

RDF

WEB

INFORMATION STANDARDS QUARTERLY

WINTER 2013 | VOL 25 | ISSUE 4 | ISSN 1041-0031

ISQ

MARC 21

BIBFRAME

TOPIC

EVOLUTION OF BIBLIOGRAPHIC DATA EXCHANGE

AUTHORITY

ARE CURRENT BIBLIOGRAPHIC
MODELS SUITABLE FOR
INTEGRATION WITH THE WEB?

A TRANSFORMATIVE OPPORTUNITY:
BIBFRAME AT GEORGE WASHINGTON
UNIVERSITY, AN EARLY EXPERIMENTER

INTERVIEW WITH GILDAS ILLIEN , BnF

SCHEMA BIB EXTEND

AUTHOR

ANNOTATION

NISO
How the information world
CONNECTS

BIBFRAME:



NOT JUST WALKING, BUT RUNNING

The BIBFRAME Initiative is a community endeavor led by the Library of Congress to re-imagine and implement a bibliographic environment for a post-MARC world.

BIBFRAME CHRONOLOGY

MAY
2011

LC Announces Bibliographic Framework Initiative
Intended to replace MARC 21

OCT
2011

Framework Plan Published
To be a linked data model with RDF vocabulary

NOV
2012

BIBFRAME Draft Model & Primer
Developed with Zepheira LLC

Early Experimenters:
US National Library of Medicine, the British Library, the German National Library, OCLC, George Washington University, and Princeton University

NOV 2012
-TODAY

Iterative Testing & Discussion Paper Developments
Includes ongoing Vocabulary updates

JAN
2013

bibframe.org Demonstration Website Launched at ALA Midwinter, Seattle
Includes draft Vocabulary & transformation tools

AUG
2013

Updated Discussion Papers
Papers on: Use Cases, Annotations, Resource Types, Authority

2014
PLANS

Experimental Implementations
Software development and demonstrations

Expert Groups
Short-term analysis of lingering issues

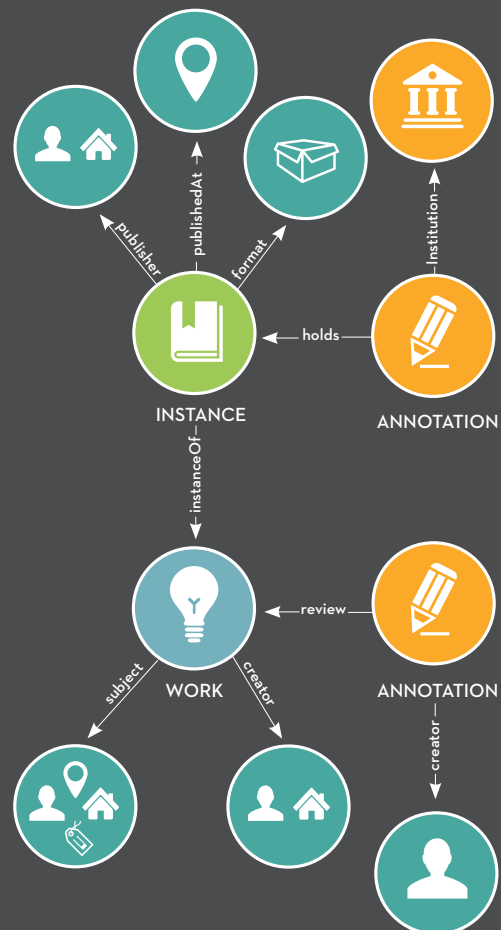
BIBFRAME DATA MODEL CORE CLASSES

WORK: A resource reflecting a conceptual essence of the cataloging item

INSTANCE: A resource reflecting an individual, material embodiment of a Work

AUTHORITY: A resource reflecting an authority concept which has a defined relationship to a Work or Instance

ANNOTATION: A resource that augments another main BIBFRAME class when knowing who asserted the Annotation is vital information



At the advent of 2014, with serious, hands-on work beginning little more than a year and a half ago, BIBFRAME—as an Initiative, as a data model, as a vocabulary—is **not just walking, its running**. For more information: Kevin Ford (kefo@loc.gov), Network Development and MARC Standards Office, Library of Congress

ISQ

WINTER 2013 | VOL 25 | ISSUE 4 | ISSN 1041-0031

NISO

How the information world
CONNECTS

INFORMATION STANDARDS QUARTERLY (ISQ) is a publication by the National Information Standards Organization (NISO). ISQ is NISO's print and electronic magazine for communicating standards-based technology and best practices in library, publishing, and information technology, particularly where these three areas overlap. ISQ reports on the progress of active developments and also on implementations, case studies, and best practices that show potentially replicable efforts.

