue to take the next step in integrating itself into the modern web of data, an RDF model is needed. MARCXML is incapable of linking library resources on more than a superficial level, either internally or with external information databases. A new framework based in RDF is required to open the library catalogue to the wider web of resources and to take advantage of linked data. The Bibframe Initiative is aiming to be one of the first major attempts to provide libraries with a well-supported RDF framework and a specialised ontology.

The Resource Description Framework, or RDF, is described as a 'standard model for data interchange on the Web'¹¹ and it facilitates the connection of data across the Web. RDF uses Universal Resource Indicators (URIs) to define individual items of data, from a web page to a

⁷ Library of Congress (2016b)

⁸ W3 Consortium (2005)

⁹ Antoniou et al (2012) p. 9

¹⁰ *Id.* pp. 17-18

¹¹ RDF Working Group (2014)

physical thing, before linking them together with a relationship that can also be described with a URI. This creates a web-like graph of data where each individual item, as well as the relationships between items, can be defined by unique URIs. Ontology is a formally defined language that can be used by RDF and it consist of classes and properties. In this case, Bibframe is the ontology used to describe bibliographic records in RDF, which allows items to be specifically referred to as classes and their relationships to be described as properties. A relationship between two items, such as an author and a book, can be described as the subject class of 'Book' with the property of 'isWrittenBy' and the object class of 'Author'.

Since the original Bibframe model was proposed, the LC has made significant changes to both the model and its ontology. The draft specifications for Bibframe 2.0 were announced in October 2015,¹² and the vocabulary has been subject to constant change and frequent updates. The major difference between the two iterations of Bibframe is the removal of the core Authority and Annotation classes and the introduction of the core Item class¹³. This brings Bibframe 2.0 closer to the Functional Requirements for Bibliographic Records (FRBR) model of Work-Expression-Manifestation-Item (WEMI),¹⁴ with the Bibframe Work representing the FRBR Work and Expression, and the Bibframe Instance and Item reflecting the FRBR Manifestation and Item respectively.¹⁵ Another change is the introduction of the

¹² Kroeger (2016) p. 7

¹³ Library of Congress <https://www.loc.gov/bibframe/docs/bibframe2-whatsnew.html>

¹⁴ International Federation of Library Associations (2009) pp. 33-48

¹⁵ Kroeger (2016) p. 14

‘Key Concepts’: Agents, Subjects and Events.¹⁶ Finally, the vocabulary of Bibframe has been significantly reviewed, with more classes being added and many properties being refined or removed; the number of classes increased from 53 to 73, while the number of properties has decreased from 289 to 137.¹⁷ These changes are assessed in more detail in chapter 3 of this dissertation.

With the release of the Bibframe 2.0 specification, the LC states that their aim is to ‘make bibliographic information more useful both within and outside the library community’.¹⁸ This highlights the core benefit of Bibframe and linked data in general: to facilitate the exchange of information between communities. The effectiveness of Bibframe 2.0 in achieving this goal shall be evaluated in chapter 4. The inclusion of the Event class demonstrates the willingness to link bibliographic records with sources outside of the library, something that has not been possible with MARC. However, in order for libraries and content producers to make the leap to Bibframe records, the mapping of legacy MARC records to the modern standard will have to be as lossless as possible. It remains to be seen whether the current mapping allows for an accurate representation of the bibliographic record without any loss of data, whilst also enriching the catalogue with genuine added usability. It is possible that Bibframe will not be the answer to introduce the library catalogue fully to the Semantic Web. It is also possible

¹⁶ Library of Congress (April 2016)

¹⁷ Taken from the vocabulary lists at <http://bibframe.org/vocab-list/> and <http://id.loc.gov/ontologies/bibframe.html>

¹⁸ Library of Congress (April 2016)

that Bibframe could be a transitional model that will at least show that the library community is finally ready to move on from a forty-year-old MARC standard that predates the internet.

The question that this dissertation proposes to answer is this:

How has Bibframe 2.0 changed from the first iteration of Bibframe and what benefits, if any, does it offer to bibliographic description over the previous model?

To answer this question, the LC's¹⁹ most recent conversion specifications that demonstrate how the MARC fields map to the corresponding Bibframe vocabulary have been assembled into a single spreadsheet.²⁰ As there are no conversion specifications available for the first iteration of Bibframe, the old converter tool²¹ has been used to discover how MARC was mapped to the first iteration of Bibframe. This is achieved by entering MARCXML fields into the converter and tracking how the data is output. This information has also been added to the spreadsheet, allowing for MARC fields, Bibframe 1, and Bibframe 2.0 to be compared side by side. This, in combination with the information on the LC website and other scholarship, clearly demonstrates the differences from the first iteration of Bibframe and Bibframe 2.0, as well as showing the success of both versions in representing the original MARC records. The spreadsheet is included in the supplementary content of this dissertation.

¹⁹ Library of Congress (2017b)

²⁰ Supplementary spreadsheet

²¹ Bibframe <http://bibframe.org/tools/transform/start>

Chapter 2 reviews the current literature on the Bibframe initiative, including its benefits over MARC, what the changes in Bibframe 2.0 will mean, and what the Bibframe model hopes to achieve. This chapter assesses the potential of Bibframe, the issues that recent scholarship has found with the previous iteration of Bibframe, and what needs to be done to answer these criticisms. Chapter 3 of this dissertation focuses more specifically on the changes between the first Bibframe model and Bibframe 2.0 and, in particular, the move from the Work-Instance-Agent-Annotation model to the Work-Instance-Item model. These differences are illustrated by comparing the mapping of MARC fields through the first iteration of Bibframe to Bibframe 2.0. Chapter 4 evaluates the current mapping of Bibframe 2.0 to MARC, a process that is still being developed by the LC, before offering recommendations on what still needs to be done. The extent to which Bibframe is able to represent the MARC data and bibliographic data in general is assessed, with the purpose of providing a clear report on the progress of Bibframe and its viability as the next bibliographic framework. The scope of this dissertation is to focus exclusively on textual records, avoiding the assessment of audio-visual material, in order to allow for a more detailed review of the Bibframe mapping.

2. Research on Bibframe: A Literature Review

This chapter reviews the scholarship that addresses the decline of MARC and the transition to Bibframe. Since Bibframe 2.0 is still a relatively new development, much of the established literature refers to the first iteration of Bibframe. For this reason, it is important to assess the criticisms levelled at both the original Bibframe Work-Instance-Annotation-Item model²² and at previous practical implementations to see if these issues are likely to remain with Bibframe 2.0. This review highlights areas where the previous model of Bibframe was weakest, allowing the later chapters to focus more specifically on key areas that needed development or remodelling to make Bibframe 2.0 a viable standard. The first section of this literature review, section 2.1, focuses on the decline of MARC, the development of RDF and linked data, and how Bibframe 1 and 2 were conceived. Section 2.2 evaluates the literature that compares the original Bibframe model to other models, specifically IFLA's FRBR, and assesses how relevant these criticisms remain with the introduction of Bibframe 2.0. Finally, section 2.3 of this chapter reviews cases where Bibframe has been practically implemented to discover if any of the issues were caused specifically by the first iteration of Bibframe, or if the problems were inherent with the theory of linked bibliographic data itself.

2.1 The Demise of MARC

Firstly, this section shall introduce what MARC is. MARC stands for MACHine-Readable Cataloguing. MARC is maintained by the LC as 'the mechanism by which computers

²² Kroeger (2013) p. 873

exchange, use, and interpret bibliographical information'.²³ It forms the standardised template in which bibliographic records can be stored, shared and interpreted. Cataloguers input information describing a bibliographic resource, such as a book, into a MARC record that is organised by fields. Library databases use data stored in these MARC fields to display information such as the author, title and publisher on an online catalogue.

Cataloguers have been predicting the demise of MARC for over a decade and, as mentioned in the introduction of this dissertation, Roy Tennant was a major influence in pushing for alternatives to MARC. Tennant published two articles in 2002: the first highlighting the deficiencies in MARC,²⁴ and the second suggesting ways to move away from the standard. Among other things, Tennant complained that 'relationships among related titles are problematic in MARC',²⁵ showing how current library records are isolated from other professions. This is demonstrated in the MARC 246 field, where the relationship of the 'Varying Form of Title' can only be described as a textual note in many cases,²⁶ meaning machines are unable to process the connection between the two titles. Since then, countless articles have discussed the future of MARC, culminating in a working group convened by the LC that stated MARC is 'no longer fit for the purpose [of being a metadata carrier]'²⁷ in 2008. This report by the Library of Congress Working Group highlighted that library bibliographic data needed to move from the 'closed database model to the open Web-based model', thus

²³ Library of Congress (2006)

²⁴ Tennant (2002a); (2002b)

²⁵ Tennant (2002a)

²⁶ Library of Congress (2014)

²⁷ Library of Congress Working Group (2008) p. 25

allowing libraries to ‘take advantage of the relationships that exist [...] on the Web’.²⁸ The importance of describing relationships between both bibliographic records and other sources of data is reflected in modern cataloguing theory. IFLA’s study group on Functional Requirements for Bibliographic Records (FRBR) was set up to describe ‘a generalized view of the bibliographic universe, intended to be independent of any cataloguing code or implementation’.²⁹ FRBR displays an entity-relationship structure³⁰ in its model, encouraging links between the conceptual ‘Group 1’, ‘2’ and ‘3’ entities, or bibliographic, individual and subject entities respectively.³¹ Such a model allows these entities to be recorded once and then assigned a universal identifier, such as with an authority record, before allowing them to be linked by this identifier as required. Along with this new theoretical model, the Resource Description and Access (RDA) Toolkit was released as a new cataloguing standard, using FRBR as a basis.³² RDA differed from its predecessor, AACR2 (Anglo-American Cataloguing Rules), in a number of ways, but, noticeably, the new rules specified the need for recording relationships.³³ These relationships include those defined by FRBR. It is unquestionable, then, that recording the relationships between data has become a necessary requirement for the future of the catalogue. The issue is that MARC records are flat, meaning that the same information must be entered into each record, which reduces the potential for links between

²⁸ *Id.* p. 26

²⁹ Tillett (2003) p. 2

³⁰ IFLA (2009) p. 9

³¹ *Id.* pp. 14-16

³² RDA Steering Committee http://rda-rsc.org/content/rda_faq#19

³³ Croissant (2012) p. 17

data to be created and described. This is highlighted by the fact that the U.S. RDA Test Coordinating Committee, set up by three national American libraries, concluded in 2011 that in order to implement RDA effectively there would need to be ‘credible progress towards a replacement for MARC’.³⁴

Developments in cataloguing theory, coming from both FRBR and RDA, suggest that representing relationships will be integral to the next cataloguing standard. While it is generally acknowledged that the future standard will be encoded in XML,³⁵ it is also acknowledged that simply migrating the MARC format into XML will not resolve the issue. Coyle³⁶ has suggested that moving to XML will ‘not necessarily encourage any modification of the fundamental content of the MARC record itself’. Essentially, the MARC design lacks any form of hierarchy or inheritance, which means data elements of higher levels cannot be ‘inherited to linked levels below them’.³⁷ As a result, machines are unable to interpret the data, since any links that can be made are flat and have no semantics.³⁸ Several articles have been published supporting the idea that MARC is not, and never will be, suitable for supporting machine-actionable metadata.³⁹ The solution is to use linked data to enable semantic links that are machine-readable.⁴⁰ Allemnag and Hendler⁴¹ state that ‘Instead of

³⁴ *Id.* p. 20

³⁵ Kroeger (2013) p. 876

³⁶ Coyle 2004 p. 167

³⁷ *Id.* p. 168; Kroeger (2013) p. 876

³⁸ Alemu *et al.* (2012) p. 551

³⁹ Coyle (2010); Coyle & Hillman (2007); Wallis (2011)

⁴⁰ Alemu *et al.* (2012) p. 551

⁴¹ Allemnag & Hendler (2008) p. 7

having one webpage point to another, one data item can point to another', and this data can be linked by a globally controlled and understood vocabulary. It is widely accepted that linked data will be realised through RDF. In fact, it is debatable whether linked data can even exist without RDF.⁴² It would appear, then, that if a replacement for MARC is to be found, it will be expressed in RDF.

However, there has also been debate among some librarians as to whether MARC needs to, or even can, be replaced. One major example, a report compiled by Karen Calhoun for the LC,⁴³ set out to interview a series of leading library professionals on the future of library catalogues. While Calhoun's report did state that libraries should 'prepare for linkages in and out of the catalogue',⁴⁴ as well as interoperability and the sharing of data,⁴⁵ it also asserted that MARC would remain. Following the responses of the interviewees, 'the consensus was that MARC is not going anywhere'.⁴⁶ Although, it is worth noting that the senior position of these interviewees,⁴⁷ and a life time of MARC may have meant they were less open to a change in cataloguing standards. The belief that MARC needs to be replaced imminently has been debated many times,⁴⁸ and it is a reasonable concern that the sheer extent and variety of legacy MARC records will require a great deal of effort to be converted. While MARC in its

⁴² Cyganiak (2009)

⁴³ Calhoun (2006)

⁴⁴ *Id.* p. 17

⁴⁵ *Ibid.*

⁴⁶ *Id.* p. 32

⁴⁷ *Id.* pp. 29-30

⁴⁸ Coyle (2010); Coyle & Hillmann (2007); Dunsire (2008, 2009, 2012); Dunsire and Willer (2011); Marcum (2011); Styles (2009); Styles *et al.* (2008); Wallis (2011)

current form is clearly not performing to the required standard to make the most of our bibliographic data, it remains to be seen if an RDF replacement, such as Bibframe, will bring enough benefits to outweigh the cost of migrating literally billions of records⁴⁹ from MARC. It is not within the scope of this dissertation to suggest when MARC should be replaced, but over the next two sections the Bibframe model and its practical application is reviewed with the aim of assessing its viability as the descendant of MARC.

2.2 Bibframe as the New Standard

The Bibliographic Framework Initiative has gained support for a number of reasons, but mainly because it has been championed by the LC, just as MARC was 40 years ago. Bibframe contains 53 classes and 289 properties, used to describe bibliographic records in RDF. The Bibframe model was conceived in light of both FRBR and RDA,⁵⁰ and, while there are some differences, the entity-relationship model of FRBR is still compatible with Bibframe.⁵¹ As mentioned in the Introduction, the main difference between the conceptual FRBR model and Bibframe is that Bibframe has taken a ‘reductionist’⁵² approach, leading to a contraction of the four FRBR entities. In Bibframe, the FRBR entities of ‘Work’, ‘Expression’, ‘Manifestation’ and ‘Item’ have been reduced to ‘Work’ and ‘Instance’, while new entities of ‘Authority’ and ‘Annotation’ have been added.⁵³ This model is visualised in Fig. 1, with ‘Work’ and ‘Instance’ featured in the centre and ‘Creator’, ‘Subject’ and ‘Publisher’ demonstrating examples of

⁴⁹ Marcum (2011)

⁵⁰ Library of Congress (2012) p. 37

⁵¹ *Id.* p. 36; Sprochi (2016) p. 133

⁵² Library of Congress (2012) p. 15

⁵³ Library of Congress (2012) p. 8

'Authority' around the outside. The consensus is that the Bibframe 'Work' roughly corresponds to the FRBR entities 'Work' and 'Expression', while 'Instance' relates to the FRBR 'Manifestation'.⁵⁴ However, there is some debate as to where the FRBR notion of 'Item' is represented in the Bibframe model. Sprochi⁵⁵ would link 'Item' to the Bibframe 'Instance', Baker *et al.*⁵⁶ would not link 'Item' to any Bibframe class, and Kroeger⁵⁷ chooses to display 'Item' as referring to either 'Instance' or 'Annotation'. It would appear from sample Bibframe records that copy-specific information, such as shelf mark and general notes, are part of the 'Annotation' class.⁵⁸ Ballegooie & Borie⁵⁹ suggest that 'Item' is attached to the Bibframe 'Instance' by 'Annotation', describing the Bibframe model as a simplified and less hierarchical interpretation of FRBR, and this explanation seems the most plausible.

⁵⁴ Sprochi (2016) p. 133; Baker, Coyle & Petiya (2014) p. 572; Kroeger (2014) p. 12

⁵⁵ Sprochi (2016) p. 133

⁵⁶ Baker, Coyle & Petiya (2014) pp. 572-3

⁵⁷ Kroeger (2014) p. 12

⁵⁸ Bibframe (2013)

⁵⁹ Ballegooie & Borie 2014 p. 82

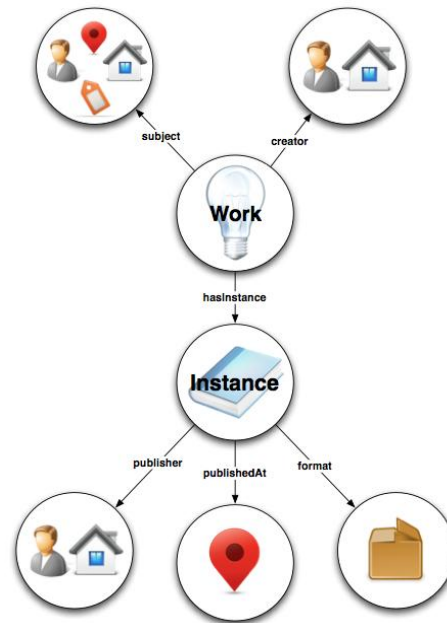


Fig. 1. The Bibframe 1 model. Library of Congress (2012) p. 8.

The implications of this difference are hard to discern, however. Ballegoie & Borie⁶⁰ speculate that the simplification will be beneficial to serials cataloguers, as the FRBR hierarchy ‘does not currently address all the needs of serial publications’. Beyond this, there is not much in the way of an evaluative comparison between the two models. According to Baker *et al.*,⁶¹ criticisms have been raised against Bibframe for not using the WEMI model, although they have not cited these criticisms. Kroeger⁶² did explain that the ‘handling of item-level information always felt woefully inadequate, like it was tacked on as an afterthought’. The relegation of ‘Item’ to a constituent of ‘Annotation’ is an obvious complaint given that libraries are still dealing with individual physical items rather than just theoretical ‘Instances’ of a conceptual ‘Work’. Equally, the FRBR WEMI model was criticised

⁶⁰ *Ibid.*

⁶¹ Baker, Coyle & Petiya (2014) pp. 573

⁶² Kroeger (July 2017)

for having too many classes. Edmunds⁶³ believes that the FRBR entity 'Expression' 'was difficult to grasp and entirely useless in its application to bibliographic description'. This idea is supported by Styles,⁶⁴ arguing that such abstractions as 'Expression' and 'Manifestation' are not in everyday use by publishers, cataloguers and library users. Wallis⁶⁵ also finds attractiveness in the simplified model as this allows 'non-domain experts to understand, reference and link to their rich resources'. Still, it would seem that in attempting to simplify the unnecessarily obtuse FRBR model, the first iteration of Bibframe removed one too many entities. The simple notion of the Bibframe 'Work' representing the abstract idea, and the 'Instance' representing the material item,⁶⁶ was unsatisfactory for expressing copy-level library holdings data.

However, Bibframe 2.0 introduced a new vocabulary and a revised model, shown in Fig. 2. The number of classes changed to 73, while the number of properties decreased to 137. The new model included 'Item' as a main class, resolving the under-represented copy-level issue mentioned above, while also relegating 'Annotation' to its constituent properties (such as 'tableOfContents') and eliminating 'Authority' altogether.⁶⁷ This latter change brings the treatment of 'Agents' and 'Subjects' more in line with other metadata standards, such as Encoded Archival Description, and this will facilitate better interoperability.⁶⁸ One of the

⁶³ Edmunds (2017) p. 1

⁶⁴ Styles (2009)

⁶⁵ Wallis (2011)

⁶⁶ Library of Congress (2012) p. 10

⁶⁷ Library of Congress <https://www.loc.gov/bibframe/docs/bibframe2-whatsnew.html>

⁶⁸ Kroeger (July 2017)

benefits of encoding in RDF is that bibliographic records can be linked with non-library metadata and vice versa, meaning similarities between standards are generally to be encouraged.⁶⁹ Another change in the Bibframe 2.0 model is the introduction of the ‘Event’ class as a key concept.⁷⁰ This allows videos, musical performances and conferences, among other things, to be linked to a specific event identified with a unique URI.⁷¹ This benefits audio-visual metadata and provides a way to link bibliographic data to information outside of the library silo. Overall, the major criticism that plagued the first iteration of Bibframe has been answered with the introduction of the ‘Item’ class. Meanwhile, other concerns that Bibframe strayed too far from the FRBR WEMI model may actually be more of a reflection on the latter model’s over-complexity and overly narrow focus on textual materials above audio-visual data.⁷² While Bibframe may have chosen not to follow FRBR directly, it has brought library data closer to other metadata standards, allowing for compatibility and shared links across professions – an ideal that Tennant called for in 2002.⁷³

⁶⁹ Kroeger (2013) p. 885

⁷⁰ Library of Congress <https://www.loc.gov/bibframe/docs/bibframe2-whatsnew.html>

⁷¹ Library of Congress (March 2017c)

⁷² The word ‘text’ is mentioned 30 times in IFLA’s (2009) *Functional Requirements for Bibliographic Records*, but ‘audio’ is mentioned only 7 times.

⁷³ Tennant (2002a)

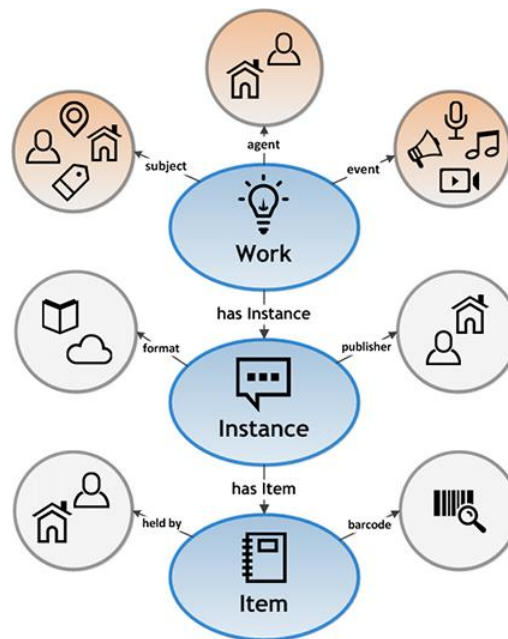


Fig. 2. The Bibframe 2.0 model. Library of Congress (April 2016).

