JLIS.it 8, 3 (September 2017)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12415

JLIS.it

BIBFRAME Development

Sally H. McCallum [a]

a) Chief, Network Development and Standards Office, Library of Congress, USA

Contact: Sally H. McCallum, smcc@loc.gov.
Received: 8 June 2017; **Accepted:** 11 June 2017; **First Published:** 15 September 2017

ABSTRACT: The Library of Congress has been exploring "linked data" for over 10 years. The genesis of this development goes back to the W3C projects in the late 1990s on SGML, then XML, HTML, and finally a linked data oriented format RDF. In 2007 the Library of Congress organized a community inquiry into the "future of bibliographic control," that became a catalyst for exploration and change. A wide ranging report was written and some major recommendations were made. Some have been acted upon, others are still on the table such as a rethinking of subject vocabularies. But two technical recommendations were ideally suited for exploration using the emerging linked data framework of the W3C: the use of technology to get broader use of library curated vocabularies and the replacement of the MARC format with a data interchange framework that makes library data more readily available on the web. The Library was investigating the linked data framework for standards and models for exposing its vocabularies such as the LCSH. Accordingly, LCSH was made publicly available as linked data in 2009 followed by name authorities, countries, languages, and many other controlled lists used in bibliographic standards such as MARC, MODS, and PREMIS. This project became the Library of Congress Linked Data Service. Its aim is to establish stable identifiers in URI form for entities and concepts that are useful for description of cultural heritage material. Then in 2011 the Library of Congress announced the start of the Bibliographic Framework Initiative (subsequently labelled "BIBFRAME") to respond to the second major technical recommendation of the future of bibliographic control report: to replace MARC for interchange and to make library resources more visible on the web. The Library of Congress, with the library community, is tackling the challenges described above and this paper looks at main aspects of that development.

AKNOWLEDGMENT: Parts of this paper are adapted from the author's earlier publication in *Linked Data for Cultural Heritage* (Chicago, ALA, 2016).

KEYWORDS

BIBFRAME; Bibliographic formats; Cataloguing; Library of Congress; Library linked data; MARC.

CITATION

McCallum, S.H. "BIBFRAME Development". *JLIS.it* 8, 3 (September 2017): 71-85. doi: 10.4403/jlis.it-12415.

JLIS.it 8, 3 (September 2017)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12415

Introduction

The Library of Congress has been exploring "linked data" for over 10 years, first providing a linked data service of major files such as Name Authorities and Library of Congress Subject Headings (LCSH) followed by in depth work on an exchange format that could replace MARC in the future. The genesis of this development goes back to the World Wide Web Consortia's (W3C) projects in the late 1990s on SGML, then XML, HTML, and finally a linked data oriented format RDF. Today "linked data" has become a key goal of our Internet-based environment and libraries are active with investigation and experimentation. However, just what is meant by "linked data" and how will it transform the user experience are not entirely clear. This is not a new concept to librarians as they have been doing "linking" for years without the tools or concepts of today, by linking with strings (such as names or subjects) or numbers (such as ISBNs). Now the W3C that makes the rules and advances the web and Internet environment has developed models, principles, tools, and components that enable linking to be more dynamic by using URIs (Universal Resource Identifiers) that can be machine-followed to reveal much more about an entity – and where further links may be found or inferences made to gather more information. This reflects the "Linked Data principles" articulated by Tim Berners-Lee in 2006: use URIs to name things; make them HTTP URIs so machines can look them up; when looked up, provide useful information using Linked Data standards such as the Resource Description Framework (RDF) and the search language, SPARQL; include links to other URIs (Berners-Lee 2006). So linking is not new to bibliographic data but Linked Data via URIs is, and it holds interesting potential for libraries.

In 2007, at about the same time that Berners-Lee was explaining linked data, the Library of Congress organized a community inquiry into the "future of bibliographic control," that became a catalyst for exploration and change. A wide ranging report, *On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control* (Library of Congress Working Group on the Future of Bibliographic Control 2008), was written by a committee and some major recommendations were made. Some have been acted upon, such as the testing and subsequent implementation of the cataloguing rules *Resource Description and Access* (RDA) (Canadian Library Association, Chartered Institute of Library and Information Professionals (Great Britain), and Joint Steering Committee for Development of RDA 2010). Others are still on the table such as a rethinking of subject vocabularies. But two technical recommendations were ideally suited for exploration using the emerging linked data framework of the W3C. They are the following.

- Use technology to get broader use of library curated vocabularies (Library of Congress Working Group on the Future of Bibliographic Control 2008, 25)
- Replace the MARC format with a data interchange framework that makes library data more readily available on the web. (*ibid.*)

Linked data concepts were not mentioned in the future of bibliographic control report as the terminology was just being coined, although the concepts had been discussed in the W3C in the context of the semantic web since the late 1990s.

JLIS.it 8, 3 (September 2017)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12415

In 2007 the Library of Congress was already starting on the first of these tasks thus the recommendations of the report were timely. The Library was investigating the linked data framework for standards and models for exposing its vocabularies such as the *Library of Congress Subject Headings* (LCSH). Accordingly LCSH was made publicly available as linked data in 2009 followed by name authorities, countries, languages, and many other controlled lists used in bibliographic standards such as MARC, MODS, and PREMIS. This project became the *Library of Congress Linked Data Service*, (LDS).[1] Its aim is to establish stable identifiers in URI form for entities and concepts that are useful for description of cultural heritage material. The URIs lead to data expressed in a way that enables its consumption in the linked data environment, i.e., in RDF.

Then in 2011, Deanna Marcum, then the Associate Librarian for Library Services at the Library, announced the start of the Bibliographic Framework Initiative (subsequently labelled "BIBFRAME") to respond to the second major technical recommendation of the future of bibliographic control report: to replace MARC for interchange and to make library resources more visible on the web (*A Bibliographic Framework for the Digital Age* 2011, 23).

The Library of Congress, with the library community, is tackling the challenges described above. This paper looks at five aspects of that development:

- Using RDF
- Library of Congress Linked Data Service
- Developing BIBFRAME
- Models
- Exploring

Using RDF

Since RDF is a key building block of this whole development, an introduction to the special characteristics that make it useful will illustrate its differences from today's common data interchange standards such as MARC.[2]

RDF triples and serializations

At its simplest, RDF is about triples which can be thought of in terms of basic grammatical sentences with three parts: subject, verb, and object. For example, the sentence: "Bambi is a deer" has three parts: subject "Bambi", verb "is a", object "deer". A triple in RDF is usually referred to as a "statement". Since the "RD" in "RDF" is "Resource Description", essentially, one starts with a resource, and describes it with statements. The resource description will be a set of triples (i.e., sentences or statements).

In order for the triples to be consumable by a computer, the triples have to be represented in a commonly understood format that a computer can recognize and process. This is a familiar concept as

[1] http://id.loc.gov.
[2] This description of RDF is adapted from material prepared by Ray Denenberg for the BIBFRAME pilot project (Denenberg 2015).

JLIS.it 8, 3 (September 2017)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12415

MARC has at least two such format structures, ISO 2709 and XML. For RDF there are a number of these formats in use; called "RDF serializations". Consider the book, *Baseball for everyone.* It is a resource, and specifically, a Work. As such, it has a URI: http://example.org/resourceX. In constructing triples describing this resource, this URI is used as the subject for these triples. Here are some example statements (in plain English, not RDF syntax).

```
http://example.org/resourceX       is a           Work

http://example.org/resourceX       is a           Text

http://example.org/resourceX       has title      "Baseball for everyone"

http://example.org/resourceX       has creator    "DiMaggio, Joe, 1914-1999"
```

Applying a serialization to this set of RDF triples that describe the resource *Baseball for everyone* we get the following. (This serialization is called "Turtle" (Terse RDF Triple Language).)

```
<http://example.org/resourceX>        a        exns:Work ;
      a        exns:Text ;

      exns:title                "Baseball for everyone" ;
      exns:creator              "DiMaggio, Joe, 1914-1999" .
```

The first two statements specify types for this resource. Any resource may have more than one type, and in this case it is a Work and it is also a Text (as opposed, for example, to a Moving Image). These types are called "classes" in RDF, and, in the example ontology, Text is a subclass of Work. The class names Work and Text are referred to as exns:Work and exns:Text to denote that they are being used within the context of the example ontology, and not some other ontology that might also use these terms but with a different meaning. Verbs in the triples are called "properties" in RDF and the property "is a" is abbreviated as just "a"; this is well-known within all of RDF to denote "is of type". URIs within RDF triples are enclosed with angle brackets. In the example above, the objects in this ontology are literals and not classes or URIs, and the convention is to place literal strings in quotation marks. There are other writing conventions, but note in the example above that the subject does not need to be repeated when triples in a statement have the same subject and the property/object pairs are separated by semi-colons (";"). One whole statement for a subject ends with a period (".").

Literals, URIs, and RDF

The triples that make up the description of *Baseball for everyone* and their serialization shown above illustrate two RDF statements as literals. However, when the object is a URI the information itself is not the object but it points with the URI to another RDF resource that does contain the information. For example, instead of the literal "DiMaggio, Joe, 1914-1999", the object of that triple might have been a URI that pointed to an RDF description of Joe DiMaggio and that included more information about him than his name.

Since many of the characteristics of RDF are commonplace and could be illustrated by the MARC format, what are the advantages to moving to this new method of expressing descriptions? First of all, this is a structure that is being embraced for data across communities. The web itself set the

JLIS.it 8, 3 (September 2017)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12415

precedent for this as HTML and its successors have been adopted universally to provide information on the web.

Also, currently, the conventional way to link to a richer description of, for example, Joe DiMaggio – information that librarians do include in some of their MARC authority records -- is by the "authorized" data string for his name or by identifiers. While these linking keys are not usually machine actionable except within an internal system, the URI system that was built for the Internet is a vehicle that enables a machine to find the richer information no matter where it is.

And finally the conventions for describing the ontology properties and classes that are used for RDF statements are also machine understandable to a large extent. Thus the amount of "thought" that a machine can bring to information, is greatly enhanced. In some situations, a machine can start to make relations and connections that are not explicit in the data; this process is called inferencing.

So using RDF and following W3C and Internet standards opens up greater possibilities for the use of library produced data. This new and complex form of linking data is in its early stages and libraries are getting into the new environment early. Library use of these conventions can be explored and grow as the Internet further develops linked data guidelines.

Library of Congress Linked Data Service (LDS)

With the above brief introduction to the cornerstone standard for the linked data environment, RDF, let's explore how it is used in the BIBFRAME project. The first step was to make the many large (LCSH, Name Authority File, etc.) and small (language codes, content and media terms, etc.) controlled lists available for use in a linked data application. These many controlled vocabularies and lists used for bibliographic description needed to be transformed from their printed or web forms into RDF so their links can be referenced by URIs and their descriptions can be made accessible in RDF. Making those links both automatic and enhanced is the aim of the Library of Congress Linked Data Service[3] which began development in 2007 with LCSH and continues in the BIBFRAME project, adding bibliographic vocabularies and lists that are needed to make RDF bibliographic descriptions for library material. These developments are part of the mandate to "Use technology to get broader use of curated vocabularies" as recommended in the On the Record report.

To transform this data to RDF a vocabulary or "ontology" with appropriate properties and classes was needed. This is a familiar concept to librarians as in the MARC environment, MARC itself is a vocabulary although expressed in a totally different way, with tags and indicators and subfield codes being roughly analogous to properties and classes.

In the Library of Congress's linked data projects two ontologies have been evolving to support traditional bibliographic data: BIBFRAME, with properties and classes related to bibliographic description data, and MADSRDF (Metadata Authority Description Schema in RDF) (*MADS/RDF Primer* 2013) with classes and properties related to expressing authority and list data. The BIBFRAME project uses these ontologies that are especially suited to library data but it also uses elements from other ontologies where that is appropriate. The W3C developed several foundation

[3] http://id.loc.gov.

JLIS.it 8, 3 (September 2017)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12415

ontologies for RDF such as OWL (Web Ontology Language), RDF, and RDFS (RDF Schema). They contain basic properties that are important in building any OWL-based RDF ontology. The W3C have also produced a basic thesaurus ontology, SKOS (Simple Knowledge Organization System), that has become widely used, and several others that may be important in the future (*SKOS Simple Knowledge Organization System Reference* 2009). The BIBFRAME and MADSRDF vocabularies draw on these and other well established and believed-to-be-stable ontologies.

The ontology closely related to the LDS service is MADSRDF. It provides a data model that is suitable for the types of authority, code, and term lists used in bibliographic data. MADSRDF is closely related to SKOS. The more complex data structures used in bibliographic data are not supported in SKOS, intentionally, as it is intended for broad application, but the MADSRDF ontology is fully mapped to SKOS in the RDF that supports the ontology. This enables interoperability of MADSRDF data with the boarder, less detailed SKOS standard. For example, a madsrdf:Authority is a sub-class of skos:Concept.

Another feature of MADSRDF is the separation of the information that provides labelling of the concept being described, from, in the case of names, information that more appropriately describes aspects of the "real world object" (RWO) that the label addresses. For example, in the case of a person, a property such as madsrdf:authoritativeLabel relates to the person's name, DiMaggio, Joe, 1914-1999, while RWO information such as madsrdf:fieldOfActivity describes the person, a baseball player.

After starting with the Library of Congress Subject Headings, LDS has been expanded to include the cooperative name authority file, other subject thesauri, and several code and term lists used in bibliographic data. LDS has also been expanded to support the value lists found in the PREMIS standard for preservation of bibliographic material, especially electronic. This enables experimentation in the community with linked data related to preservation as these lists are accessible and referenceable in RDF for institutions to use in data format and storage configurations.

Developing BIBFRAME

Changing environment

With the terminology component started the next step toward a linked data environment was to restructure the bibliographic descriptions themselves, which are the heart of bibliographic control. There were many motivators for taking this step, one being MARC[4] itself. It was certainly the case that MARC had moved as best it could with the times. It had been adapted to different cataloguing conventions, archival descriptions, different media needs, holdings complications, and conventions of other countries such as Canada, the United Kingdom, and Germany. It had even developed technically, moving from limited character sets to embrace UNICODE and enabling an XML structure in addition to the traditional ISO 2709 (*ISO 2709:2008. Information and Documentation -- Format for Information Exchange* 2008) structure for bibliographic descriptions. And because of the adaptations it fully

[4] *MARC standards.* Washington, Network Development and MARC Standards Office, Library of Congress. http://www.loc.gov/marc.

JLIS.it 8, 3 (September 2017)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12415

permeated the bibliographic environment as it provided the core data for integrating acquisitions, cataloguing, online public catalogues, record sharing and resource lending systems.

But this extensive reach and long term use also bred a large accumulation of data elements – that ran up against some structural limitations in the tagging and subfield coding in the ISO 2709 structure. MARCXML[5] provided structural expansion but not without loss when converting to the traditional structure. The ISO 2709 data structure with the MARC application has also not been easy to adapt to replacing data with URI links.

At the same time cataloguing practices have been changing with the most disruptive change being the introduction of a new cataloguing norm for libraries, RDA. RDA has an increased emphasis on relationships, making it compatible for linked data. Bibliographic control practices are also adding into records pointers to many "non-traditional" resources such as cover images, tables of contents, reviews, abstracts, author biographies, and content excerpts – sometimes called the Amazon effect.

Other developments in the broader community have also provided impetus for format redevelopment. There have been several resource modelling projects the last 15 or so years: FRBR (Functional Requirements for Bibliographic Records, 1998), <indecs> metadata framework (Rust and Bide 2000), and CIDOC Conceptual Reference Model (International Council of Museums 2014), to name a few. FRBR was used as the model behind RDA.

The nature of library collections has also been shifting to electronic resources and to more media resources like recorded sound and moving images, even as the printed resource production has not yet diminished. And finally library systems have been challenged to provide more functionality: access management (licensing and rights) for electronic resources and object management to preserve both analogue and electronic resources.

All of the above factors have both stressed the MARC format itself and argued for a new resource description and interchange environment.

BIBFRAME goals

With these motivations the community was ready to launch an exploration of a new framework for bibliographic control, which was called for in the future of bibliographic control report, *On the Record*, described earlier. The goals of the Bibliographic Framework Initiative (BIBFRAME) are ambitious. Just as MARC had tried to accommodate different cataloguing norms and various media, the new initiative needs to be as open to different norms and models and media as possible.

In a new environment, the key areas of description, authority, and holdings might be reconfigured, coded data appears to be a natural for conversion to links, and the relationship between pre-coordinated subject headings and subject terms could be studied. And all that data that had often been pushed into MARC as the only carrier solution needed to be rethought -- technical metadata, preservation metadata, rights, and special archival data.

[5] *MARCXML: MARC 21 XML Schema.* Washington, Library of Congress. http://www.loc.gov/standards/marcxml.

JLIS.it 8, 3 (September 2017)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12415

Also the Internet environment might yield different configurations for data exchange, for internal storage, and for input interfaces and techniques. In the MARC environment the internal storage bore similarities to the format, so that the data could be output without loss. And the input interfaces also had the flavour of the MARC field and subfield tags. It is likely that exchange formats will be less obvious at the interface level in the BIBFRAME environment.

A general requirement was that whereas traditional linking had been done with textual data strings and identifier, URIs had to be the linking goal of any new environment. Other goals included accommodation of different types of libraries -- large, small, research, public, and specialized. Since scholarship requires access to the knowledge generated in the past, so existing bibliographic descriptions -- e.g., over 330 million at OCLC – needed to be able to be brought forward to the new environment. It was also realized the change was so enormous and potentially costly that there would be a long transition, so MARC will need adequate maintenance and support for many more years.

BIBFRAME development

The Library of Congress tackled the project via various initiatives. The first was contractual support to develop a high level model for a new environment (*Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services* 2012). Eric Miller, who had worked with OCLC and its Dublin Core initiative and with the W3C during its development of RDF, was the contractor, with his company Zepheira. As soon as his report was complete and published on the web in late 2012, a small group of institutions were invited to work with filling it out and taking it to the next level. This expert group included representatives from OCLC, National Library of Medicine, Princeton, George Washington University, British Library, and the Deutsche National Bibliothek. For a year there was simultaneous consideration of the model by this group and by the community via a listerv.[6] The community was just beginning to take serious notice of linked data requirements for a new environment so the year was one of more questions than answers.

In order to enable the community to have a glimpse of what the BIBFRAME model meant, the Library of Congress developed a "starter" BIBFRAME ontology, taking advice from needs expressed in the MARC environment, from RDA which was just beginning to be implemented, and from RDF conventions that were current at the time. The Library then developed various tools based on that ontology to enable a concrete look at BIBFRAME data: a transformer for MARC to BIBFRAME conversion and a simple editor. The tools were put into GitHub,[7] a public web-based online hosting environment used primarily for program code, for others to download and experiment with them, and the community was very responsive to a request to submit comments and corrections in the GitHub reporting space.

With those tools the library of Congress also mounted a pilot project in 2015 with 40 Library of Congress cataloguers creating BIBFRAME descriptions for multiple forms of material – books, serials, maps, moving images, still images, and music. The pilot was planned for 6 months and had modest goals. Can cataloguers input BIBFRAME descriptions into a BIBFRAME oriented system? The

[6] Bibliographic Framework Transition Initiative Forum. http://listserv.loc.gov/archives/bibframe.html.
[7] https://github.com.

JLIS.it 8, 3 (September 2017)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12415

cataloguers were very positive about the experience. Will type-ahead and drop downs make work easier? These were indeed popular but they also provided data linking URIs without keying them and made input more efficient. Was the labelling on the editor clear and useful? Interestingly the cataloguing staff choose to label the elements on the editor screens with RDA terms where possible, adding also links to the actual rules in RDA. The cataloguers liked this but it also allowed them to ignore somewhat the BIBFRAME data model of Work and Instance and diminished learning whether the Work/Instance dichotomy was clear and useful for cataloguers. Very elementary search capability was provided with a focus on known item search, but the pilot illustrated the necessity to provide better searching for unknown items, especially some forms of browsing. Enabling a bibliographic liaison to add ontology elements as needed the pilot yielded a list of ontology improvements for the BIBFRAME vocabulary.

So the pilot along with the experimentation of others in the community provided important information for the redevelopment of the ontology. Over 200 listerv comments based on examining the ontology and many GitHub comments from experimentation with the tools were reviewed, several RDF and linked data experts were consulted, and pilot participant additions based on actual experience all contributed. As a result of these experiences and reviews the ontology was reworked with positive results. The BIBFRAME 2 ontology is much better integrated with the RDF environment, yet it is also more in synch with the RDA cataloguing rules even while staying rule agnostic. The pilot also showed that as long as the concepts were identifiable in the ontology, the editor could take the approach of using RDA labels and RDA rule links for the Library of Congress pilot cataloguers.

MARC, BIBFRAME and RDA Models

What is the BIBFRAME model and how does it relate to other data models, especially that of RDA? The MARC model is a good starting point for understanding the fundamentals of the BIBFRAME model. The MARC model at a "high level" has three principal components: Bibliographic, Authority, and Holdings. The data in the bibliographic component is a description of the resource being catalogued – both its conceptual (BIBFRAME Work) and its physical embodiment (BIBFRAME Instance). Authority data holds the authorized forms of names of persons, corporate bodies, and conferences, and of subject headings made up of topics, temporals, and places. Its emphasis is on the names of those resources, as the name strings are used to both link resources and to present browse displays to users. But in the United States of America (USA) and some other cataloguing environments it also contains the authorized names for some resource titles, called uniform titles, which relate to works and expressions in the BIBFRAME and FRBR/RDA contexts. These authorized forms of names, subjects, and titles are used in the bibliographic records as needed. And finally the MARC holdings component provides the information on how much an institution holds of the resource and where it is located, along with lending and acquisition information.

The BIBFRAME model, however, moves the bibliographic data closer to the FRBR/RDA view of bibliographic data. The BIBFRAME Work class, bf:Work, which is the conceptual view of a resource, relates to the FRBR/RDA Work and Expression entities. The BIBFRAME Instance class is the physical embodiment of a Work which corresponds to FRBR/RDA manifestations. The BIBFRAME Agent, Topic, Temporal, Place, and Event classes (commonly called name and subject authority data)

JLIS.it 8, 3 (September 2017)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12415

provide information for key concepts with defined relationships to Works and Instances. The Item class, bf:Item, covers the typical information concerning items that are held. Figure 1 illustrates the BIBFRAME model, with references to MARC and RDA counterparts.

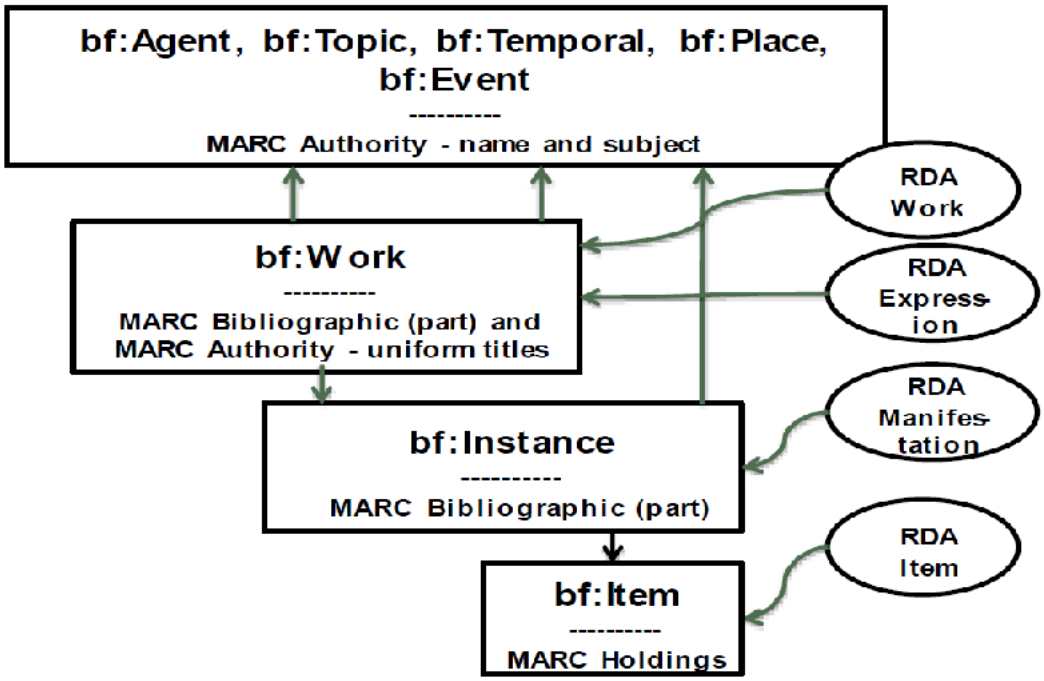

Figure 1: BIBFRAME data model with MARC and RDA mapping

There are several differences from MARC in the BIBFRAME model, and the most prominent is modelling MARC uniform title authority records as BIBFRAME Work descriptions. Treating them as Works is appropriate and it enables them to have subjects associated with them. The different manifestations of a Work are separately described as Instances and then linked to the Work information. Thus all manifestations (BIBFRAME Instances) must have a Work to which to relate. Also in the MARC environment it was common to try to describe on one record the manifestation and all the different carriers for it. With the BIBFRAME model the expectation is that major differences in carriers, such as print and electronic, would be separate Instances of a Work and characteristics of each can then be clearly recorded in the Instance descriptions.

In BIBFRAME there is the intent to treat agent descriptions, i.e., "name authorities" in MARC, as more than a place to record an authorized name label as was noted in an earlier section. So in the case of a person entity, information about the person behind the name would also be available, such as the, occupation, gender, etc. While some of these attributes of the person may also be used as a component of the preferred form of the name (per RDA rules), that choice is up to the cataloguer. Thus a benefit of using the URI rather than the name label string to identify and link entities is that information about the entity is available (in RDF) at the place where the URI points.

JLIS.it 8, 3 (September 2017)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12415

Many questions

BIBFRAME is modelled with a careful look at the various emerging models of the bibliographic community, fitting them into the current world view of the Internet environment using linked data "rules". But some of the common practices for bibliographic data do not match well with the linked data world view. This transition to a more Internet aware format for bibliographic data will bring up issues for how the community carries out its business. One issue will be the use of links. Will there be more linking out to full descriptions or bits of descriptions (such as a language or a subject) rather than bringing the then in and storing them locally? How will indexing and retrieval work if the data is stored remotely? If data is stored locally, can it be kept up-to-date in a more dynamic way? What links can be trusted? The power and limitations of links will have to be worked out with experience and experimentation.

To explore the fit of various media to the basic BIBFRAME model the Library of Congress had two studies prepared analysing moving image and recorded sound material (*BIBFRAME AV Modeling Study: Defining a Flexible Model for Description of Audiovisual Resources* 2014, and *BIBFRAME AV Assessment: Technical, Structural, and Administrative Metadata* 2016). The Library of Congress is particularly interested in these media as it has very large and varied collections of them due to the Library's preservation commitment at its Packard Campus for Audio Visual Conservation in Culpeper, Virginia. The modelling study pointed to a general fit with the BIBFRAME model with an important exception – the treatment of events. Based on this work Events were redescribed to indicate that the class could be a subject of a Work or its content(*Bibframe 2.0 Event Model* 2017). Reconciling those media, and other resources that are very different from "book-oriented" material, with a common model will be a challenge, but this could greatly improve access to those materials.

Another issue or opportunity will be to use multiple authority files when creating descriptions. This could have the potential to reduce the cost of maintaining authority files and, as a "bonus", specialized community authority files might contain more information than libraries can afford to gather. Currently BIBFRAME even enables the specification of multiple authority pointers for one entity, but how many and which ones may not be easy to determine.

As part of its focus on relationships, RDA put more emphasis on specifying the role a person or corporate body played in the creation of a resource. Specification of roles when creating bibliographic descriptions has the potential to enhance end user retrieval and understanding of those descriptions. While the cataloguing community has had an extensive role list for many years associated with MARC, it was only sporadically employed in cataloguing operations. Now with new prominence for roles, the RDA list of roles has been reconciled with the larger MARC list and the new environment could make it easier to apply role information more consistently. However, questions remain. There are other specialized role lists, for example, for moving image, music, and other specialized communities. Do guidelines need to be determined to establish consistency; are hierarchies needed; what are the retrieval aspects of roles from different lists?

Another type of relationship that is also more prominent in RDA and in linked data is between resources such as Work to Work and Instance to Instance. They have been recorded at a high level in MARC (e.g., other edition), with the serial relationships in MARC being the most explicit (e.g.,

JLIS.it 8, 3 (September 2017)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12415

absorbed in part). There is an extensive list in RDA, some with great detail (e.g., musical variations based on). And of course the new environment needs to be able to accommodate new lists from other communities. How much reconciliation of different lists will be needed? How do we want to treat these relationships for retrieval and for display?

There are myriad smaller issues that linked data surfaces for consideration. Should we split BIBFRAME Instances by carrier such as different BIBFRAME Instances for the paperback, hardback, library binding or just by different media: print, electronic and audio? Can we develop authority descriptions for more entities, such a publisher names? One can imagine an environment where publishers maintain such lists of their imprints and make them available as linked data with information about and a standard name for the imprint.

These and other issues should be explored in the coming years as libraries try out and analyse their data from the linked data point of view. Some changes will be needed in the way and what data is recorded, and adjustments to the traditional systems will need to be worked out.

Exploring

There were a number of experiments carried out by institutions in the USA and other countries using the initial BIBFRAME ontology, now called BIBFRAME 1.0. This was a "get our feet wet" period and it inspired institutions to analyse vocabularies, try out conversions, and make data available in a BIBFRAME form. The Library of Congress joined in this with a first Pilot described above and by making its records available as BIBFRAME RDF in response to a Search/Retrieve Protocol (SRU) search of the Library of Congress catalo. The BIBFRAME development also stimulated in the USA more investigation of training options for Linked data and RDF to give technical and bibliographic staff more understanding of these technologies.

Another Library of Congress Pilot

With the release of BIBFRAME 2.0 in 2016, the Library of Congress launched the development of a new Pilot, more ambitious than the first. The goal was to create a Pilot environment that a cataloguer can work in without reference to MARC system. The pilot has two important parts. The first part is the conversion of the whole Library of Congress MARC catalo to BIBFRAME against which the cataloguer will create descriptions of new resources. The other part is a BIBFRAME description creation facility that interacts with the BIBFRAME catalogue.

For the first part an initial step was to map the current MARC catalogue data to the BIBFRAME ontology and create conversion programs to convert it. Much more detailed conversion specifications were constructed than were used for the first pilot, with an attempt not to lose data in the transformation. Then conversion programs were developed that not only convert the data but in many cases supply a URI for data. Even though the conversion of the 19 million record MARC catalogue has been made the published specs and programs continue to be adjusted in response to our own discoveries and those from other who are testing or using the tools. Early in 2017 the specifications

JLIS.it 8, 3 (September 2017)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12415

were published on the BIBFRAME web site[8] and the conversion programs are downloadable from GitHub.[9]

The next and most difficult step for the first part has been to convert this data as fully as possible into the data model preferred by BIBFRAME and RDA. This meant creating Work, Instance, Item descriptions that are linked to Agent, Topic, Place, etc. descriptions. For the Library of Congress where MARC authority files are maintained for uniform titles, those title records were converted to BIBFRAME Work descriptions and then additional Work descriptions were created with appropriate data pulled from the MARC bibliographic records. The Instance descriptions were generated also from the MARC bibliographic records. This was complicated by the fact that MARC cataloguing at the Library of Congress had combined the various Instances of a bibliographic item, e.g., a printed text Instance and an electronic Instance onto a single MARC record. After the Instance descriptions were made from the bibliographic records they were linked to their respective Work descriptions and the subject information from the Instances was transferred from the BIBFRAME Instances to the BIBFRAME Works.

The second part, a description input tool, was at the same time designed for input of descriptions for new material received by the cataloguer. The cataloguer is working with input screens that do not use numbers and codes for data elements, as with the typical MARC-based system input screen, but with labels for the elements. The input screen displays were developed by the bibliographic liaisons and they chose again to use labels for the input screen that corresponded to the terminology used in the cataloguing rules they use, RDA – and as in the first pilot, to link the screen labels to the rules themselves. Cataloguers can create new Instances if BIBFRAME Works are found, or create both BIBFRAME Works and BIBFRAME Instances if not. Unlike the first Pilot, input screens were also created to enable the cataloguers to make descriptions of agents as part of their process, to further simulate a complete BIBFRAME cataloguing environment.

That Pilot began with the 40 cataloguers who participated in the first pilot and the plan to add 20 more in July 2017.

Other explorations

Other institutions are also experimenting with the BIBFRAME tools and ontology with different approaches. A few are indicated below.

Access: A number of projects have been working primarily with access, generally by providing a conversion of a Library's data to BIBFRAME RDF and then offering the descriptions for web access. Colorado College currently leads two interesting projects. One is to convert not only MARC but other data they hold in formats like MODS, Dublin Core, and other XML file formats to BIBFRAME RDF for access across these files. Another converts MARC records to BIBFRAME and then converts BIBFRAME to schema.org for sending to Google. Zepheira is working with a number of libraries, many of them public or state libraries, to convert their data to a simplified form of BIBFRAME and

[8] MARC21 to BIBFRAME 2.0 Conversion Specifications. http://www.loc.gov/bibframe/mtbf.
[9] https://github.com/lcnetdev/marc2bibframe.

JLIS.it 8, 3 (September 2017)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12415

then making it available on the web. The SirsiDynix BLUEcloud Visibility project is another that converts MARC data to BIBFRAME and then makes it available for access.

Workflow: Stanford University and the University of California at Davis have separate projects that focus on workflow. This is also the appropriate category for the Library of Congress Pilot.

Ontology extension: Several universities, Harvard, Columbia, Stanford, and Cornell, are also exploring extensions to the BIBFRAME 2.0 ontology to accommodate details relating to cartographic, still image, music, sound recordings, and other media. Princeton University is working with BIBFRAME and the W3C's new annotation specifications to describe some rare resources. These are important explorations to determine what is necessary, desirable, and affordable for different types of libraries to record.

Data interchange: Casalini Libri, the resource supplier headquartered in Florence, Italy, has a very ambitious project underway, one aspect of which is to supply bibliographic data from their system in BIBFRAME RDF. This is very interesting to USA libraries as many of them, including the Library of Congress, purchase library material from Casalini along with the matching bibliographic records in MARC. Stanford and the Library of Congress are both interested in exploring use of those descriptions in their BIBFRAME projects in the near future.

The National Libraries of Finland and Sweden are also exploring BIBFRAME and providing feedback on the conversion tools from different perspectives.

Going forward

The future of bibliographic control report, *On the Record*, that recommended serious reconsideration of the library data interchange environment came just as the W3C was giving new life to its earlier semantic web concepts. The present paper has described briefly a major component of that new environment, RDF, and the attempt through the Bibliographic Framework Initiative of the Library of Congress to test how linked data could enhance the library experience. The amount of change will be enormous but the Library of Congress is trying by sharing tools, ontology work, and experiences to facilitate the exploration by the community. In just a few years librarians have become much more knowledgeable about the precepts of a linked data scenario, placing them in a position to delve into this change. 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. In the 1960's and 1970's the AACR cataloguing rules and MARC format for bibliographic data were developed. Forty years later we are in the transition to new cataloguing rules and also a new carrier environment, with RDA and BIBFRAME.

References

A Bibliographic Framework for the Digital Age. 2011. Washington, D.C.: The Library of Congress. http://www.loc.gov/bibframe/news/framework-103111.html.

Berners-Lee, Tim. 2006. "Linked Data. Designed Issues." http://www.w3.org/DesignIssues/LinkedData.html.

JLIS.it 8, 3 (September 2017)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12415

Bibframe 2.0 Event Model. 2017. Washington, D.C.: The Library of Congress. https://www.loc.gov/bibframe/docs/pdf/bf2-eventmodel-march2017.pdf.

BIBFRAME AV Assessment: Technical, Structural, and Administrative Metadata. 2016. Washington, D.C.: The Library of Congress. http://www.loc.gov/bibframe/docs/bibframe-avassessment.html.

BIBFRAME AV Modeling Study: Defining a Flexible Model for Description of Audiovisual Resources. 2014. Washington, D.C.: The Library of Congress. http://www.loc.gov/bibframe/pdf/bibframe-avmodelingstudy-may15-2014.pdf.

Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services. 2012. Washington, D.C.: The Library of Congress. http://www.loc.gov/bibframe/pdf/marcld-report-11-21-2012.pdf.

Canadian Library Association, Chartered Institute of Library and Information Professionals (Great Britain), and Joint Steering Committee for Development of RDA. 2010. *Resource Description & Access: RDA*. Chicago: American Library Association.

Denenberg, Ray. 2015. *RDF for BIBFRAME*. Unpublished.

IFLA Study Group on the Functional Requirements for Bibliographic Records, International Federation of Library Associations and Institutions, Section on Cataloguing, and Standing Committee. 1998. *Functional Requirements for Bibliographic Records Final Report*. München: K.G. Saur. http://catalog.hathitrust.org/api/volumes/oclc/39964684.html.

International Council of Museums. 2014. *CIDOC Conceptual Reference Model (CRM). (ISO 21127:2014)*. Geneva: ISO. http://www.cidoc-crm.org/index.html.

ISO 2709:2008. Information and Documentation -- Format for Information Exchange. 2008. Geneva: International Standard Organization.

Library of Congress Working Group on the Future of Bibliographic Control. 2008. *On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control*. Washington, DC: Library of Congress. http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf.

MADS/RDF Primer. 2013. Washington, D.C.: The Library of Congress. http://www.loc.gov/standards/mads/rdf/.

Rust, Godfrey, and Mark Bide. 2000. *The <indecs> Metadata Framework, Principles, Model and Data Dictionary*. London: Indecs. http://www.doi.org/topics/indecs/indecs_framework_2000.pdf.

SKOS Simple Knowledge Organization System Reference. 2009. Cambridge, Mass.: W3C. http://www.w3.org/TR/2009/REC-skos-reference-20090818.