NISO MANAGING DIRECTOR and ISQ PUBLISHER | Todd Carpenter

ISQ MANAGING EDITOR | Cynthia Hodgson

NISO ASSOCIATE DIRECTOR FOR PROGRAMS | Nettie Lagace

NISO EDUCATIONAL PROGRAMS MANAGER | Juliana Wood

DESIGN | B. Creative Group, Inc.

NISO Publications Committee Members

Marshall Breeding, *Marshall Breeding Consulting Services*

Priscilla Caplan, *Florida State University Libraries*

Liam Earney, *JISC Collections*

Corey Harper, *New York University, Division of Libraries*

Helen Henderson, *Information Power Ltd.*

Sheila Morrissey, *ITHAKA*

Peter Murray, *LYRASIS*

Andrew Pace, *OCLC*

Kristen Ratan, *Public Library of Science (PLOS)*

Susan Stearns, *Boston Library Consortium*

Elisabeth Winter, *Georgia Institute of Technology Libraries*

2014 AD RATES

	1 ISSUE	2-3 ISSUES	4 ISSUES
Full page (8.5" x 11")	\$375	\$350	\$325
Half page (4.25" x 5.5")	\$250	\$225	\$200
Back cover (8.5" x 11")	\$700	\$600	\$550
Inside Cover Front (8.5" x 11")	\$500	\$450	\$400
Inside Cover Back (8.5" x 11")	\$500	\$450	\$400

For more information on advertising, visit www.niso.org/publications/isq/ads/

©2013 National Information Standards Organization. [Authors retain the copyright for the content of their articles and all rights beyond the agreed-upon license with NISO to use their work.]

REUSE: For permission to photocopy or use material electronically from Information Standards Quarterly, ISSN 1041-0031, please access www.copyright.com or contact Copyright Clearance Center, Inc. (CCC) 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users.

CONTENTS

i LC BIBFRAME: Not Just Walking But Running

2 2014 NISO Educational Programs

4 Introduction from the Guest Content Editor

FE FEATURE 6

6 Are Current Bibliographic Models Suitable for Integration with the Web?

OP OPINION 14

14 Replacing MARC: Where to Start

IP IN PRACTICE 17

17 A Transformative Opportunity: BIBFRAME at the George Washington University, an Early Experimenter

22 Interview with Gildas Illien, Director, Bibliographic and Digital Information Department, Bibliothèque nationale de France (BnF)

30 Schema Bib Extend

NR NISO REPORTS 33

33 Charting a Course through a New Exchange Environment: The NISO Bibliographic Roadmap Initiative

NW NOTEWORTHY 35

35 COUNTER Releases First PIRUS Code of Practice for Recording and Reporting Usage at the Individual Article Level

36 AAP Publishes EPUB 3 Implementation White Paper

36 ANSI Launches Online Portal for Standards Incorporated by Reference

SD STANDARDS IN DEVELOPMENT 37

NISO

“Wonderful topic
and perspectives.
Very interesting
and relevant.”

Webinar Subscription Package Discounts

- » Buy 5 NISO webinars
& get 4 free
- » Buy 9 NISO webinars
& get 5 free - attend
the whole series
- » Buy 4 NISO virtual
conferences & get 2 free

JANUARY

- 8 From Device to Device: Adaptive Interfaces for Content (*Webinar*)

FEBRUARY

- 12 We Know it When We See It: Managing "Works" Metadata (*Webinar*)
- 19 The Semantic Web Coming of Age: Technologies and Implementation (*Virtual Conference*)

MARCH

NISO Two-Part Webinar: The Infrastructure of Open Access

5 **Part 1:** The Infrastructure of Open Access: Knowing What is Open

12 **Part 2:** The Infrastructure of Open Access: Toward a Functioning Business Ecosystem

APRIL

- 9 Back from Marrakesh: Implementing an Accessible Content World (*Webinar*)
- 23 Dealing with the Data Deluge: Successful Techniques for Scientific Data Management (*Virtual Conference*)

MAY

- 14 Getting to the Right Content: Link Resolvers and Knowledgebases (*Webinar*)
- 21 Playing the Numbers: Best Practices in Acquiring, Interpreting, and Applying Usage Statistics (*NISO/NASIG Webinar*)

“One of the best webinars I've ever attended.”



www.niso.org/news/events

2014 EDUCATIONAL EVENTS

JUNE

- 11 Fragmented Publishing: The Implications of Self-Publishing (*Webinar*)
- 18 Transforming Assessment: Alternative Metrics and Other Trends (*Virtual Conference*)

JULY

Summer Break

AUGUST

- 13 Streamlining and Simplifying: Advances in Consortial Licensing (*Webinar*)

SEPTEMBER

NISO Two-Part Webinar: E-books for Education

- 10 **Part 1:** E-books for Education: Electronic Textbook: Plug in and Learn
- 17 **Part 2:** E-books for Education: Open Textbook Initiatives
- 24 Library Data in the Cloud (*Virtual Conference*)

OCTOBER

- 15 21st Century Resource Sharing: Which Inter-Library Loan Standard Should I Use? (*Webinar*)
- 23–24 Web Technologies as a Distribution Platform: Implications and Opportunities (*In-Person Meeting with Live Streaming*)

NOVEMBER

- 12 Keyword Search = "Improve Discovery Systems" (*Webinar*)
- 19 Can't We All Work Together? Interoperability & Systems Integration (*Virtual Conference*)

DECEMBER

- 3 Connecting the Library to the Wider World: Successful Implementations of Linked Data (*NISO/NFAIS Virtual Conference*)
- NISO Two-Part Webinar: Sustainable Information
- 10 **Part 1:** Sustainable Information: Digital Preservation for Text
 - 17 **Part 2:** Sustainable Information: Digital Preservation of Audio-Visual Content



“Terrific presentations by all, and directly keyed in to what many libraries are going through.”



NISO Open Teleconferences



Join us each month for NISO's Open Teleconferences—an ongoing series of calls held on the second Monday of each month as a way to keep the community informed of NISO's activities. The call is free.



INTRODUCTION FROM THE GUEST CONTENT EDITOR

It is with great pleasure that I introduce this issue on the Evolution of Bibliographic Data Exchange—full of thoughtful and informative articles describing new metadata initiatives across the library landscape. As guest content editor, I am fortunate to have a set of keen observers who have applied their knowledge, experience, and critical thinking skills to help bring us insight into the background and compelling forces for change. I am grateful to all of the authors for taking the time and effort to share their ideas with ISQ's readers.

As is clear from the themes throughout the issue, the success of the web as a research tool has dramatically changed the library's role in the exposure of library catalogs. As librarians have increasingly professionalized and improved the core mandates of selection, acquisition, preservation, and description of library collections, there has been a corresponding fracturing and loss of effectiveness in another of our responsibilities: exposure. The user has generally moved away from the library catalog as the tool used early in the research process—it is now used, if at all, as a source for availability or fulfillment in the last mile of the research process.

A companion theme throughout this issue is the widespread recognition that our current model for data exchange between library organizations has outlived its usefulness and is ripe for replacement with something with lower barriers to entry for library developers and partners.

Imperatives for Data Exchange

The rise of new metadata initiatives reflects the need to respond to this change and to increase our effectiveness in the exchange and management of library metadata.

As we proceed, we need a metadata model that allows us to achieve the following outcomes:

- 1 Effective exposure of library collections on the web
- 2 Efficient sharing of data between libraries and library organizations
- 3 Promotion of data quality to enable effective library workflows

The key word in the first item is “effective.” Our goal should be to find methods that will maximize the full disclosure of unique and commodity library collections on the web. That includes taking risks with the formats and methods that the web search engines prefer. It also suggests that we respond to the technical requirements of the web by aggregating data whenever possible and using canonical identifiers to make our assets efficiently identifiable in the linked data ecosystem.

The second item echoes the work that libraries have been quite good at over the last half-century—collaborating on standards that facilitate data sharing among library organizations. Our willingness to achieve near-universal adoption of data exchange standards is one of our greatest assets. We can leverage that collaborative spirit as we design the next generation of exchange standards and shed the inefficiencies and high barriers to entry of the current MARC 21 model.

Finally, the last imperative encourages us to look broadly at the data we manage (books, journals, collections, articles, etc.) and welcome new models for managing all of the data assets we care about. Data quality doesn't just mean accuracy; it also means breadth and depth of data. Catalog librarians and library systems developers are comfortable managing the books and journals that information seekers use. It is well past time, though, to recognize that library users care about more than just books and journals. They also care about collections, parts of books, articles, and parts of articles including tables and charts. We must be willing to address the management of the metadata

describing those things and accept the need for new shared methods for exchanging this data. We should also accept the possibility of allowing social input to the management of library data. We shouldn't automatically assume that it's unacceptable for end users to make assertions about our metadata. Social input could improve the accuracy of both the metadata itself and the relationships between elements, such as manifestation clustering and collection memberships. We have a tremendous opportunity to build on our expert communities of practice and the vast potential of motivated end users.

Themes: “One size can't fit all”

The articles in this issue all echo two common themes. First, that the mandate to effectively expose our data on the web calls for changes in the way we describe and manage that data. In our feature article, **Lars Svensson** from the German National Library reminds us that: “The bibliographic world still very much mirrors card catalogs. The problem is that the card display was not built around the concept of pivot points (e.g., authorities) but for sequential display organized according to certain criteria (title, headings)... To enable a better integration into modern, web-based workflows—be it the identification of a book for private reading or the construction of a bibliography for a PhD thesis—it is important that library (meta) data is not only available on the web, but really an integral part of it.” **Jackie Shieh** from the George Washington University echoes that reminder and writes: “In the last two decades, information professionals have been under pressure to remain relevant in the world of web data. Information professionals, in particular those who provide bibliographic description, have had to rethink and retrain themselves in the face of a new data service model for the records that they create and curate.” **Richard Wallis** (OCLC) endorses the call for change in his description of the Schema Bib Extend Working Group.

The second theme to emerge in this set of articles, and the one that I hope is a contribution to the dialog about library data exchange, is best expressed in the BnF Director Gildas Illien's response to one of my interview questions: “In the past 40 years, be it with MARC or other formats such as Dublin Core, we have experienced the limitations of trying to answer all functional and community requirements with

a single format or implementation scheme. *One size can't fit all and doesn't need to.* [Emphasis added.]...I would say we are ideally looking for a scenario where we could meet the joint requirements of a) internal metadata management, including the management of legacy data not only for descriptive purposes, but also for digitization, rights management, and long term preservation of collections; b) rich bibliographic data exchange services with no loss of granularity in description; and c) standard data exchange and exposure on the web the people and search engines use.”

Paul Moss from OCLC is more emphatic on this point when he writes: “The library is not in a position to define its own standard for interoperability with those [search engine] players, but rather should accept that the price of getting their materials in front of users is to do what is necessary to get where the users are.”

The suggestion that we will need multiple exchange models or layered exchange models for different use cases offers a pragmatic recommendation for the way forward. Our task is to develop these models in an orderly and efficient way. If we do, we can maximize the potential of libraries to provide the information needs of library users while avoiding ineffective and costly responses to current demands.

A Modest Hope

It is my hope that this set of thoughtful essays provides you with some insight into the landscape of new metadata initiatives. Indeed, it is my hope that this is a useful continuation of the dialog on how we can improve data exchange and that we see more recommendations and experiments inspired by the pragmatic and optimistic spirit of these authors. doi: 10.3789/isqv25no4.2013.01



Ted Fons | