

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



BIBFRAME Linked Data: A Conceptual Study on the Prevailing Content Standards and Data Model

Jung-Ran Park, Andrew Brenza and Lori Richards

Abstract

The BIBFRAME model is designed with a high degree of flexibility in that it can accommodate any number of existing models as well as models yet to be developed within the Web environment. The model's flexibility is intended to foster extensibility. This study discusses the relationship of BIBFRAME to the prevailing content standards and models employed by cultural heritage institutions across museums, archives, libraries, historical societies, and community centers or those in the process of being adopted by cultural heritage institutions. This is to determine the degree to which BIBFRAME, as it is currently understood, can be a viable and extensible framework for bibliographic description and exchange in the Web environment. We highlight the areas of compatibility as well as areas of incompatibility. BIBFRAME holds the promise of freeing library data from the silos of online catalogs permitting library data to interact with data both within and outside the library community. We discuss some of the challenges that need to be addressed in order to optimize the potential capabilities that the BIBFRAME model holds.

Keywords: linked data, functional requirements for bibliographic records (FRBR), resource description and access (RDA), semantic web, machine readable cataloging (MARC)

1. Introduction

Over the last several decades, the library community has been faced with the challenge of remaining relevant as an authoritative source of bibliographic data within the larger networked environment of the Web. This relevance has particularly been tested by what a number of information professionals see as the library community's reliance on resource description such as Machine Readable Cataloging (MARC), which do not fully support the establishment of relationships between resources across the Web at large nor optimize library data for machine readability. As a result, the vast majority of bibliographic data held in libraries has been locked in library catalogs, which, although automated, essentially function as electronic equivalents of the physical card catalogs of a hundred years ago [1].

However, due to the rapidly changing technology environment, there is now the opportunity for the library community to expose the data created by cataloging

and metadata professionals and to establish interconnections to related resources across the Web [2]. Newer technologies, such as developed by the World Wide Web Consortium's (W3C) linked open data (LOD) initiative under the banner of the Semantic Web, offer libraries the potential to permit library data to be read and indexed by major online search engines, enhancing user access to authoritative sources of bibliographic data, as has been the library community's historic role to create. As the World Wide Web Consortium defines it, the Semantic Web "is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation" [3]. In other words, the Semantic Web is a method whereby those who are creating content on the Web can markup this content with specific types of metadata in such a way that machines, meaning Web browsers and other applications, can better understand it and use it in novel ways.

Already a number of prominent libraries have developed projects that have published library data that are in compliance with Semantic Web principles, including the Swedish National Library, the French National Library (BnF), the British Library, the Spanish National Library, the German National Library as well as the OCLC [2]. Additionally, implementation of Semantic Web technologies like W3C's Resource Description Framework (RDF) within the library community holds the potential for enriching user experience by permitting users to explore the diverse interconnections between resources through optimizing the machine readability of library data. Lastly, by altering the cataloging process to conform to LOD standards, libraries are afforded the opportunity to reduce cataloging costs through a reduction in duplicate cataloging efforts and to better leverage existing bibliographic data produced elsewhere.

In response to these challenges and opportunities, the Library of Congress (LOC) has developed a high-level model of bibliographic description called the Bibliographic Framework Initiative or BIBFRAME, which aims not only to replace MARC but to provide a framework for optimizing library data within the networked environment. BIBFRAME is essentially an entity-relationship model which uses the Web as architecture and a Resource Description Framework/Extensible Markup Language (RDF/XML) serialization for the description of bibliographic resources. It involves a radical reconceptualization of bibliographic description, eliminating the static, bibliographic record as the product of cataloging in favor of a series of machine readable statements that result in a graph of interconnected entities.

The purpose of this paper will be to examine the development of BIBFRAME through a comprehensive review of relevant literature. We will begin with an overview of BIBFRAME by LOC, outlining the history and structure of the model [in Section 2]. We will then examine the relationship of BIBFRAME to other relevant bibliographic models and content standards including MARC [in Section 3.1], Functional Requirements for Bibliographic Records (FRBR) [in Section 3.2], Resource Description and Access (RDA) [in Section 3.3], and Semantic Web [in Section 3.4]. We will highlight areas of compatibility as well as areas of incompatibility when known. Then, we will end the paper with some concluding remarks.

2. History and overview of BIBFRAME

Officially established in 2011 by the Library of Congress, the Bibliographic Framework Initiative, or BIBFRAME, is a high-level model designed to facilitate the bibliographic description of information resources as well as the exchange of bibliographic data in the networked environment. In 2012 the Library of Congress contracted Zepheria, a consulting firm that specializes in the deployment of semantic

web technologies, to assist with the development of the model. In addition to its work with the Library of Congress, Zepheria has also played, in partnership with Google, Yahoo, and Bing, a key role in the development of Schema.org, a common set of web developer metadata schemas designed to describe websites in support of the indexing efforts of the Internet's major search engines. Over its brief history, BIBFRAME has produced and published a vocabulary for the model, a number of discussion papers related to the vocabulary or other aspects of BIBFRAME implementation, and tools for data conversion.

In its essence, BIBFRAME is an entity-relation model similar to the model put forth in the Functional Requirements for Bibliographic Description. As such, it consists of entities and attributes designed for the description of resources typically managed by cultural heritage institutions. As a result of this entity-relation model, BIBFRAME emphasizes its focus on capturing data elements relevant to bibliographic description, such as title, author, publisher, etc., instead of the creation of complete bibliographic records, which has historically been the focus of the library community. In this way, BIBFRAME establishes a framework for bibliographic description that clearly separates information related to the intellectual contents of resources from their physical properties.

Within this entity-relation model, BIBFRAME is further modeled within RDF/XML in order to bring the model in-line with Semantic Web principles. The use of RDF/XML allows users of the model to identify entities and to describe the relationships between them more clearly and completely. Moreover, it permits these relationships be processed more easily by machines, making library data more conducive to the Web environment. In other words, it allows library data to be found more easily by Internet search engines and, by extension, users. At the heart of this development is the use of Universal Resource Identifiers, or URIs, to name entities and data values, instead of text strings. Thus, the entire BIBFRAME vocabulary of entities and properties has been rendered in URI form.

In summary, BIBFRAME utilizes Web architecture for the description, maintenance, and exchange of bibliographic data in order to accomplish three primary goals [4]:

1. Differentiate clearly between conceptual content and its physical manifestation(s) (e.g., works and instances).
2. Focus on unambiguously identifying information entities.
3. Leverage and expose relationships between and among entities.

2.1 The BIBFRAME model

The newest BIBFRAME model, version 2.0, consists of three core class entities [5, 6]. These are defined below:

- Work: “a resource reflecting a conceptual essence of the cataloged resource” [5]
- Instance: “a material embodiment of a work” [5]
- Item: “an actual copy (physical or electronic) of an instance” [5].

As these entities and their definitions make clear, BIBFRAME, like FRBR, separates the intellectual content of a resource (creative work) from its physical

realization (instance). However, instead of FRBR's four entity classes (work, expression, manifestation, and item), BIBFRAME models only three. Thus, although BIBFRAME and FRBR are conceptually related, it appears that BIBFRAME has simplified the number of entity classes required for bibliographic description.

Below (**Figure 1**) is a graphical depiction of the BIBFRAME model that highlights the relationships between these core entities.

While presenting the evolution of the latest version of BIBFRAME 2.0 from the previous version, McCallum reports the participation of vendors in linked data: "Another major step is now beginning to happen as the vendors who supply many of the services in the community have started to explore linked data, and they are the community's essential innovators" [7, p. 84].

BIBFRAME offers a significant amount of flexibility with resource description. However, per the BIBFRAME documentation, other relationships can also be described. Namely, works can be related to works, instances to instances, works to instances, and instances to works [8]. Beyond the main classes of entities, BIBFRAME also includes a number of properties that are related to each entity. For instance, the creative Work class contains properties that, as one researcher notes, reflect traditional bibliographic elements such as title, creator, language, etc. [9] as well as specific resource Work types that can be used to increase the granularity of a work's description. These properties include resource-type concepts like audio, text, and movingimage.

The instance class contains properties which serve to describe the physical "embodiment" of resources. These properties include terms that overlap with those of the work class such as title and creator, as well as those that describe the aspects of a resource at the manifestation level, such as publisher [9]. Although there is overlap in terminology between the work and instance class, the modeling of these properties in RDF/XML serves as a means to disambiguate terms with the same name through the assignment of a specific URI. Thus, despite identical text names, the use of URIs serves to identify properties within their specific classes.

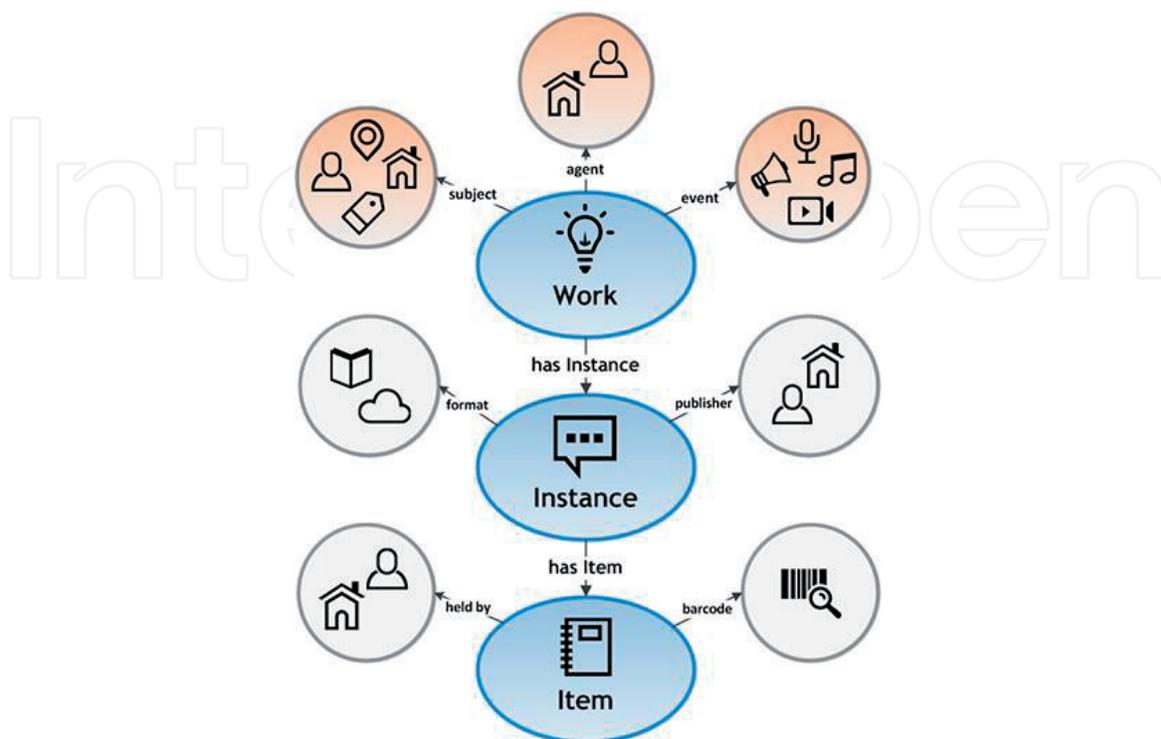


Figure 1.
Graphical depiction of BIBFRAME model [5].

To put it plainly, BIBFRAME attempts to be content standard and model agnostic. Its framework is intended to be flexible enough to accommodate existing models (FRBR, MARC, etc.) and content standards (RDA, VRA, DACS) as well as models and standards that have yet to be developed. Thus, it appears that BIBFRAME appears to be poised to provide the library community with a new model of bibliographic description and exchange that takes full advantage of the Web as architecture. Furthermore, the model also promises to make library data more visible on the Web, not only to the benefit of users looking for library resources but also for re-use in contexts outside of the library community. Finally, it appears that BIBFRAME will permit the full description of relationships between and among resources, enhancing user experience of library information.

2.2 BIBFRAME profiles

It is worth noting that the high degree of flexibility and extensibility built into the model comes with a cost. The under-specification of the model, which is what lends it flexibility, means that there are no built in mechanisms within the model or its RDF schemas that guide and constrain the generation of BIBFRAME data [10]. Nevertheless, the initiative proposes the use of BIBFRAME profiles to address this issue. A BIBFRAME profile can be understood as “a document, or set of documents, that puts a Profile (e.g. local cataloging practices) into a broader context of functional requirements, domain models, guidelines on syntax and usage, and possibly data formats” [10]. In other words, a BIBFRAME Profile serves as a kind template for the generation of BIBFRAME descriptions through the establishment of metadata structure and value constraints. BIBFRAME data can be validated against relevant profiles in order to ensure conformance to an established metadata structure.

However, it should be noted that BIBFRAME profiles exist externally to the model and must be developed within the context of local needs and practices, likely within an application used by cataloguers to capture bibliographic data. In other words, a BIBFRAME profile matches the metadata structures needed within a given context. As long as the overall structure of the data conforms to the BIBFRAME model, then that data should remain interoperable on the Web. Thus, it appears that the initiative is attempting to balance the need for a flexible structure within the model itself and the need to contain that flexibility within a viable framework that can produce consistent and reliable data at the local level.

The study in [11] compares locally created Dublin Core metadata scheme-based application profiles from a number of institutions and digital projects (n = 8). The results of the study present the commonalities and variations of locally developed application profiles and shed light on the effects of resource type and subject domain on naming conventions. The experiences and lessons drawn from the implementation processes of locally developed metadata application profiles are invaluable in the sense that they offer insights and efficient mechanisms for metadata planning and reuse. Thus, the study may shed light on the development of BIBFRAME application profiles in local practice settings.

3. Relationship of BIBFRAME to prevailing content standards and models

It is the intention of the BIBFRAME initiative to design the model in such a way that it not only can serve as the standard encoding and interchange format of bibliographic data within the library community but also to be a model for integrating

library data within the Web environment more generally. As such, the model is designed with a high degree of flexibility in the hope that it can accommodate any number of existing models as well as models yet to be developed. Put simply, the model's flexibility is intended to foster extensibility. The following sections will discuss the relationship of BIBFRAME to the prevailing content standards and models employed by cultural heritage institutions, or those in the process of being adopted by cultural heritage institutions, in an effort to determine the degree to which BIBFRAME, as it is currently understood, can be a viable and extensible framework for bibliographic description and exchange in the Web environment.

3.1 Machine readable cataloging (MARC)

BIBFRAME is intended to replace MARC as the encoding and exchange format for the bibliographic data produced by the library community. But why? What is it about MARC's design that requires the format to be replaced?

First of all, the design of MARC can perhaps be best understood as an exchange format which emphasizes the display of bibliographic information about specific library holdings within electronic catalogs. As a result of this emphasis, MARC records can be conceived as aggregates of information that include descriptions of both the conceptual essence of resources as well as aspects of their physicality [4]. These aggregates are realized in the cataloging process through the application of content standards such as AACR2 and now RDA and are captured, for the most part, in a series of tagged literals or tagged text strings. Ultimately, the overarching structure of MARC records and the content rules used to realize them serve as means to display bibliographic data in much the same way as the physical card catalogs which were its predecessor [1]. MARC's design has served the library community well over the years and has, as the Library of Congress points out in their introductory paper on the BIBFRAME model, allowed librarians to accomplish three important bibliographic tasks [4]:

1. To capture information about the intellectual essence of resources
2. To capture information on the physical aspects of resources
3. To capture information about the management of resources such as control numbers and record handling codes

However, within the current context of the Web environment coupled with the increased processing capabilities of modern computers and applications, MARC's design presents the library community with a number of structural difficulties that limit the potential uses of bibliographic data. First of all, MARC's reliance on the use of literals as identifiers for resources and the elements that compose bibliographic records limits the ability of machines to process MARC information [4]. As a result, variations or equivalences of literals are difficult for machines to parse. Secondly, MARC does not separate information regarding the intellectual content of a resource and its physical carrier clearly enough [4]. Even with adjustments to MARC, such as those included in RDA, an FRBR-based content standard that makes a clearer distinction between the content and carrier, the very format of MARC will not allow machines to utilize it fully [12]. Thirdly, the structure of MARC records, although information rich, are poor at expressing relationships between bibliographic elements in ways that machines can easily understand [13]. Again, even with adjustments to MARC, such as MARC/XML, a serialization intended to increase the machine readability of MARC records, the use of content standards

like AACR2 which were developed primarily with display issues in mind prevents the processing of MARC data significantly [14]. Ultimately, this means that library data is unable to interact with the vast majority of computer applications automatically, limiting the exposure of bibliographic data on the Web, preventing the rich relationships between data elements from being realized and effectively hiding bibliographic information from online users.

BIBFRAME is designed to address these issues. To begin, as one researcher notes, BIBFRAME is not only designed to replace MARC as an encoding and exchange format but to offer a complete re-conception of bibliographic description itself, one that is in-line with the capabilities of the Web environment [15]. BIBFRAME accomplishes this in a number of ways. First, BIBFRAME replaces the idea of the catalog record with the notion that a resource is defined by a discrete series of bibliographic elements. These elements clearly distinguish between the intellectual content of a resource, its physical carrier, and the various entities responsible for its production. Freed from the record as a bundle of data elements, the individual elements are better able to interact in computer applications, and the cataloguer is better able to describe relationships between elements. Secondly, text strings or literals are replaced by URIs or Universal Resource Identifiers. By using URIs to identify bibliographic elements and their values, machines are better able to process the bibliographic information and to utilize the relationships described between them. These two elements, when built upon a Web-based architecture and serialized in RDF/XML, permit BIBFRAME bibliographic data to interact more freely on the Web.

However, despite these changes and the claim that it is standard agnostic, the BIBFRAME initiative also claims that BIBFRAME will be backwards compatible with MARC, meaning that MARC will be mapped to BIBFRAME in such a way that MARC data can be automatically converted to BIBFRAME data without loss of information. Indeed, the BIBFRAME initiative has already developed tools that are available on its website which can translate MARC data into BIBFRAME 2.0 (Figure 2) [16]. As the relationship between MARC elements and BIBFRAME entities may be complex, may even be many-to-many, as one researcher notes [17], the success of such a mapping remains to be seen.

The screenshot shows the BIBFRAME Comparison Tool interface. At the top, it says "BIBFRAME Comparison Tool" and "Compare MARCXML to BIBFRAME2". Below this is a search bar with the value "5226" and a "Search" button. There are also tabs for "Identifier", "Bib ID", "LCCN", "Serialization", "Turtle", and "RDF/XML". Below the search bar, there is a "Lookup Name URIs" section with "No" selected. The main content area is split into two columns. The left column is labeled "MARC" and contains a list of MARC records for Terry Flanagan's "Snoopy on wheels". The right column is labeled "BIBFRAME (Turtle)" and contains the corresponding RDF/XML representation of the same data. The interface also includes a "Lookup Name URIs" section with "No" selected.

Figure 2. Screenshot of the BIBFRAME comparison service results page showing MARC data (left) and BIBFRAME RDF/XML data (right) for Terry Flanagan's Snoopy on wheels.

3.2 Functional requirements for bibliographic records (FRBR)

Published in 1998 by the International Federation of Library Associations (IFLA), the final draft of the Functional Requirements for Bibliographic Records provided a radical re-conception of bibliographic description. In essence, FRBR is an entity-relation model which is composed of four primary classes (work, expression, manifestation, and item) that separate the intellectual content of resources from various aspects of their physical properties, resulting in a new emphasis on the component pieces of bibliographic data rather than the bibliographic record as a whole [15]. As BIBFRAME, with its three primary entity classes (work and instance and tem), is related, at least superficially to FRBR, and considering the likelihood of FRBR's international acceptance as the standard model of bibliographic description, it is useful to compare the two models to determine the degree of compatibility and potential interoperability.

At least on the surface, BIBFRAME and FRBR appear to be closely related. Both models employ the entity-relation approach to bibliographic description and divide the bibliographic record into component pieces which are attached as attributes to entities. As noted, FRBR defines four primary entities for bibliographic description. These are as follows:

- **Work:** “a distinct intellectual or artistic creation” [18]. As such, a work is abstract, pertaining to the intellectual content of a resource as separate from its physical existence. For example, Shakespeare's *Romeo and Juliet* is a work apart from all of the various editions (print and electronic), performances, and films that have embodied it.
- **Expression:** “the intellectual or artistic realization of a work in the form of alpha-numeric, musical, or choreographic notation, sound, image, object, movement, etc., or any combination of such forms” [18]. For example, the English text of *Romeo and Juliet*, as separate from the various ways it is presented in different editions is an expression of the work.
- **Manifestation:** “the physical embodiment of an expression of a work” [18]. For example, the 1998 Signet Classics edition of *Romeo and Juliet* is a manifestation. In other words, when the expression of a work takes on a physical form, as text, film, sound recording, etc., it becomes a manifestation.
- **Item:** “a single exemplar of a manifestation” [18]. For example, an item is a single copy of the 1998 Signet Classics edition of *Romeo and Juliet*.

As can be seen, the FRBR main entities represent a hierarchical movement from abstraction to specificity of a particular information resource [17]. In a similar fashion, BIBFRAME is constructed of entities in a hierarchical fashion, but instead of FRBR's four levels, BIBFRAME defines three [4]:

1. **Work:** “a resource reflecting a conceptual essence of the cataloged resource”
2. **Instance:** “a material embodiment of a work”
3. **Item:** “an actual copy (physical or electronic) of an instance”

Thus, although BIBFRAME only uses three main entity classes, there is still the same movement from abstraction to specificity as represented in the FRBR

hierarchy. Nevertheless, the lack of conformance to the FRBR hierarchy has resulted in much discussion, and, perhaps, even some confusion about how BIBFRAME relates to FRBR. For instance, there appears to be some disagreement in the literature regarding the exact relationship between BIBFRAME and FRBR entities, especially with regard to how the BIBFRAME entities may represent confluations of FRBR entities. Although a number of researchers espouse a correspondence between the BIBFRAME work entity and the FRBR entities work and expression [13, 15, 16, 19], at least one researcher sees a correspondence only between BIBFRAME Work and FRBR Work [20]. Similarly, it appears that most researchers see a correspondence between BIBFRAME instance and FRBR manifestation entities [13, 15, 19], while others see a correspondence between BIBFRAME instance and FRBR manifestation and expression [20].

Perhaps some of the difficulty of mapping BIBFRAME to FRBR lies in the basic ambiguity of the meaning of the respective concepts. For instance, as is noted by IFLA, the FRBR concept of work is an abstraction, meaning that it is hard to define its “precise boundaries” and that the divisions between works and between works and expressions may in fact be culturally dependent [18]. Furthermore, as other researchers have noted, efforts at operationalizing the concept of work have led to at least two different conceptions of the concept. For instance, some have argued that a work can be conceived as the intellectual content of an endeavor with no “assumptions about how it is physically realized,” while, from a different point of view, a work can be conceived as the sum of all common attributes (author, title, etc.) from a set of manifestations [17]. Perhaps complicating the matter is fact that neither BIBFRAME’s nor FRBR’s hierarchy constitutes a definable bibliographic whole. For instance, although FRBR’s entities are organized hierarchically, and are often pictured within a box, there is no single concept to which this hierarchy relates [19]. The need for a kind of super-entity has been noted well in the literature [19]. It would seem that these questions regarding FRBR are equally applicable to BIBFRAME since BIBFRAME does not include a super-entity that encapsulates the work and instance entities. Thus, it appears that there may still be some serious conceptual difficulties that need to be overcome if BIBFRAME, as an entity-relation model, is to be a viable framework for bibliographic description.

Nevertheless, because BIBFRAME appears to be a simplified version of FRBR, perhaps some of the conceptual difficulties regarding FRBR will not negatively affect BIBFRAME as much. For instance, perhaps BIBFRAME’s conflation of FRBR’s work and expression concepts is useful since it is sometimes difficult to determine the boundaries between a work and its expression. However, since the BIBFRAME initiative has suggested that its model is agnostic, meaning that it can be applied to any model, it must be able to be mapped clearly to other models if it is to foster interoperability. Yet, as one researcher notes, to make the model completely agnostic may be unrealistic, since to be perfectly interoperable, both models require almost equivalent semantics and granularity, a situation which would suggest the redundancy of one of the models [2]. This does not seem to be the case between FRBR and BIBFRAME, which means that the initiative may need to re-examine the possibilities of BIBFRAME working with other models.

3.3 Resource description and access (RDA)

BIBFRAME is designed to be content standard agnostic, meaning that the model does not include requirements or specifications for the use of any particular content standard for bibliographic description. In fact, per the initiative, BIBFRAME is intentionally underspecified so that any content standard may be applied successfully within the context of the model, including those that have yet to be developed [4].

Thus, this intentional under-specification is designed to maximize the extensibility of the model and to help ensure its usefulness in a wide range of extant and future information management contexts and use scenarios, as well as for the widest variety of current and future resource types [4].

However, since the BIBFRAME initiative has positioned the model to be the replacement for MARC as the primary method of bibliographic description and data exchange between libraries, the initiative is doing more than simply ensuring the openness of the model to accommodate RDA and other content standards. Per the initiative, the designers are planning on taking an active look at the elements in RDA and other content standards, including the *Anglo-American Cataloging Rules, Second Edition* (AACR2). As a number of researchers have noted, it appears that BIBFRAME is also being designed to specifically accommodate RDA [1, 13, 20], which suggests that this particular content standard may be playing a stronger role in the design of the model than may have been suggested initially. As BIBFRAME is still under development, it remains to be seen exactly to what degree RDA plays a role in the design of the model and what effects this might have on the model's extensibility.

Nevertheless, BIBFRAME designers suggest that the use of profiles will be another way to accommodate a variety of content standards within the model. A BIBFRAME profile is “a document, or set of documents, that puts a Profile (e.g., local cataloguing practices) into a broader context of functional requirements, domain models, guidelines on syntax and usage, and possibly data formats” [10]. According to the initiative, such profiles can be used to define constraints in the creation of BIBFRAME records such as those required by any content standard, including RDA.

As other researchers have noted, RDA may not have gone far enough in distinguishing the content from the carrier of information resources [1, 14]. This potential fundamental flaw in the content standard may pose further difficulties in mapping RDA to BIBFRAME. Such difficulties are presented in the study [21] which shows the uneven mapping between existing RDA classes and BIBFRAME 2.0— particularly the RDA Expression class. The study demonstrates many-to-many relationships in the mapping between RDA and BIBFRAME. Nevertheless, as BIBFRAME is in a relatively early stage of development, the nature and magnitude of these difficulties remain to be seen.

3.4 Semantic web

The current Web environment is structured in such a way that machines, and thus users, are unable to take full advantage of the links that are established among and between resources. In other words, the Web is an environment composed of Web pages and hypertext links that do not describe the nature of the links that connect pages together nor the nature of the data (content) contained in Web pages. In other words, as many researchers note, the current web is a “Web of Documents” versus a “Web of Data” [22, 23]. As a result, current search mechanisms, such as the major search engines, are limited in their ability to utilize information on the Web, relying almost solely on harvesting algorithms to index the content of Web pages and then to match this indexed information against the search terms entered by users. While, as one researcher notes, this method has served the Web well, permitting users to locate needed resources within the vast sea of online information, it lacks the ability to lead users to related content, even when complex and intelligent relevancy algorithms are employed [14]. Furthermore, within the context of the library community, it means that most library data remains relatively difficult to locate online and relatively static with regard to other online resources relevant to

library holdings. In other words, library data, in its current form, remains in the proverbial silo of its online catalogs.

However, through the employment of Semantic Web technologies, there is the potential to expand the uses of library data in the Web environment and thereby to enhance user experience of this data. As is commonly the current case on the Web, a typical hyperlink connects resources but the nature of the connection remains unexplained. However, through the use of Semantic Web and Linked Data principles, such as the use of URIs to identify resources and the embedding of URIs in RDF statements, the nature of these connections can be exposed. In this scenario, a hyperlink can then be defined in almost any way that the user can imagine, indicating the link points to a reference, an author, a subject an authority, etc. Machines can then use this data to “infer” other resources that have been described similarly, such as resources with the same subject heading as the one in question, and permit users to explore these relationships more readily.

At the heart of the Semantic Web are four principles that Tim Berners-Lee, inventor of the World Wide Web and founder and director of the W3C, set forth in his paper entitled “Linked Data” [24]. These principles define the nature of Linked Data as it can be implemented in the current Web environment. Furthermore, they serve as a framework and guide for those interested in making their Web content viable within the Semantic Web, as some conformance to a standard model is required for successful implementation. These principles are as follows:

1. Use URIs as names for things [24].
2. Use Hypertext Transfer Protocol (HTTP) URIs so that people can look up those names [24].
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL) [24].
4. Include links to other URIs, so that they can discover more things [24].

Perhaps most significantly, the conception of Linked Data requires the use of URIs to identify resources or, more specifically, the data elements of resources (Principle 1). In other words, as was mentioned in the discussion on MARC above, the use of text strings to identify resources makes machine processing difficult. The shift to URIs as identifiers means that machines can better understand the identity of resources, especially if they are known by different names or to disambiguate different resources known by the same name. Furthermore, the shift to URIs also signals the shift in understanding in regards to the nature of information resources as described in the above FRBR section. It emphasizes the identification of discrete data elements within information resources versus the identification of the resource as a whole. In other words, it emphasizes the atomization of resources into their relevant components.

Principle 2 emphasizes the need for a common schema for the definition of URIs. Since HTTP is already the foundation of data transfer on the Web and since it appears to be serving its function well, Berners-Lee suggests that using this common protocol for the definition of URIs will increase the usefulness of data described in Semantic Web compliant ways. Furthermore, as the BIBFRAME initiative notes, these URI schemes should not be obscure, even if they are represented in HTTP, in order to facilitate data interaction and reuse [4].

Principle 3 emphasizes the need for a common framework for the exchange of information described with URIs. Typically this means the use of RDF for the

modeling of data, which, as the BIBFRAME initiative notes, is the most common framework within the LOD community [4]. As a conceptual framework for representing resources on the Web [15], RDF can be understood as a kind of syntax for structuring data in such a way that it fosters the machine readability of that data through the use of URIs and the delineation of relationships between data elements. RDF is typically rendered in XML, but other languages, such as N3, Turtle, and N-Triples, are also used [22]. In its basic format RDF consists of statements, called triples, which, like sentences, contain subjects, predicates, and objects. A basic RDF statement might read as “Book A (subject)—Written By (predicate)—Author A (object),” where Book A, Written By, and Author A are all identified by URIs, with the possible exception of the object, which could be populated with a text string [22]. The power of this model is that the type of relationships between resources (Book A and Author A) is defined (Written By). **Figure 3** illustrates this statement graphically. Thus, as a result of delineating relationships between data elements, tools called “reasoners” can make inferences about the data [19].

A reasoner is a software application that can make logical inferences based on a set of statements, or axioms, provided to it through queries. Although there are many query languages that can be used to access and manipulate data modeled in RDF, the SPARQL Protocol and RDF Query Language (SPARQL) has emerged as the most popular [23]. For instance, a reasoner, beginning with a SPARQL query to a database that contained the above RDF statement, could use that statement to make inferences about other books written by Author A and present those to users without the user specifically querying the system to do so (**Figure 4**). Furthermore, there are no restrictions on the number of RDF triples that can be created for a particular resource, which fosters the development of rich data graphs, or the decentralized interconnections between data elements, within the Web environment. Although RDF is not a data format, but a model for representing data elements on the Web, it has been serialized in a number of ways. For instance, BIBFRAME has been modeled in RDF/XML, but other languages, like N-Triples, ATOM, and JSON, also exist. Although BIBFRAME has been modeled in RDF/XML, the Initiative claims that any data format that conforms to the standard model of URIs embedded in triples should be compliant with the BIBFRAME model [4].

Principle 4 encourages broad use of the connections established through the first three principles [4]. Thus, data that has been described in conformance with the above principles can be considered Linked Data and Semantic Web compliant. However, if the URIs expose, point to, or otherwise include information that is made freely available for reuse on the Web, such as through a Creative Commons license, this data can be considered Linked Open Data, not just Linked Data.

As stated earlier, a number of prominent libraries have published library data in compliance with Semantic Web principles [2]. Even though these projects are not BIBFRAME projects, they are generally in-line with FRBR principles of bibliographic description. It is worth examining the degree to which the model conforms to the current understanding of Linked Data and the Semantic Web. To begin, BIBFRAME has defined URIs for all BIBFRAME entities and properties within the BIBFRAME namespace. This is particularly important as some properties that belong to different classes have identical names. The use of URIs serves as a clear



Figure 3.
Graphical depiction of a basic RDF statement.

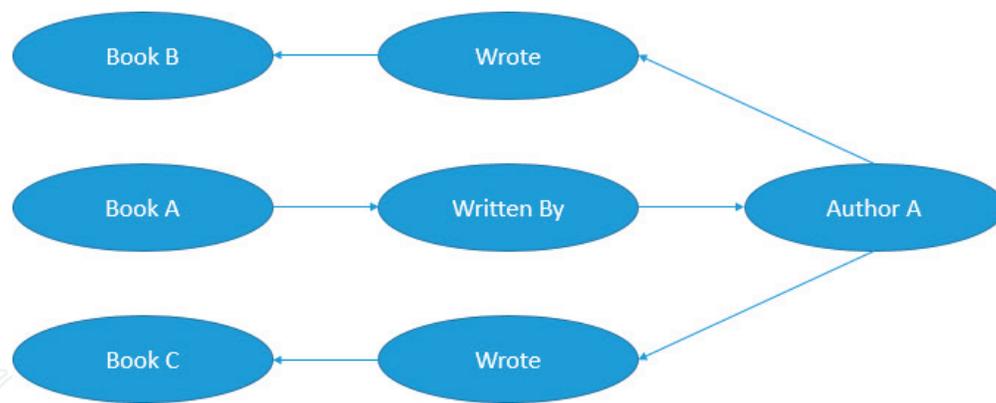


Figure 4.
Graphical depiction of a reasoner using RDF statements to infer additional resources.

means to disambiguate these properties. Secondly, as has been noted, BIBFRAME has been modeled in RDF/XML [25].

In addition to these two factors, the BIBFRAME model, like FRBR, deconstructs bibliographic records into their component pieces through the entity-relation conception of bibliographic description. Taken together, these elements suggest that BIBFRAME conforms well to the current understanding of Linked Data and the Semantic Web. Furthermore, even though the initiative has rendered the model in RDF/XML, BIBFRAME is also designed to be compliant with other data formats which conform to the structured use of URIs within syntax of triples statements. Thus, it also appears that BIBFRAME is, at least in principle, poised to integrate library data with other data produced within contexts outside the library community. This aspect too suggests that BIBFRAME is Semantic Web friendly.

4. Discussion

There are challenges that may hinder the widespread adoption of BIBFRAME within the library community. In addition to the modeling difficulties and potential conceptual misalignment of BIBFRAME in relation to MARC, FRBR, RDA, Linked Data, and RDF, there are difficulties posed by complex resource types such as audiovisual materials, manuscript, and serial publications [26]. Additionally, although MARC is in essence an exchange format for bibliographic data, it has become so intertwined with the content standards applied to it, first AACR2 and now RDA; this union of the two may further entrench it within the library community. Without consensus regarding the fate of MARC, it may be difficult to persuade MARC's adherents, even if BIBFRAME proves to offer more capabilities to catalogers.

There may be significant conceptual difficulties with mapping RDA to BIBFRAME. For instance, RDA was developed within the context of the FRBR entity-relationship model. As such, RDA separates resources into FRBR's four main entity classes: Work, Expression, Manifestation and Item. However, as has already been noted, BIBFRAME's main entity classes do not align with FRBR's classes in an exact manner [20]. This lack of alignment may make the mapping between RDA and BIBFRAME difficult.

Although it appears that BIBFRAME conforms to current conceptions of Linked Data and the Semantic Web, there are still a number of issues worth considering. First, since the usefulness of the relationships delineated through the RDF triples depends on the quality and stability of the resources to which they are linked, the BIBFRAME initiative will have to determine the degree to which it will maintain

its own controlled vocabularies and ontologies versus relying on others to do so. Ontologies suitable for the Linked Data environment are taxonomies and thesauri that meet the W3C Web Ontology Language (OWL) standard [22]. For example, the Library of Congress Subject Headings modeled in the Simple Knowledge Organization System (SKOS) framework is an OWL-compliant ontology.

The existence of high-quality, stable ontologies is particularly a relevant concern with regard to the use and reuse of Linked Open Data resources. For instance, as one researcher notes, many LOD ontologies and vocabularies are developed in the context of research projects, which means that for a particular moment they may be up-to-date, accurate, and in compliance with current standards, though it does not ensure continued governance and maintenance [12]. Thus, the reliance on such vocabularies could present the threat of obsolescence should governing bodies discontinue their activities. Thus, it appears that BIBFRAME will need to assess the stability of ontologies and vocabularies, such as those for resource type, and determine if it is better to develop and maintain its own within the BIBFRAME namespace or to link to resources outside the initiative.

Secondly, although BIBFRAME claims that the model should be interoperable with any serialization using triples and URIs, the fact that the initiative has serialized the model in RDF/XML may be a limitation. In other words, because the initiative has limited its serialization within a single framework, it may discourage implementation in other formats. As one researcher notes, it may be better for the initiative to provide potential implementers with examples from a number of possible serializations in order to demonstrate the model's flexibility, extensibility, and potential for interoperability [2].

Thirdly, there may be difficulties with viably implementing the BIBFRAME model which are rooted in the nature of RDF itself. As the study in [19] notes in their comparison of BIBFRAME, FRBR, and RDA, there is nothing in RDF that prevents people from making nonsensical RDF triples. In other words, there are no validation mechanisms for the creation of RDF statements, as there are for well-formed XML or HTML documents. While, as the researchers note, BIBFRAME has proposed the use of profiles in order to establish content rules and constraints on the creation of BIBFRAME records, these do not prevent potential difficulties with the integration of BIBFRAME data elements with data elements modeled in other frameworks such as FRBR.

However, perhaps the biggest threat to BIBFRAME as a mechanism to expose library data in a Semantic Web friendly way lies in the fact that, like the framework itself, the Semantic Web is still under development. For instance, as has been noted in the literature, understanding of what actually constitutes Linked Data is still under debate [19]. Since the very underpinning of the Semantic Web is still in flux, there is a possibility that any operationalization of the concept will change in the future. Thus, if the current methods for creating Linked Data alter significantly in the future, and if data described with current methods cannot be easily translated into the newer modes, then BIBFRAME Linked Data could potentially become obsolete, resulting in the relegation of library data to yet another, but different, silo.

This final point may also be exacerbated by the very fact that BIBFRAME is a model for the description of bibliographic data within the library community itself. For instance, as some researchers have noted, for data to be truly integrated in the Web, what is required is a common model for data description that includes not only bibliographic data but data of all types [2]. In other words, BIBFRAME, as a model for the description of bibliographic data, may not be intuitively understood by others outside the library community, which may result in a lack of implementation and difficulties with the integration of data embedded in other frameworks.

This is particularly important as BIBFRAME data is intended for use outside of the library community, especially with regard to the authority data such as controlled subject headings that have been the province of the library community for so long [2, 13]. Thus, while BIBFRAME holds the promise of freeing library data from the silos of online catalogs and to permit library data to interact with data both within and outside the library community, there may still be challenges to overcome in order to optimize these capabilities.

5. Conclusion

It is the intention of the BIBFRAME initiative to design the model in such a way that it not only can serve as the standard encoding and interchange format of bibliographic data within the library community but also be a model for integrating library data within the Web environment more generally. As such, the BIBFRAME model is designed with a high degree of flexibility that can accommodate any number of existing models as well as models yet to be developed within the Web environment. The model's flexibility is intended to foster extensibility.

However, regarding the model itself, there appears to be a significant need to consider the creation of a super-entity that would encapsulate the work and instance entities. With regard to the cataloging requirements for the description of complex resources such as audiovisual materials and serial publications, the creation of such a super-entity would solve a number of bibliographic description challenges. The existence of a super-entity would permit the description of resources and relationships that are currently difficult to model within the existing framework. Resources that do exhibit intellectual content or that are primarily event based would be easier to depict if such a super-entity was present.

BIBFRAME attempts to be content standard and model agnostic. Its framework is intended to be flexible enough to accommodate existing models. While increasing its extensibility, the framework may also result in an uncertainty of its application in specific cataloging contexts. This too may limit the willingness of the library community to invest in its adoption. Furthermore, even though BIBFRAME's potential for extensibility is intended to foster its adoption in a wide range of bibliographic contexts and to work equally well for divergent descriptive needs, its ability to accommodate most if not all modeling and content standards currently in use or yet to be invented may be optimistic. In this regard, BIBFRAME's ability to support widespread interoperability needs to be further addressed.

In this study we discussed the relationship of BIBFRAME to the prevailing content standards and models employed by cultural heritage institutions in order to determine the degree, to which BIBFRAME can be a viable and extensible framework for bibliographic description and exchange in the Web environment. Despite the promise of improved data management, sharing, and usage offered through the BIBFRAME model, there are various challenges that must be overcome for its adoption within the library community. However, if the initiative can overcome what will likely be significant challenges to the implementation of the model, BIBFRAME appears to be poised to become the next standard of bibliographic description and exchange for the library community and beyond. Furthermore, the model also promises to make library data more visible on the Web, not only to the benefit of users looking for library resources but also for reuse in contexts outside of the library community. Finally, it appears that BIBFRAME will permit the full description of relationships between and among resources, enhancing and enriching the user experience of library information.

IntechOpen

IntechOpen

Author details

Jung-Ran Park*, Andrew Brenza and Lori Richards
The College of Computing and Informatics, Drexel University, Philadelphia, USA

*Address all correspondence to: jp365@drexel.edu

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Dean JW, Charles AC, Edward T. Coming to a library near you, via BIBFRAME. In: *The Library with the Lead Pipe*. 2013. Available from: <http://www.inthelibrarywiththeleadpipe.org/2013/charles-a-cutter-and-edward-tufte-coming-to-a-library-near-you-via-bibframe/> [Accessed: 02 June 2020]
- [2] Svensson LG. Are current bibliographic models suitable for integration on the web? *Information Standards Quarterly*. 2013;25(4):6-13
- [3] World Wide Web Consortium. The semantic web made easy [Internet]. Available from: <http://www.w3.org/RDF/Metalog/docs/sw-easy.html> [Accessed: 02 June 2020]
- [4] Library of Congress. Bibliographic Framework as a Web of data: Linked Data model and supporting services. 2012. Available from: <http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf> [Accessed: 02 June 2020]
- [5] Library of Congress. Overview of the BIBFRAME 2.0 model [Internet]. 2016. Available from: <http://www.loc.gov/bibframe/docs/bibframe2-model.html> [Accessed: 02 June 2020]
- [6] Library of Congress. BIBFRAME 2.0 Vocabulary [Internet]. 2016. Available from: <http://www.loc.gov/bibframe/docs/index.html> [Accessed: 02 June 2020]
- [7] McCallum S. BIBFRAME development. *JLIS.it*. 2017;8(3):71-85
- [8] Library of Congress. BIBFRAME relationships [Internet]. 2014. Available from: <http://www.loc.gov/bibframe/docs/bibframe-relationships.html> [Accessed: 02 June 2020]
- [9] Mitchell ET. Three case studies in linked open data. *Library Technology Reports*. 2013;49(5):26-43
- [10] Library of Congress. BIBFRAME profiles: Introduction and specification [Internet]. 2014. Available from: <http://www.loc.gov/bibframe/docs/bibframe-profiles.html> [Accessed: 02 June 2020]
- [11] Park J-r, Andrew B, Lu C. A comparative analysis of metadata best practices and guidelines: Issues and implications. *International Journal of Metadata, Semantics and Ontologies*. 2015;10(4):240-260
- [12] Fallgren N, Lauruhn M, Reynolds RR, Kaplan L. The missing link: The evolving current state of linked data for serials. *The Serials Librarian*. 2014;66(1-4):123-138
- [13] Kroeger A. The road to BIBFRAME: The evolution of the idea of bibliographic transition into a post-MARC future. *Cataloging & Classification Quarterly*. 2013;51:873-890
- [14] Breeding M. Linked data: The next big wave or another tech fad? *Computers and Libraries*. 2013;33(3):20-22
- [15] Ballegoie MV, Borie J. From record bound to boundless: FRBR, linked data, and new possibilities for serials cataloguing. *The Serials Librarian*. 2014;66(1-4):76-87
- [16] Library of Congress. MARC to BIBFRAME comparison viewer [Internet]. Available from: <http://id.loc.gov/tools/bibframe/compare-id/full-ttl> [Accessed: 02 June 2020]
- [17] Godby CJ. The relationship between BIBFRAME and OCLC's linked-data model of bibliographic description: a working paper [Internet]. 2013. Available from: <http://www.oclc.org/content/dam/research/publications/library/2013/2013-05.pdf> [Accessed: 02 June 2020]

[18] IFLA. Study group on the functional requirements for bibliographic records. In: *Functional Requirements for Bibliographic Records*. Munich: K.G. Saur Verlag; 1998

[19] Baker T, Coyle K, Petiya S. Multi-entity models of resource description in the semantic web: A comparison of FRBR, RDA, and BIBFRAME. *Library Hi Tech*. 2014;**32**(4):562-582

[20] Meehan TP. BIBFRAME. *Catalogue and Index*. 2014;**174**:43-52

[21] Taniguchi S. Examining BIBFRAME 2.0 from the viewpoint of RDA metadata schema. *Cataloging and Classification Quarterly*. 2017;**55**(6):387-412

[22] Yang S, Lee YY. Organizing bibliographic data with RDA: How far have we stridden towards the semantic web? In: Park JR, Howards L, editors. *New Directions in Information Organization*. Bingley, UK: Emerald Insight; 2013. pp. 3-27

[23] Yoose B, Perkins J. The linked open data landscape in libraries and beyond. *Journal of Library Metadata*. 2013;**13**(2/3):197-211

[24] Berners-Lee T. Linked data [Internet]. 2006. Available from: <http://www.w3.org/DesignIssues/LinkedData.html> [Accessed: 02 June 2020]

[25] Taniguchi S. Is BIBFRAME 2.0 a suitable schema for exchanging and sharing diverse descriptive metadata about bibliographic resources? *Cataloging and Classification Quarterly*. 2018;**56**(1):40-61

[26] Park J-R, Richards L, Brenza A. Benefits and challenges of BIBFRAME: Cataloging special format materials, implementation, and continuing educational resources. *Library Hi Tech*. 2019;**37**(3):549-565