2.3 Bibframe in Practice

As Edmunds⁷⁴ states in an otherwise scathing review of Bibframe, while the model itself may be perfectly functional, its success depends entirely on its practicalities and adoption rate. This assessment also mentioned that the support given by the LC should provide Bibframe with an advantage. However, libraries and library system vendors will need to see concrete evidence of the practical applications of Bibframe before it can be widely adopted as a new standard. Since the conception of Bibframe 1, a number of institutions have published findings from pilot projects, but the results are often preliminary due to low funding and

⁷⁴ Edmunds (2017) p. 7

commitment at such an early stage. Edmunds⁷⁵ summarises a number of projects that were unsuccessful or inconclusive in attempting to adopt Bibframe in the early stages, with BIBFLOW in particular standing out. BIBFLOW⁷⁶ was a two-year project working with Zepheira, the software company that helped create Bibframe,⁷⁷ to ‘investigat[e] the future of library technical services’. A report is yet to be published, although their initial findings⁷⁸ demonstrated the difficulty of implementing Bibframe into a library system serviced by ‘over 30’ software applications. With that in mind, it is important to understand that many pilot studies are conducted in isolated conditions, eliminating a great deal of complexity that a total conversion to Bibframe would require. Even the National Library of France (BnF),⁷⁹ using its own linked open data model, was only able to create such an expansive database because of previously established ARK (Archival Resource Key) links⁸⁰ while still creating and maintaining bibliographic data using UNIMARC.⁸¹

Despite this, it is still worth reviewing practical tests conducted with Bibframe to assess its strengths and weaknesses of representing data. The University of Washington compared RDA/RDF to Bibframe and found that, while RDA/RDF was stronger in some areas such as notes and series, Bibframe handled identifiers, URI’s, transcription and subject headings

⁷⁵ *Id.* pp. 4-5

⁷⁶ BIBFLOW (2016a)

⁷⁷ Zepheira <https://zepheira.com/>

⁷⁸ BIBFLOW (2016b)

⁷⁹ Bibliothèque nationale de France <http://catalogue.bnf.fr/index.do>

⁸⁰ Illien (2013) p. 26

⁸¹ Edmunds (2017) p. 6

better.⁸² Arguably, URI's and identifiers can be seen as an integral part of linked data, as they link the bibliographic data together. Tharani⁸³ assessed the feasibility of harvesting and sharing bibliographic metadata with Bibframe, using The Harvard Collection of Israeli Literature as a sample, and found a great deal of success. The case study reports that 'seemingly disparate library systems and data were integrated to provide a unified and [accessible] thematic research collection'.⁸⁴ Furthermore, Tharani believes that Bibframe can 'accommodate cultural and indigenous resources',⁸⁵ as well as allowing users to uncover credible data on the web more easily.⁸⁶ Shieh⁸⁷ describes how George Washington University Library was able to test the vocabulary and data modelling of Bibframe. The report was cautiously optimistic of the model and vocabulary, stating that it 'appeared to have passed and validated the original goal',⁸⁸ while also expressing that more work was needed if Bibframe hoped to replace MARC entirely.⁸⁹ University College London began the Linked Open Bibliographic Data Project as an educational resource, finding that the 'binary choice' between 'Work' and 'Instance' was pragmatic, while also noticing that the use of RDA principles meant cataloguers would often have to use their judgement.⁹⁰ Drummond compared Bibframe with two other data models, FRBR-aligned Bibliographic Ontology

⁸² Kiegel (2015)

⁸³ Tharani (2015) p. 10

⁸⁴ *Id.* p. 16

⁸⁵ *Id.* p. 17

⁸⁶ *Ibid.*

⁸⁷ Shieh (2013) p.20

⁸⁸ *Ibid.*

⁸⁹ *Id.* p. 21

⁹⁰ Welsh *et al.* (2015) p. 4

(FaBiO) and The Europeana Data Model (EDM), finding that Bibframe benefitted from a 'cataloguing focus'.⁹¹ This reflects the benefit of having the LC design the data model. However, Drummond also found that the structure of Bibframe meant it could be confusing to use,⁹² reflecting the need for improvements in the vocabulary. The University of Illinois transformed a major collection of 300,000 e-books to Bibframe records to try to create a new discovery system.⁹³ The results were mixed, with financial and technical limitations meaning that they were unable to create 'Work' - 'Work' or 'Instance' - 'Instance' relationships and occasionally transformations were hampered by overloaded APIs.⁹⁴ These issues highlight how much work it will take to not only implement Bibframe fully, but also just to convert billions of MARC records into the Bibframe standard. On the other hand, Jin *et al.* did find Bibframe to be a 'profound step for the library community',⁹⁵ suggesting that Bibframe would vastly improve discovery for users.

2.4 Conclusion

Overall, many case studies were unable to be implemented on a scale that truly demonstrated how Bibframe could work between institutions. The initial signs seem promising, but it may take years for Bibframe 2.0 to be tested fully. While there did not seem to be much issue translating the first Bibframe model into practice, it is clear that moving away from MARC will be a massive undertaking. Much bigger studies will need to be

⁹¹ Drummond (2015) p. 50

⁹² *Ibid.*

⁹³ Jin, Hahn & Croll (2016) p. 223

⁹⁴ *Id.* pp. 231-232

⁹⁵ *Id.* p. 233

conducted on Bibframe 2.0 in order to test its practicality fully. In addition, these studies began by transforming MARC data in some way, so it is worth assessing how effectively Bibframe and Bibframe 2.0 are able to do this. Chapter 3 assesses whether concerns over the simplistic Bibframe model have been addressed. This chapter aims to compare how the LC has mapped first iteration of Bibframe and Bibframe 2.0 to MARC and to discern how the new model will influence this transformation.

3. Comparing MARC Transformations: Bibframe 1 to Bibframe 2.0

This chapter compares how MARC records have been converted to the first iteration of Bibframe, with the conversion specifications for Bibframe 2.0.⁹⁶ The aim is to view the differences between the two Bibframe models and assess how they are able to represent the MARC data. The spreadsheet provided as supplementary content is used to show how the vocabulary differs between Bibframe 1 and Bibframe 2.0. This provides a clear point of reference when addressing how key aspects of bibliographic data, such as names and titles, are handled by these two models. First, section 3.1 of this chapter introduces the spreadsheet and explains how it was compiled and the decisions taken in creating it. Next, section 3.2 compares the major structural difference between the two vocabularies: specifically, how ‘types’ are represented and the use of classes and properties. This explains the noticeable change in classes and properties described in the Introduction. Then, the treatment of Authorities and contributors is evaluated in section 3.3. This reflects how the removal of the ‘Authority’ core class, described in section 2.2, has affected the Bibframe 2.0 vocabulary. The following section, section 3.4, highlights the treatment of titles, including how the cover page is represented, how multiple titles are handled and how both models are able to cope with changing titles in a series. Finally, section 3.5 of this chapter addresses how

⁹⁶ Library of Congress (2017b)

notes are used following the elimination of the Annotation class and the introduction of the Item class described in section 2.2.

3.1 Methodology

The spreadsheet that has been compiled as supplementary contents has been created from scratch, using the data provided in the LC conversion specifications and information taken from the Bibframe 1 transformation tool. It shows every MARC field and subfield relevant to bibliographic data, alongside the Bibframe 1 and Bibframe 2.0 classes and properties used to represent this data. The final two columns also provide any comments made by the LC in the conversion specification for Bibframe 2.0, as well as any issues found in reviewing these specifications for this dissertation. The first page of the spreadsheet can be found in the Appendix to demonstrate the structure. The spreadsheet provides visual feedback on the progress made in converting the MARC fields. The cells filled in red show where the LC have made ‘no attempt to convert’ (nac) the MARC fields into the Bibframe 2.0 vocabulary, suggesting that they will be reviewed at a later date.⁹⁷ The red cells also highlight mistakes and areas that require revisiting in the Bibframe 2.0 conversion specifications. These issues will be addressed in chapter 3. The cells filled in green show MARC fields that the LC has chosen to ignore when converting records to Bibframe 2.0, meaning they are seen as unnecessary.

⁹⁷ *Ibid.*

The first four columns, columns A – D of the spreadsheet, contain the MARC data, including the first and second indicators, and the subfield. MARC fields 006 and 008 have a fixed number of characters rather than subfields and this has been represented in column A. The next three columns, columns E – G, contain the Bibframe 1 vocabulary, with the classes and properties placed in the same row as the MARC field that they represent. This is not the full extent of the vocabulary as a detailed conversion specification is not available, but it provides enough information in key fields to highlight the differences between the two models. Column E describes whether the properties and classes from columns F and G are used on a Work, Instance, Annotation, or Authority level. Columns F and G show the Bibframe 1 property and the relevant class in its range, respectively. The data within these three columns was created using the Bibframe 1 transformation service.⁹⁸ MARCXML data was input manually, before the Bibframe 1 output was reviewed and entered into the spreadsheet. Columns H – K show the Bibframe 2.0 vocabulary and comments. This data was aggregated from the conversion specifications provided by the LC.⁹⁹ Again, column H shows the conceptual level being described (Work, Instance or Item), while Columns I and J display the property and related class respectively. Column K contains the comments provided by the LC to help aid conversion, as well as the expected value or rdfs:label. Occasionally, exposition has been added to improve clarity, but the information within these columns remains unchanged. The final column, column L, is used to highlight issues in the Bibframe 2.0

⁹⁸ Bibframe <http://bibframe.org/tools/transform/start>

⁹⁹ Library of Congress (2017b)

specification and to draw attention to some areas that should be revisited and why. These instructions have been summarised in the Appendix.

3.2 The Representation of Type in Property and Class

The most notable difference between the vocabularies of the first iteration of Bibframe and Bibframe 2.0 is how ‘types’ are indicated. In the Bibframe context, ‘type’ refers to different kinds of a certain resource, such as the different ‘types’ of ‘Identifier’.¹⁰⁰ Types of identifiers are represented as properties in the first iteration of Bibframe, including ‘isbn10’, ‘legalDeposit’ and ‘fingerprint’.¹⁰¹ This means that, when describing an Instance of a bibliographic record, the Instance would have the property ‘isbn10’ with the value of either a literal string or, preferably, a URI. This method means that it was the relationship between the subject and the object that represented ‘type’. However, in Bibframe 2.0, the LC changed its approach to RDF conventions. ‘Type’ is now represented by different classes rather than by different properties as it was previously.¹⁰² Using the example of ‘Identifier’, there is now one identifier property, ‘identifiedBy’, while there are multiple classes for the different identifier types, such as ‘Isbn’, ‘CopyrightNumber’ and ‘Fingerprint’.¹⁰³ As a result, instead of the ‘type’ being represented by the relationship between two resources, it is represented by the resource itself. This means that the URI or value of an ISBN is understood as an ‘Isbn’ that

¹⁰⁰ Library of Congress (March 2017a) p. 2

¹⁰¹ Supplementary spreadsheet, cell F:190, F:179, F222

¹⁰² Library of Congress (March 2017a) p. 2

¹⁰³ Supplementary spreadsheet, cell J:190, J:179, J222

‘identifies’ an ‘Instance’, rather than being an unknown ‘Identifier’ URI or value with a relationship that describes it as an ‘isbn’.

The benefit of representing different types of a resource as a class is that it allows for greater interoperability. Firstly, specific classes can be more easily linked to by other communities outside of the Bibframe context. The class of ‘Isbn’ reflects the identifier source, meaning that this source will be known by external ontologies if they want to use ‘Isbn’ as an identifier.¹⁰⁴ In comparison, the first iteration of Bibframe would record ISBN as simply an ‘Identifier’, meaning the source would only be known in a Bibframe context. Furthermore, if an external namespace created a new type of identifier, it will resist degradation better if this external identifier can be received as a class. For example, if the external namespace (ex:) introduces a property for the new identifier, the resource will have the property ‘ex:newProperty’ with the object ‘Identifier’. If this namespace fails, the relationship between the resource and the ‘Identifier’ will be unknown and the statement will be meaningless. Whereas, if the resource has the property ‘identifiedBy’ and the object is the external class ‘ex:NewClass’, then it is at least clear how the content of ‘ex:NewClass’ is related to the resource, even if the class itself is not understood.¹⁰⁵ In both cases, these improvements mean that Bibframe 2.0 is more flexible and can be more easily integrated outside of the library community. Reducing the number of properties simplifies Bibframe and makes it

¹⁰⁴ Library of Congress (March 2017a) p. 2

¹⁰⁵ *ibid.*

easier to understand and more attractive for external members of the linked data community to engage with.¹⁰⁶

However, with simplicity comes a loss of specificity. By reducing the number of properties and simplifying them, Bibframe 2.0 is no longer able to specify the 'Domain' or 'Range' of its properties. Range and Domain constraints are used for 'future-proofing'¹⁰⁷ as they clearly define how certain properties should be used, allowing for a more consistent use of the vocabulary. Originally, Bibframe 1 was able to define the Domain of many of its identifiers to one of the core classes, usually 'Instance', with the Range of 'Identifier'. This meant that the property 'isbn10' could only be used to relate the object of 'Identifier' to the subject of an 'Instance'. The rigidity of these constraints can cause unexpected issues, as shown when Drummond¹⁰⁸ was prevented from applying the property of 'editionStatement' to two distinct Works by its 'Domain' of 'Instance'. However, these issues can be resolved by improving the vocabulary, albeit at the expense of adding complexity. The LC's response to losing this ability to define a property's role formally is to record 'expected values' in documentation, thus providing informal guidance for the cataloguer.¹⁰⁹ Ironically, the LC states that these rules 'should be well-document',¹¹⁰ and this typo demonstrates the risk of relying on informal constraints rather than incorporating them into the vocabulary. Within

¹⁰⁶ Library of Congress (September 2016) timestamp: c.1:10:00 Transcript available from: <https://stream-media.loc.gov/webcasts/captions/2016/160906dfy1000.txt>

¹⁰⁷ W3C (1999)

¹⁰⁸ Drummond (2015) p. 37

¹⁰⁹ Library of Congress (March 2017a) p. 4

¹¹⁰ *ibid.*

the Bibframe 2.0 conversion specifications, there are already a few instances of human error, such as a mix up between ‘rdf:value’ and rdfs:label’¹¹¹ and at least two incorrect URIs.¹¹² Future revisions to the conversion specifications will definitely be needed if the LC hopes to keep the vocabulary use consistent among external library cataloguers.

3.3 Authorities and Relationships

The comparison between the first iteration of Bibframe and Bibframe 2.0 has highlighted the change in the way that authority is handled. A minor change is the removal of the property ‘authorizedAccessPoint’,¹¹³ as this was rendered superfluous when a URI or label is used.¹¹⁴

The major change has been in displaying contributor relationships. There were two properties used to describe agent contributions in Bibframe 1: ‘creator’ and ‘contributor’.¹¹⁵

These would link from the Work or Instance to an Authority, which represented the agent. If the role of the agent could be more specified further, the property ‘relator’ was used to define the role. If a URI for the role was available, the prefix of ‘relators:’ allowed the role to be defined by the relator namespace, before linking it to the relevant Authority. Otherwise, the property ‘relator’ linked the Work to the class ‘Relator’ and defined this class with the property ‘relatorRole’ and literal string. The relator was then linked to an Authority.¹¹⁶ This method involved three different ways of linking Works to agents¹¹⁷ and this could lead to

¹¹¹ Supplementary spreadsheet, cell K:427, K429

¹¹² *Id.* cell K:94, K:115

¹¹³ *Id.* cell F:443

¹¹⁴ Library of Congress (September 2016) timestamp c.0:26:00

¹¹⁵ Supplementary spreadsheet, cell F:439, F:1147

¹¹⁶ Library of Congress (April 2014)

¹¹⁷ Library of Congress (October 2015) p. 1

inconsistencies between records. This idea is supported by Drummond,¹¹⁸ who felt that ‘the variety of methods to express the same information could lead to some inconsistency in the data set’. Similarly, using a URI to define the role could create issues in the future if these URIs or roles are to be redefined, as this will cause the relationship between the Work and the Authority to be unknown and meaningless.

In Bibframe 2.0, there is just one property describing the relationship between the resource and an agent: ‘contribution’.¹¹⁹ This property links the Work, Instance or Item to the class ‘Contribution’, which is then described by the properties ‘agent’ and ‘role’. The property ‘agent’ links ‘Contribution’ to the agent, which can be assigned a URI or an rdfs:label, while the property ‘role’ links ‘Contribution’ to a URI or rdfs:label that describes the role of the agent. This change represents a departure from the ‘main entry’ and ‘added entry’¹²⁰ fields in MARC and moves away from RDA, which specifies the ‘Creator’ as being associated with the FRBR Work and the ‘Contributor’ as being associated with the FRBR Expression.¹²¹ The change means Bibframe 2.0 can simplify how contributors are treated, allowing them to be linked at the necessary conceptual level without needing to consider whether an editor is deemed a ‘creator’ or a ‘contributor’. The ‘Work’ will simply have a ‘contribution’ from a ‘Person’, with the ‘role’ of ‘rdfs:label “Editor”’. This is easier for the cataloguer to understand and will improve consistency across bibliographic records.

¹¹⁸ Drummond (2015) p. 31

¹¹⁹ Supplementary spreadsheet, cell I:439, I:1147

¹²⁰ Library of Congress (1999)

¹²¹ University of Washington (2016) p. 14

3.4 Representing the Title

Drummond has already criticised how Bibframe 1 represented title data, finding that it was unclear whether the properties ‘instanceTitle’ and ‘workTitle’ should both be used in cases where the Instance and Work titles are identical.¹²² Furthermore, Drummond found two properties with the name of ‘title’, where one property seemed capable of fulfilling the role.¹²³ Since Drummond’s dissertation was published, the Bibframe 1 vocabulary had updated the names of these two properties to ‘title’ and ‘titleValue’,¹²⁴ but their definitions were still unclear. This problem of over-complexity has already been shown in the previous two sections, but it demonstrates the importance of good documentation, especially if the LC hopes to move away from specifying the Range and Domain of its properties.¹²⁵ The Bibframe 2.0 vocabulary has once more condensed the number of title properties and converted ‘types’ of title, such as variant titles, from property to class. There is now a single ‘title’ property defined which links either the Work or the Instance to the class ‘Title’. This ‘Title’ class has several subclasses, such as ‘VariantTitle’, as well a number of properties that describe its parts, such as ‘subtitle’.¹²⁶ The benefit of this change is that Work and Instance titles can now be more consistently described as the same process is used for both. However, as has been mentioned previously, this approach means that the Domain and Range of ‘title’ cannot be specified and shows the LC is relying on informative instructions to help external

¹²² Drummond (2015) pp. 30-1

¹²³ *Id.* p. 31

¹²⁴ Bibframe <http://bibframe.org/vocab-list/>

¹²⁵ Library of Congress (March 2017b) p. 4

¹²⁶ *Id.* p. 2

cataloguers. Again, this is troublesome, as the subtitle property is spelled ‘subtitle’¹²⁷ in the vocabulary, but ‘subTitle’ in the *Title Notes*.¹²⁸

Another issue that Drummond found when creating Bibframe 1 records was that there were no properties for describing the ‘format and content of the book’.¹²⁹ This refers to the MARC 245 \$c field, or the statement of responsibility, which transcribes the information after the title and subtitle on the title page of a book. The Bibframe 1 vocabulary does now support this information, using the property ‘label’ to include the contents of the 245 field as a text string under the Authority class.¹³⁰ The conversion of this information into a textual string is not an effective way of utilizing the data, regardless of whether this property was added before or after the observation made by Drummond. For the information contained within the 245 \$c field to be actionable, Bibframe 1 relies on the data being displayed elsewhere, such as in an added entry or a bibliographical note. However, this is often where the relationship between contributors and the resource are identified¹³¹ and, without recording this semantically, these links will be wasted. Bibframe 2.0 does provide a specific ‘responsibilityStatement’ property, linking the Instance to the isolated literal string of the 245 \$c field, but this again does not allow the data to be used fully. Unfortunately, resolving this

¹²⁷ Library of Congress (2017a)

¹²⁸ Library of Congress (March 2017b) p. 2

¹²⁹ Drummond (2015) p. 42

¹³⁰ Supplementary spreadsheet, cell F:553

¹³¹ See the record for *Pictures in Verse*, where the illustrator Maud Stumm is recorded as an added entry with no further information:

<https://catalog.loc.gov/vwebv/staffView?searchId=17472&recPointer=0&recCount=25&searchType=1&bibId=7339140>

issue may not be possible using batch conversions, since the difficulty of turning text strings into actionable data is one of the reasons MARC is being left behind in the first place. While Bibframe 2.0 shows little improvement in the criticism levied by Drummond,¹³² it is still no worse at displaying title page information than MARC was and the data itself will remain intact, even if it is not machine-readable.

3.5 The Treatment of Notes in Annotation and Item Classes

The removal of the Annotation class and the introduction of the Item class is widely regarded to be the biggest improvement in Bibframe 2.0. Kroeger,¹³³ Sanderson¹³⁴ and McGrath¹³⁵ all describe the Bibframe 1 use of Annotation as indirect and convoluted. To describe copy-specific data, such as the shelf mark, the class of 'HeldItem' is used, which is a subclass of 'HeldMaterial', and this is a subclass of Annotation.¹³⁶ The class 'HeldItem' can then have the property of 'shelfMarkLcc'.¹³⁷ The criticisms levied against this method warn that there are too many links between the Instance and the 'HeldItem' description, as a result of the subclasses, yet in practice this may not be the case. The output from the Bibframe 1 converter tool displays the class 'HeldItem' with its own URI, with the property of 'holdingFor' linking it directly to the Instance. The super-classes of 'HeldItem' are not shown, suggesting that the impact of these links would be mostly superficial. Fig. 1 below shows how

¹³² Drummond (2015) p. 42

¹³³ Kroeger (2016) p. 15; (July 2017)

¹³⁴ Sanderson (2015) slides 8-10

¹³⁵ McGrath (2017) pp. 9-10

¹³⁶ Supplementary spreadsheet, cell G:354

¹³⁷ *Id.* cell F:354

the Bibframe 1 transformation tool¹³⁸ and the Bibframe 2.0 comparison tool¹³⁹ output the shelf mark of the same bibliographic record. Other than refinements to the vocabulary and the fact that Bibframe 2.0 has added a source and a class for 'ShelfMarkLcc', the structure remains largely the same. This demonstrates that the Bibframe 1 class of 'HeldItem' may not have been as big a problem as originally thought. However, it also highlights the importance of clarity and simplicity in vocabulary structure, since creating such widespread criticism will seriously impact whether Bibframe can be successfully implemented in future.

```
<http://bibframe.org/resources/sGm1503876258/15741651helditem38> a bf:HeldItem ;
  bf:holdingFor <http://bibframe.org/resources/sGm1503876258/15741651instance18> ;
  bf:label "PA4037 .A5955 2009" ;
  bf:shelfMarkLcc "PA4037 .A5955 2009" .
```

```
<http://bibframe.example.org/15741651#Item050-15> a bf:Item ;
  bf:itemOf <http://bibframe.example.org/15741651#Instance> ;
  bf:shelfMark [ a bf:ShelfMarkLcc ;
    rdfs:label "PA4037.A5955 2009" ;
    bf:source <http://id.loc.gov/vocabulary/organizations/dlc> ] .
```

Fig. 1. Comparison of Bibframe 1 (above) and Bibframe 2.0 (below) output for a shelf mark.

For other aspects of the Bibframe 1 Annotation class, such as general notes and summaries, the Bibframe 2.0 vocabulary has separated them completely from copy-specific information. The spreadsheet compiled in the appendix found only twelve occurrences of the Annotation class,¹⁴⁰ while the vocabulary for Bibframe 1 lists thirty-six properties with the Domain of

¹³⁸ Bibframe <http://bibframe.org/tools/transform/start>

¹³⁹ Library of Congress <http://id.loc.gov/tools/bibframe/compare-liccn/full-ttl?find=2009020160>

¹⁴⁰ Supplementary spreadsheet, column E

Annotation or any of its subclasses (such as 'CoverArt', 'HeldMaterial', 'Review', 'Summary' and 'Annotates').¹⁴¹ When the total number of properties is considered, this means that Annotation properties make up roughly 12% of the vocabulary. Meanwhile for Bibframe 2.0, thirteen specific uses of the Item class were found, despite the much more complete set of specifications.¹⁴² In the Bibframe 2.0 vocabulary, only eight properties are specifically linked to the Item class, which equals just fewer than 6% of the total properties.¹⁴³ The reason for this decrease in properties is caused by reducing the Item class to only copy-specific information while general information, such as the summary, is linked to either the Work or Instance.¹⁴⁴ The benefit of this is that the structure is clearer, since now notes can be linked directly to the level of resource they are describing, rather than notes of different levels of specificity being attached to different Annotations. The removal of the Annotation class simplifies Bibframe 2.0 by taking away an unnecessary layer of abstraction. The elevation of Item to a core class does not have many practical implications, but rather it reflects the importance of the physical bibliographical item to the library community on a conceptual level.

3.6 Conclusion

Overall, Bibframe 2.0 has been consistently simplified and refined from the first iteration of Bibframe. The change of the representation of 'type' from property to class, described in

¹⁴¹ Bibframe <http://bibframe.org/vocab-list/>

¹⁴² Supplementary spreadsheet, Column H

¹⁴³ Library of Congress (2017a)

¹⁴⁴ Supplementary spreadsheet, cell H:822

section 3.2, allows for greater interoperability with external ontologies. By reducing the number of properties and limiting their restrictions, Bibframe 2.0 has become more flexible while making its vocabulary and easier to implement. Bibframe 1 often provided multiple ways to express the same information, such as with title fields and Authorities, while Bibframe 2.0 has refined these descriptions and reduced the possibilities for inconsistency. However, the downside of increasing the flexibility of Bibframe is that the use of vocabulary can often be unclear. Bibframe 2.0 has chosen not to use 'Domain' or 'Range' to specify the use of its properties, choosing instead to rely on clear instruction and documentation to guide cataloguers in appropriate use. The Bibframe primer document¹⁴⁵ states that the LC hopes the Bibframe model will open library resources up to the Web to allow information users access to good quality data. If this is the case, it is imperative that the data entered into these records is consistent and accurate, because something as simple as attributing a variant title to a Work could spread misinformation across the Web. It is therefore vital that the LC provides clear instructions for implementing the Bibframe 2.0 vocabulary if it is to be successful. The progress of these instructions is evaluated in chapter 4.

¹⁴⁵ Library of Congress (2012) p. 5

4. The Current Status of the Bibframe 2.0 Conversion Specifications

This chapter reviews the current state of the LC's conversion specifications. The aim is to assess how well the conversion specifications provided by the LC¹⁴⁶ are able to convert MARC data into Bibframe 2.0 records. The spreadsheet provided as supplementary content combines all the relevant information from these conversion specifications, previously contained in 21 separate documents, and presents the data alongside the relevant MARC fields. It should be understood that these specifications are still under development, with the most recent updates posted in late July 2017,¹⁴⁷ and so this has been taken into consideration when evaluating the progress. However, these specifications are being used in the Bibframe 2.0 pilot,¹⁴⁸ so therefore they should be considered worthy of careful scrutiny if they are already being used in practical application. Section 4.1 of this chapter reviews the issues found within these documents, focusing on any mistakes or areas that have been poorly explained. The implications of these mistakes are considered and improvements are suggested. Next, section 4.2 evaluates the conversion itself, assessing how well MARC data is preserved in the transition to Bibframe 2.0. This looks at any areas where data is lost and considers potential alternatives. Section 4.3 looks at a number of occasions where the conversion specifications refer to non-existent external documentation. On these occasions,

¹⁴⁶ Library of Congress (2017b)

¹⁴⁷ *Ibid.*

¹⁴⁸ *Ibid.*

the communication of the Bibframe 2.0 mapping breaks down completely, often making it impossible to follow. This section looks at the LC's approach to retaining, updating and disseminating documentation. If Bibframe 2.0 is to be successful, it needs to be implemented by institutions using the instructions provided by the LC, and so it is important to assess the provision of this information. Finally, Section 4.4 evaluates areas that the LC has made 'no effort to convert'¹⁴⁹ MARC fields, both in terms of the validity of these judgements and how these judgements influence potential implementers of Bibframe 2.0. This section considers whether the LC has been too introspective with its conversion specifications and whether this will prevent other institutions from implementing Bibframe 2.0.

4.1 Consistency within the Conversion Specifications

The MARC to Bibframe conversion specifications provided by the LC contains a huge amount of data. The 21 documents detailing every MARC field and its relevant Bibframe conversion has taken up 1,830 rows after being aggregated in the supplementary spreadsheet. Naturally, the risk of providing so much manually created data is that human error can lead to mistakes. As has been mentioned in section 3.2 of this dissertation, the LC intends to instruct external cataloguers through the use of good documentation.¹⁵⁰ The current conversion specifications show that, so far, this approach does not seem realistic. Firstly, there are several occasions where incorrect information has been displayed. There are two cases in the conversion of the MARC 008 field where the wrong URI has been provided, with one using the wrong URI and

¹⁴⁹ *Ibid.*

¹⁵⁰ Library of Congress (March 2017a) p. 4

the other containing a typo.¹⁵¹ The latter case is easily noticeable, since the URI does not work, meaning that the record would just contain a dead link with the correct label. This could be resolved by simply adding the correct URI once the problem has been discovered. However, the former error provides URI for 'abstract'¹⁵² in place of the URI for 'dictionary'.¹⁵³ Therefore, the link provided is valid and, consequently, information users could be given incorrect search results or they could be pointed in the wrong direction. Furthermore, there were two occasion in the MARC 086 field where 'rdf:value'¹⁵⁴ and 'rdfs:label'¹⁵⁵ were used incorrectly. The RDF conventions supplied by the LC do not specify how to use 'rdf:value' and 'rdfs:label',¹⁵⁶ but generally the former is used for numerical values and the latter is used for text strings.¹⁵⁷ Mistakes like this can lead to inconsistency in the bibliographic records and will create problems in the future. Without the ability to define all values as numbers, or all 086 \$a fields as an 'rdf:value',¹⁵⁸ the power to manipulate the data in future is diminished. However, both these errors came from documents that are yet to undergo the first round of revisions.¹⁵⁹ This means that the LC still has time to fix these errors, although it remains to be seen if they will.

¹⁵¹ Supplementary spreadsheet, cell K:94 and cell K:115

¹⁵² Supplementary spreadsheet, cell K:91

¹⁵³ *Id.* K:94

¹⁵⁴ *Id.* K:427

¹⁵⁵ *Id.* K:429

¹⁵⁶ Library of Congress (March 2017a) p. 3

¹⁵⁷ Supplementary spreadsheet, cell K:394 and cell K:435. These cells show how the 'rdf:value' and 'rdfs:label' are treated elsewhere.

¹⁵⁸ Supplementary spreadsheet, row 427

¹⁵⁹ Library of Congress (2017b)

Another issue that has been found with the conversion specifications is that the LC often struggles to provide enough detail when describing the conversions. The worst example of this is for the MARC 600 – 630 fields.¹⁶⁰ Here, the specification points to a document that describes the treatment of all added entry fields, which includes those from the MARC 100, 110, 700 and 800 fields.¹⁶¹ However, while the document entitle ‘Process notes’ describes in detail some examples of added names and uniform titles, there is no reference to the MARC 6XX fields or subject added entries other than in a brief description of relationships. It is unclear whether the subject added entry for agents attaches to the Work or Instance and, while it is assumed it attaches to the Work like the MARC 650 field,¹⁶² this creates the potential for inconsistent cataloguing. Furthermore, the class of ‘Subject’ is not part of the Bibframe 2.0 vocabulary,¹⁶³ meaning it is very difficult to interpret how the property ‘subject’ is used and what its value should be. Clearly, this poor standard of documentation is more than just the result of a lack of revision. Subject is one of the ‘key concepts’ in the Bibframe 2.0 model,¹⁶⁴ yet the class has been overlooked from the vocabulary and MARC 600 – 630 fields have not been given individual attention. The cause of this is likely due to the overly introspective approach taken by the LC, a theme that is discussed again in section 4.4. The LC has not made this documentation suitable for external use and this is evident by the level of prior knowledge required to convert even a key concept of Bibframe 2.0.

¹⁶⁰ Supplementary spreadsheet, row 1023 - 1041

¹⁶¹ Library of Congress (2017b)

¹⁶² Supplementary spreadsheet, cell H:1056

¹⁶³ Library of Congress (2017a)

¹⁶⁴ Library of Congress (April 2016)

4.2 Preservation of MARC Data

Outside of the errors mentioned in section 4.1, the language of the MARC to Bibframe conversion is generally quite simple. The properties are often broad since, as described in section 3.2, type is now represented by class.¹⁶⁵ The benefit of this is that properties are clearly defined, since they describe simple concepts such as ‘agent’, ‘title’ and ‘identifiedBy’, meaning there is little overlap between them. This should reduce inconsistency as, even with little explanation, it is often obvious which property to use when describing a resource based on its name alone. Class, too, is often well defined, reflecting either the property name or the name of the MARC field in many cases. An example of this is seen in the title fields, where ‘CollectiveTitle’ is used for the MARC collective uniform title field,¹⁶⁶ ‘VariantTitle’ is used for the varying title field,¹⁶⁷ and ‘AbbreviatedTitle’ is for the abbreviated title field.¹⁶⁸ However, while the simplicity allows for more consistent cataloguing, it does mean that some of the specifics of MARC data will be lost. Examples of this loss of data can be seen early on in the conversion specifications. In the MARC 008 field, ‘questionable dates’¹⁶⁹ and ‘probable dates’¹⁷⁰ are recorded in the same way as confirmed dates, with the property of ‘provisionActivity’ and the class of ‘Publication’. This means that, after the conversion, it will be impossible to ascertain whether the dates taken from the 008 field are firm, questionable, or probable. Similarly, many MARC fields have been ignored or generalised to reduce the

¹⁶⁵ Library of Congress (March 2017a) p. 2

¹⁶⁶ Supplementary spreadsheet, row 538

¹⁶⁷ *Id.* row 567

¹⁶⁸ *Id.* row 494

¹⁶⁹ *Id.* row 33

¹⁷⁰ *Id.* row 38

vocabulary needed to describe the records. There are 243 cells marked as 'To be ignored' in the conversion specification,¹⁷¹ which does not take into account occasions where MARC subfields have been contracted, and this demonstrates a substantial loss of data. Furthermore, certain MARC fields have been incorporated generally, meaning the information is retained without necessarily maintaining the context. An example of this can be seen in the MARC 547 field, where the former title complexity note is recorded simply with the property 'note' and the class 'Note'.¹⁷² The same is true for the 'information about documentation note' in the MARC 556 field¹⁷³ and the 'linking entry complexity note' in the MARC 580 field.¹⁷⁴ In all three cases, these notes are attached directly to the Work or Instance, meaning that the relationship between the note contents and the resource can be specified no further than 'note'. Once again, specificity has been sacrificed in place of simplicity.

4.3 The Structure of the Specification

This section focuses on how the LC has chosen to disseminate its conversion specifications and the issues that have been caused by this method. Specifically, the use of several individually published documents that often seem to be updated in isolation. As mentioned in section 4.1, the conversion specification page contains 21 separate documents, each at

¹⁷¹ *Id.* column K

¹⁷² *Id.* row 909

¹⁷³ *Id.* row 924

¹⁷⁴ *Id.* row 945

various stages of revision.¹⁷⁵ The spreadsheets generally contain guidelines for the simple conversions, while two word documents contain complex instructions for converting data or repetitive fields.¹⁷⁶ Presumably, this approach has been taken to save time, since MARC fields can be repetitive. However, as mentioned in section 4.1, this means that often the instructions are simplified to allow one description to be used for several MARC fields, such as the multiple title fields. Furthermore, the spreadsheets have fields that point to the more complicated word documents, despite the fact that these documents are updated separately. An example of this can be seen in the document entitled 'Process notes'.¹⁷⁷ This document, which mostly describes the conversion of name and title authorities, has been updated twice at July 27. Meanwhile, the spreadsheet document that treats name fields (1XX, 6XX, 7XX etc) has not been updated since March and is still in its first draft. As a result, the spreadsheet refers to 'process 1.5' of the 'Process notes' on eight separate occasions,¹⁷⁸ yet process 1.5 does not exist in the 'Process notes' document. This means it is impossible to follow the conversion of these MARC fields. Again, it is understood that these conversion specifications are under development, but the problem is caused by how the LC has chosen to display the information. By dividing the MARC fields across so many documents, this means that each spreadsheet would need to be checked every time a change is made, otherwise further issues will continue to develop. The supplementary spreadsheet demonstrates that all of the

¹⁷⁵ Library of Congress (2017b)

¹⁷⁶ *Ibid.* The third document labelled as 'word' is actually a spreadsheet.

¹⁷⁷ *Ibid.*

¹⁷⁸ Supplementary spreadsheet, cell K:441, K:453 K:1005, K:1042, K:1149, K:1164, K:1706, K:1720

data can be compiled into one spreadsheet. While this is a daunting amount of information, at least the entire specification can be searched, ordered and viewed at the same time. A better option might be for the LC to host these conversion specifications on a Webpage, thereby allowing for descriptions to be given in full, while also actually linking to separate documents if need be. This is recommended if the LC hopes to allow others to use these specifications, even if they are still under development.

The LC hopes to promote consistent cataloguing based on the merit of its documentation.¹⁷⁹ However, the issue of poor documentation is something that has been touched on at several points in this dissertation. Section 2.4 described how Drummond¹⁸⁰ found difficulty interpreting the Bibframe 1 vocabulary. Sections 3.2 and 3.4 highlighted areas of carelessness in the documentation and section 4.1 demonstrated a fault in the Bibframe 2.0 vocabulary. On top of this, it is worth considering the reliability of the LC in updating and maintaining this documentation. Since the announcement of Bibframe 2.0 in 2015, there has been very little information posted on the LC website. A series of 'Examples and Notes' were posted in March 2016¹⁸¹ which provide a good explanation of some of the key concepts used in Bibframe 2.0. Next, a brief overview of the model, published in April 2016, gave a short summary of how the Bibframe 2.0 focus had shifted.¹⁸² Finally, a webcast was published in

¹⁷⁹ Library of Congress (March 2017a) p. 4

¹⁸⁰ Drummond (2015) p. 50

¹⁸¹ Library of Congress <https://www.loc.gov/bibframe/docs/index.html>

¹⁸² Library of Congress (April 2016)

September 2016 that explained the reasons behind moving to Bibframe 2.0.¹⁸³ However, since then updates are quietly concentrated on the conversion specifications and the vocabulary. The ‘What’s New’ sidebar on the Bibframe homepage lists a forum held in June 2017 (which links to four presentations with little new value), the updated specifications, and the Bibframe 1 pilot from June 2016.¹⁸⁴ What this shows is a lack of narrative on the progress of Bibframe 2.0, making it difficult for external institutions to follow. It poses a risk for potential adopters to begin experimenting on Bibframe 2.0 when it is unclear where the model is heading and how much it is likely to change in the future. As Edmunds¹⁸⁵ states, Bibframe will not become successful just by being the best model, it also needs to be widely adopted. This can only happen if the LC is able to convince the library community that it is worth committing to Bibframe 2.0. Such a task will be difficult when the updates and progress of Bibframe 2.0 remain hidden from the library community. Trust and understanding will be better gained through consistent updates and transparency, such as with a blog or social media. By keeping the community informed, the LC will have a much better chance of convincing other institutions to migrate to Bibframe 2.0, as well as providing a more dynamic forum for community feedback. Otherwise, external institutions will have to commit to a model that seems to be updated slowly, without warning, and often without spellchecking.

¹⁸³ Library of Congress (September 2016)

¹⁸⁴ Library of Congress <https://www.loc.gov/bibframe/>

¹⁸⁵ Edmunds (2017) p. 7

4.4 Introspection and the Library of Congress

Finally, this section evaluates the approach taken by the LC to ignore certain MARC fields purposefully. The supplementary spreadsheet documents the MARC fields that the LC has chosen to ignore by using the term 'nac', or No Attempt to Convert. The use of 'nac' has been qualified in the conversion specifications as for when 'there is little or no use of an element in the LC records', continuing with 'others may want to augment the specifications'.¹⁸⁶ This suggests that the LC does not plan to provide a conversion for these records, as they advise other institutions to attempt their own conversions if they believe these fields are necessary. There are 152 MARC fields that have not been converted by the LC, without counting subfields that have been contracted.¹⁸⁷ One example of a field that has been ignored is the first indicator of the 541 MARC fields,¹⁸⁸ which represents whether the acquisition source information is displayed privately or publically. Clearly, information that has been labelled 'sensitive or restricted'¹⁸⁹ should remain private, especially now that Bibframe records have the potential to be linked to across the Web. The fact that the LC still provided a conversion for this field is an oversight and an issue that will certainly need to be resolved before the final release of Bibframe 2.0. Another example of an ignored field that deserves more attention is the MARC 080 field. This field contains the Universal Decimal Classification (UDC) number, a classification scheme used in over 130 countries,¹⁹⁰ for which a great

¹⁸⁶ Library of Congress (2017b)

¹⁸⁷ Supplementary spreadsheet, column K

¹⁸⁸ *Id.* row 886

¹⁸⁹ Library of Congress (2005)

¹⁹⁰ UDC (2017a)

number of institutions will use to classify their holdings. Before these libraries can even consider adopting Bibframe 2.0, a conversion will need to be made for this field, otherwise they risk losing the ability to display the class mark of their bibliographic records online. This could be a major factor in whether certain libraries decide to commit to Bibframe 2.0 or not. The LC risks alienating a huge number of libraries by ignoring a major aspect of bibliographical data, with over 140,000 institutions using UDC in Europe alone.¹⁹¹ This demonstrates that the LC is guilty of taking an introspective approach to Bibframe 2.0, focusing more on its own benefits rather than on those of the library community as a whole, and this could be detrimental to the success of Bibframe 2.0. For external libraries to move on from MARC they will need to see evidence that Bibframe 2.0 is the best option for the future. Currently, for libraries that use UDC, Bibframe 2.0 is not an option.

4.5 Summary of Findings

This chapter has provided harsh criticism of Bibframe 2.0 in the hope that it can be refined and improved in the future. The evaluation of the conversion specifications shows that, while there is still work to be done to eliminate human error, the conversion process itself works well. This supports the conclusion from the first chapter that the Bibframe model seems well thought out and functional. However, the LC does need to improve how its dissemination of information, as the documentation available is often in the form of a rough draft. With this improvement, external libraries will be able to better prepare for the future and, when Bibframe is finally released, it is more likely that they will be ready to accept the transition.

¹⁹¹ UDC (2017b)

Furthermore, the LC should take into account the needs of libraries from all over the world, rather than just focusing on the MARC fields that it uses. For Bibframe 2.0 to be successful, it must be widely adopted, and this can only happen if the needs of all libraries are taken into account. Otherwise, if global libraries use inconsistent data models, the potential for exploiting information worldwide is vastly diminished. It is recommended that the LC establish a blog to keep the library community up to date with the progress of Bibframe 2.0. The conversion specifications will need to be displayed in a more versatile format if they are to be widely adopted. It is also advised that the LC should focus Bibframe 2.0 globally in an attempt to provide a conversion for fields used by external institutions.

5. Conclusion

The question that this dissertation set out to answer is this: How has Bibframe 2.0 changed from the first iteration of Bibframe and what benefits, if any, does it offer to bibliographic description over the previous model? Chapter 2 reviewed the literature describing the conception of Bibframe and the issues with the vocabulary and implementation of the first iteration. This chapter showed that the vocabulary and model were generally successful. However, Drummond found flaws in the complexity of the vocabulary, and the practical tests were not of a scale large enough to test the first iteration of Bibframe fully. From this review, it was clear that vocabulary complexity and clarity were the main issues affecting the success of Bibframe 1. Chapter 3 compared the vocabulary from Bibframe 1 with Bibframe 2.0, using the results from the first chapter to focus specifically on the changes to properties and class in section 3.2, Authorities in section 3.3 and Item in section 3.5. This chapter concluded that Bibframe 2.0 had indeed been simplified and refined. Interestingly, it was the change to the representation of 'type' from property to class that had the biggest impact on the vocabulary, not the change in core classes. This change can be seen as a definite positive, although it brought with it a new issue: a heavy reliance on accurate documentation. The flexibility of Bibframe 2.0 came with a loss of specificity, something the LC acknowledged. Their response to this problem was to replace the specificity lost in the vocabulary with clear and concise instructions for cataloguers. Chapter 4 then looked at the vocabulary of Bibframe 2.0 and evaluated the conversion specifications provided by the LC. The previous chapter made it clear that good documentation was needed for the successful implantation of Bibframe 2.0.

Unfortunately, chapter 4 showed that the conversion specifications still need some work, with isolated updates causing the separate documents to become disjointed. These mistakes should be fixed with future revisions, although the approach of the LC in general to posting updates and providing information needs to be reviewed. It has been two years since the release of Bibframe 2.0, which should have provided enough time for a more serious description of Bibframe and its vocabulary. Finally, it would appear that the scope of the current Bibframe 2.0 vocabulary is to enrich the records held at the LC, but it would certainly benefit both Bibframe and the library community if the vocabulary was updated to address the needs of all MARC records on a more global scale. Overall, Bibframe 2.0 is a definite improvement over its predecessor, but it still needs some development. The hard work is done, the LC now needs to listen to the needs of the library of community and work with other institutions to achieve a new bibliographic standard that can be accepted worldwide.

In future, studies will need to monitor the documentation provided by the LC for signs of improvement. The scope of this dissertation only covered bibliographic records, and a further study on audiovisual material would be beneficial. Due to the nature of the bibliographic records, the introduction of the Event class was not covered, and this certainly deserves the attention of future work. The supplementary spreadsheet will be useful for those wanting to assess the conversion specifications, although it could be improved with the addition of more Bibframe 1 vocabulary.

Appendix

Supplementary Spreadsheet Format

Columns A-D:	MARC Fields and Subfields
Columns E-G:	Bibframe 1 Vocabulary
Columns H-J:	Bibframe 2.0 Vocabulary
Column K:	Comments provided by the Library of Congress
Column L:	Issues in the Conversion Specification

Supplementary Spreadsheet Example Page

MARC field	Ind 1.1	Ind 1.2	Subfield	Work (W) / Instance (I) / Annotation / Authority	BIBFRAME 1.0 Property	BIBFRAME 1.0 Class	Work (W) / Instance (I) / Item	BIBFRAME 2.0 Property	BIBFRAME 2.0 Class	Comment	Issues
001				adminMetadata	adminMetadata	AdminMetadata	W	adminMetadata	AdminMetadata	Control number	
				source	source	Local rdf:resource		source	Source	Local rdf:value	
				adminMetadata	adminMetadata	AdminMetadata	W	adminMetadata	AdminMetadata	Control number identifier	
				source	source	Source		source	Source	Control number identifier	
				code	code	Literal		code	Literal	Literal	
005				Annotation	changedDate	AdminMetadata	W	adminMetadata	AdminMetadata	Literal - YYYY-MM-DD-Thrmm:ss [Eg. 1995-07-17T17:30:00 for 5:30pm on July 17 1995]	
006 - 00				nac	nac	nac		nac	nac	For textual material the first character will either be 'a' (books) or 'r' (manuscript language material)	
006 - 01-17				same as 008 - 18-34	same as 008 - 18-34	Fixed string characters, for characters 01-17 refer to 008 characters 18-34		same as 008 - 18-34	Physical Description - 00 for category of material (text), 01 for specific material designation (a=regular print, b=large print, c=braille, d=looseleaf, u=unspecified, z=other, no attempt)		
007 - 00/01				nac	nac	Date record was created - Literal - ccy-rm-dd (MARC format is only yymmdd so infer year, Eg. 801102 = 1980-11-02, 051102 = 2005-11-02)		nac	Date record was created - Literal - ccy-rm-dd (MARC format is only yymmdd so infer year, Eg. 801102 = 1980-11-02, 051102 = 2005-11-02)		
008 - 00-05				creationDate	creationDate	AdminMetadata	W	creationDate	AdminMetadata	Type of date	
008 - 06						No date/B/C date for publication etc - To be ignored					
-b				provisionalActivity	provisionalActivity	Continuing resource currently published	I	provisionalActivity	Continuing resource currently published	Use editf (example in 005) at characters 07-10 for date of first publication - Convert all 'V' to 'X'	
-c				date	date	Publication	I	date	Publication	Note rdfs:label "Currently published"	
-d				note	note	Note rdfs:label "Currently published"	I	note	Publication	contribution resource ceased publication	
-e				date	date	Publication same as above, with last publication at characters 11-14	I	date	Publication	same as above, with last publication at characters 11-14	
-f				note	note	Note rdfs:label "Ceased publication"	I	note	Publication	Detailed date - same as above, format YYYYmmdd in characters 07-14, or YYYYmm in characters 07-12 if no day provided.	
-g				provisionalActivity	provisionalActivity	Inclusive dates of collection	I	provisionalActivity	Production	Same as above, first date 07-10, second date 11-14	
-h				note	note	Note "Inclusive collection dates"	I	note	Production	Same as above, first date 07-10, second date 11-14	
-i				note	note	Note Exactly the same as "i" field, with note "bulk collection dates"	I	note	Production	Exactly the same as "i" field, with note "bulk collection dates"	
-j				note	note	Note Exactly the same as "i" field, with note "bulk collection dates"	I	note	Production	Exactly the same as "i" field, with note "bulk collection dates"	
-k				note	note	Note Exactly the same as "i" field, with note "bulk collection dates"	I	note	Production	Exactly the same as "i" field, with note "bulk collection dates"	
-l				note	note	Note Exactly the same as "i" field, with note "bulk collection dates"	I	note	Production	Exactly the same as "i" field, with note "bulk collection dates"	
-m				provisionalActivity	provisionalActivity	Multiple dates	I	provisionalActivity	Publication	Same as above	
-n				provisionalActivity	provisionalActivity	Dates unknown - To be ignored	I	provisionalActivity	Publication	Dates unknown - To be ignored	
-o				provisionalActivity	provisionalActivity	Date of distribution/release/issue AND production/recording session when different - 1	I	provisionalActivity	Production	Date of distribution/release/issue AND production/recording session when different - 1	
-p				provisionalActivity	provisionalActivity	Distribution	I	provisionalActivity	Production	Date of production at characters 07-10	
-q				provisionalActivity	provisionalActivity	Production	I	provisionalActivity	Production	Date of production at characters 11-14	
-r				date	date	Questionable date - Same as other publication fields	I	date	Publication	Questionable date - Same as other publication fields	
-s				provisionalActivity	provisionalActivity	Reprint/Reissue date	I	provisionalActivity	Publication	Reprint/Reissue date	
-t				provisionalActivity	provisionalActivity	Record 07-10 date, ignore second date	I	provisionalActivity	Publication	Record 07-10 date, ignore second date	I - Ignoring original date?
-u				provisionalActivity	provisionalActivity	Single known date/probable date	I	provisionalActivity	Publication	Single known date/probable date	I - Recording probable dates as true dates
008 - 07-14				provisionalActivity	provisionalActivity	Date at 07-10	I	provisionalActivity	Publication	Date at 07-10	
008 - 15-17				provisionalActivity	provisionalActivity	Publication date and copyright date	I	provisionalActivity	Publication	Publication date and copyright date	
				copyrightDate	copyrightDate	Publication	I	copyrightDate	Publication	Publication date and copyright date	
				provisionalActivity	provisionalActivity	AT 11-14	I	provisionalActivity	Publication	AT 11-14	
				date	date	Same as other Publication fields	I	date	Publication	Same as other Publication fields	
				provisionalActivity	provisionalActivity	Literal conversion following above standards	I	provisionalActivity	Publication	Literal conversion following above standards	
				provisionalActivity	provisionalActivity	Place of publication, production or execution	I	provisionalActivity	Publication	Place of publication, production or execution	I - Is production different to publication here?
008 - 35-37				place	place	Convert location code to URI from http://id.loc.gov/vocabulary/locations	W	place	Convert location code to URI from http://id.loc.gov/vocabulary/locations		
008 - 38				language	language	Language - W - Convert language code to URI from http://id.loc.gov/vocabulary/languages	W	language	Language - W - Convert language code to URI from http://id.loc.gov/vocabulary/languages		
008 - 39				nac	nac	Modified record - To be ignored	I	nac	Modified record - To be ignored		
008 - 18-21				illustrativeContent	illustrativeContent	BOOK SPECIFIC - Illustrations	I	illustrativeContent	BOOK SPECIFIC - Illustrations	Cataloguing source - nac	
-#				illustrativeContent	illustrativeContent	No illustrations - To be ignored	I	illustrativeContent	BOOK SPECIFIC - Illustrations	No illustrations - To be ignored	
-a				illustrativeContent	illustrativeContent	## - http://id.loc.gov/vocabulary/mlil/us/ill - Add rdfs:label "Illustrations"	I	illustrativeContent	BOOK SPECIFIC - Illustrations	## - http://id.loc.gov/vocabulary/mlil/us/ill - Add rdfs:label "Illustrations"	
-b				illustrativeContent	illustrativeContent	## - http://id.loc.gov/vocabulary/mlil/us/map - Add rdfs:label "maps"	I	illustrativeContent	BOOK SPECIFIC - Illustrations	## - http://id.loc.gov/vocabulary/mlil/us/map - Add rdfs:label "maps"	
-c				illustrativeContent	illustrativeContent	## - http://id.loc.gov/vocabulary/mlil/us/por - Add rdfs:label "portraits"	I	illustrativeContent	BOOK SPECIFIC - Illustrations	## - http://id.loc.gov/vocabulary/mlil/us/por - Add rdfs:label "portraits"	
-d				illustrativeContent	illustrativeContent	## - http://id.loc.gov/vocabulary/mlil/us/chr - Add rdfs:label "charts"	I	illustrativeContent	BOOK SPECIFIC - Illustrations	## - http://id.loc.gov/vocabulary/mlil/us/chr - Add rdfs:label "charts"	

Bibliography

Alemu, G.; Stevens, B.; Ross, P. & Chandler, J. (2012). "Linked Data for Libraries: Benefits of a Conceptual Shift from Library-Specific Record Structures to RDF-Based Data Models." *New Library World* [online]. 113(11/12), 549-570. Doi: <https://doi.org/10.1108/03074801211282920> [Accessed 11 August 2017]

Allemnag, D. & Hendler, J. (2008). *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Amsterdam: Morgan Kaufmann.

Antoniou, Grogoris; Groth, Paul; van Harmelen; Frank & Hoekstra, Rinke (2012) *A Semantic Web Primer*. Third ed. London: The MIT Press.

Baker, T.; Coyle, K. & Petiya, S. (2014) "Multi-Entity Models of Resource Description in the Semantic Web: A Comparison of FRBR, RDA and Bibframe." *Library Hi Tech* [online]. 32(4), 562-582. Doi: <https://doi.org/10.1108/LHT-08-2014-0081> [Accessed 28 August 2017].

van Ballegoie, M. & Borie, J. (2014) "From Record-Bound to Boundless: FRBR, Linked Data, and New Possibilities for Serials Cataloguing." *The Serials Librarian* [online]. 66(1-4), 76-87. Doi: <http://dx.doi.org/10.1080/0361526X.2014.879527> [Accessed 28 August 2017].

BIBFLOW (2016a) "BIBFLOW: About" [online]. Available from: <https://bibflow.library.ucdavis.edu/about/> [Accessed 3 August 2017].

BIBFLOW (2016b) "It's Complicated" [online]. Available from: <https://bibflow.library.ucdavis.edu/its-complicated/> [Accessed 28 August 2017].

Bibframe "MARC to BIBFRAME Transformation Service" [online]. Available from: <http://bibframe.org/tools/transform/start> [Accessed 28 August 2017].

Bibframe "BIBFRAME Vocabulary" [online]. Available from: <http://bibframe.org/vocab-list> [Accessed 28 August 2017].

Bibframe (2013) "BIBFRAME Test Case" [online]. Available from: <http://bibframe.org/tools/tests/AnnotationHolding.html> [Accessed 2 August 2017].

Bibliothèque nationale de France "Catalogue General" [online]. Available from: <http://catalogue.bnf.fr/index.do> [Accessed 22 August 2017].

Calhoun, Karen (2006) *The Changing Nature of the Catalog and its Integration with Other Discovery Tools* [online]. Ithica, NY: Cornell University Library. Available from: <https://www.loc.gov/catdir/calhoun-report-final.pdf> [Accessed 28 August 2017].

- Coyle, Karen (2004) "Future Considerations: The Functional Library Systems Record." *Library Hi Tech* [online]. 22(2), 166-174. Doi: <https://doi.org/10.1108/07378830410524594> [Accessed 23 August 2017].
- Coyle, Karen (2010) "Library Data in a Modern Context." *Library Technology Reports*, 46(1), 5-13.
- Coyle, K. & Hillman, D. (2007) "Resource Description and Access (RDA): Cataloguing Rules for the 20th Century." *D-Lib Magazine*, 13(1/2).
- Croissant, Charles (2012) "FRBR and RDA: What They Are and How They May Affect the Future of Libraries." *Theological Librarianship* [online]. 5(2), 6-22. Available from: <https://theolib.atla.com/theolib/article/view/234/565> [Accessed 24 August 2017].
- Cyganiak, Richard (2009) "What's in a Name? And the Linked Data Police." *Cygri's Notes on Web Data* [blog]. Available from: <http://richard.cyganiak.de/blog/2009/11/whats-in-a-name-and-the-linked-data-police/> [Accessed 18 August 2017].
- Drummond, Sophie (2015) *Is Bibframe Best? A Comparison of Linked Data Models for Bibliographic Data*. Dissertation (MA), University College London.
- Dunsire, G. (2008) "Semantic Web and Expert Metadata: Pull Apart then Bring Together." Paper presented at the Archives, Libraries, Museums (AKM12) [online]. Available from: <http://eprints.rclis.org/bitstream/10760/7443/1/akm2008semanticweb.pdf> [Accessed 24 July 2017].
- Dunsire, G. (2009) "UNIMARC, RDA and the Semantic Web." *International Cataloguing and Bibliographic Control* [online]. 39(2) Available from: <http://www.ifla.org/files/hq/papers/ifla75/135-dunsire-en.pdf> [Accessed 24 July 2017].
- Dunsire, G. (2012) "Representing the FR Family in the Semantic Web." *Cataloging and Classification Quarterly*, 50(5-7), 1-18.
- Dunsire, G. & Willer, M. (2011) "Standard Library Metadata Models and Structures for the Semantic Web." *Library Hi Tech*, 28(3), 1-12.
- Edmunds, Jeff (2017) "Bibframe as an Empty Vessel" [online]. Available from: <https://scholarsphere.psu.edu/downloads/rbn999824s> [Accessed 23 August 2017].

Illien, Gildas (2013) "Interview with Gildas Illien", *Information Standards Quarterly* [online]. 25(4), 22-29. Available from: http://www.niso.org/apps/group_public/download.php/11938/IP_Illien_BnF_isqv_25no4.pdf [Accessed 13 August 2017].

International Federation of Library Associations (2009) *Functional Requirements for Bibliographic Records* [online]. Amended Edition. Available from: https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf [Accessed 28 August 2017].

Kiegel, Joseph (2015) "BIBFRAME Projects at the University of Washington." Paper presented at the annual meeting for American Library Association, San Francisco, June 25 to June 30, 2015 [online]. Available from: <http://slideplayer.com/slide/5383520/> [Accessed 19 August 2017].

Kroeger, Angela (2013) "The Road to BIBFRAME: The Evolution of the Idea of Bibliographic Transition into a Post-MARC Future." *Cataloging and Classification Quarterly* [online]. Doi: <http://dx.doi.org/10.1080/01639374.2013.823584> [Accessed 8 August 2017].

Kroeger, Angela (2014) "The Progress of BIBFRAME." Presentation given at the OLAC-MOUG 2014 Conference, Kansas City [online]. Available from: <https://www.slideshare.net/akroeger/kroeger-bibframe2014olacmoug> [Accessed 12 August 2017].

Kroeger, Angela (2016) "Hello Bibframe 2.0: Changes from 1.0 and Possible Directions for the Future." *Criss Library Faculty Proceedings & Presentations* [online]. Paper 65. Available from: <http://digitalcommons.unomaha.edu/crisslibfacproc/65> [Accessed 23 August 2017].

Kroeger, Angela (akroeger@unomaha.edu) 31/07/2017. "Bibframe 2.0 Dissertation." Email to (alex.keane.16@ucl.ac.uk).

Library of Congress "What's New in BIBFRAME 2.0" [online]. Available from: <https://www.loc.gov/bibframe/docs/bibframe2-whatsnew.html> [Accessed 28 August 2017].

Library of Congress "BIBFRAME Comparison Tool" [online]. Available from: <http://id.loc.gov/tools/bibframe/compare-lccn/full-ttl?find=2009020160> [Accessed 28 August 2017].

Library of Congress "Pictures in Verse." MARC Bibliographic record for *Pictures in Verse* [online]. Available from:

<https://catalog.loc.gov/vwebv/staffView?searchId=17472&recPointer=0&recCount=25&searchType=1&bibId=7339140> [Accessed 26 August 2017].

Library of Congress “BIBFRAME Model, Vocabulary, Guidelines, Examples, Notes, Analysis” [online]. Available from: <https://www.loc.gov/bibframe/docs/index.html> [Accessed 30 August 2017].

Library of Congress “Bibliographic Framework Initiative” [online]. Available from: <https://www.loc.gov/bibframe/> [Accessed 30 August 2017].

Library of Congress (1999) “MARC 21 Format for Bibliographic Data.” Updated May 2017 [online]. Available from: <https://www.loc.gov/marc/bibliographic/> [Accessed 28 August 2017].

Library of Congress (2005) “541 – Immediate Source of Acquisition Note” [online]. Available from: <https://www.loc.gov/marc/bibliographic/bd541.html> [Accessed 29 August 2017].

Library of Congress (2006) “Frequently Asked Questions” [online]. Available from: <https://www.loc.gov/marc/faq.html> [Accessed 28 August 2017].

Library of Congress (2012) “*Bibliographic Framework as a Web of Data*” [online]. Washington, DC: Library of Congress. Available from: <http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf> [Accessed 23 August 2017].

Library of Congress (April 2014) “BIBFRAME Authorities” [online]. Available from: <https://www.loc.gov/bibframe/docs/bibframe-authorities.html> [Accessed 26 August 2017].

Library of Congress (2014) “246 – Varying Form of Title” [online]. Available from: <https://www.loc.gov/marc/bibliographic/bd246.html> [Accessed 25 August 2017].

Library of Congress (October 2015) “Agents and Roles” [online]. Available from: <https://www.loc.gov/bibframe/docs/pdf/bf2-draftspecagentsroles-10-29-2015.pdf> [Accessed 26 August 2017].

Library of Congress (April 2016) “Overview of the BIBFRAME 2.0 Model” [online]. Available from: <https://www.loc.gov/bibframe/docs/bibframe2-model.html> [Accessed 27 August 2017].

Library of Congress (September 2016) “BIBFRAME on the Move” [online]. Available from: https://www.loc.gov/today/cyberlc/feature_wdesc.php?rec=7571 [Accessed 23 August 2017].

Library of Congress (2016a) “The Semantic Web and Linked Data Concepts.” Presentation [online]. Available from: <https://www.loc.gov/catworkshop/bibframe/Module1-Part1.pptx> [Accessed 23 August 2017].

Library of Congress (2016b) “MARCXML” [online]. Available from: <http://www.loc.gov/standards/marcxml/> [Accessed 14 August 2017].

Library of Congress (March 2017a) “RDF Conventions” [online]. Available from: <https://www.loc.gov/bibframe/docs/pdf/bf2-conventions-march2017.pdf> [Accessed 23 August 2017].

Library of Congress (March 2017b) “Title Notes” [online]. Available from: <http://www.loc.gov/bibframe/docs/pdf/bf2-titles-march2017.pdf> [Accessed 24 August 2017].

Library of Congress (March 2017c) “Event Model” [online]. Available from: <https://www.loc.gov/bibframe/docs/pdf/bf2-eventmodel-march2017.pdf> [Accessed 22 August 2017].

Library of Congress (2017a) “BIBFRAME 2.0 Vocabulary List View” [online]. Available from: <http://id.loc.gov/ontologies/bibframe.html> [Accessed 28 August 2017].

Library of Congress (2017b) “MARC 21 to BIBFRAME 2.0 Conversion Specifications” [online]. Available from: <https://www.loc.gov/bibframe/mtbf/> [Accessed 22 August 2017].

Library of Congress Working Group (2008) “*On the Record*” [online]. Washington, DC: Library of Congress. Available from: <https://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jano8-final.pdf> [Accessed 24 August 2017].

Marcum, Deanna (2011) “Transforming our Bibliographic Framework” [online]. Available from: <https://www.loc.gov/bibframe/news/framework-051311.html> [Accessed 22 August 2017].

McGrath, Kelley (2017) “A Brief Introduction to BIBFRAME 2.0” [online]. Available from: http://www.academia.edu/33540247/A_Brief_Introduction_to_BIBFRAME_2.0 [Accessed 28 August 2017].

RDA Steering Committee (2017) “RDA Frequently Asked Questions” [online]. Available from: http://rda-rsc.org/content/rda_faq [Accessed 23 August 2017].

- RDF Working Group (2014) "RDF: Resource Description Framework (RDF)" [online]. Available from: <https://www.w3.org/RDF/> [Accessed 28 August 2017].
- Sanderson (2015) "Linked Data Best practices and BibFrame." Presentation [online]. Available from: <https://www.slideshare.net/azaroth42/linked-data-best-practices-and-bibframe> [Accessed 27 August 2017].
- Shieh, Jackie (2013) "A Transformative opportunity: BIBFRAME at the George Washington University, an Early Experimenter." *Information Standards Quarterly* [online]. 25(4), 17-21. Doi: <http://dx.doi.org/10.3789/isqv25no4.2013.04> [Accessed 13 August 2017].
- Sprochi, Amanda (2016) "Where Are We Headed? Resource Description and Access, Bibliographic Framework, and the Functional Requirements for Bibliographic Records Library Reference Model." *International Information & Library Review* [online]. 48(2), 129-136. Doi: <http://dx.doi.org/10.1080/10572317.2016.1176455> [Accessed 12 August 2017].
- Styles, R. (2009) "Bringing FRBR Down to Earth" [online]. Available from: <https://dynamicorange.com/2009/11/11/bringing-frbr-down-to-earth/> [Accessed 3 August 2017].
- Styles, R.; Ayers, D. & Shabir, N. (2008) "Semantic MARC, MARC21 and the Semantic Web" [online]. Available from: <http://events.linkedata.org/ldow2008/papers/02-styles-ayers-semantic-marc.pdf> [Accessed 2 August 2017].
- Tennant, Roy (2002a) "MARC Must Die." *Library Journal* [online]. Available from: <http://lj.libraryjournal.com/2002/10/ljarchives/marc-must-die/> [Accessed 5 August 2017].
- Tennant, Roy (2002b) "MARC Exit Strategies." *Library Journal* [online]. Available from: <http://lj.libraryjournal.com/2002/11/ljarchives/marc-exit-strategies/> [Accessed 5 August 2017].
- Tharani, Karim (2015) "Linked Data in Libraries: A Case Study of Harvesting and Sharing Bibliographic Metadata with BIBFRAME." *Information Technology & Libraries* [online]. 34(1), 5-19. Doi: <https://doi.org/10.6017/ital.v34i1.5664> [Accessed 9 August 2017].
- Tillett, Barbara (2003) "What is FRBR?" [online]. Washington, DC: Library of Congress. Available from: <https://www.loc.gov/cds/downloads/FRBR.PDF> [Accessed 28 August 2017].

UDC Consortium (2017a) “Home” [online]. Available from: <http://www.udcc.org/> [Accessed 29 August 2017].

UDC Consortium (2017b) “UDC Users Worldwide” [online]. Available from: http://www.udcc.org/index.php/site/page?view=users_worldwide [Accessed 29 August 2017].

University of Washington (2016) “Mapping of RDA Core to BIBFRAME 2.0” [online]. Available from: <https://www.lib.washington.edu/msd/pubcat/ld/rda-bibframe-mapping> [Accessed 27 August 2017].

W3 Consortium (1999) “Resource Description Framework (RDF) Schema Specification” [online]. Available from: <https://www.w3.org/TR/PR-rdf-schema/> [Accessed 19 August 2017].

W3 Consortium (2005) “The XML Data Model” [online]. Available from: <https://www.w3.org/TR/PR-rdf-schema/> [Accessed 14 August 2017].

Wallis, Richard (2011) “Library of Congress to Boldly Voyage to Linked Data Worlds.” *Data Liberate* [online]. Available from: <http://dataliberate.com/2011/11/02/library-of-congress-to-boldly-voyage-to-linked-data-worlds/> [Accessed 7 August 2017].

Welsh, A.; Bikakis, A.; Garea Garcia, N.; Mahony, S. & Inskip, C. (2015) “Work in Progress: The Linked Open Bibliographic Data Project.” *Cataloguing and Index* [online]. 178(15-19), 1-6. Available from: <http://discovery.ucl.ac.uk/1466469/1/LOBD.RPS.pdf> [Accessed 28 August 2017].

Zepheira “Home Page” [online]. Available from: <https://zepheira.com/> [Accessed 29 July 2017].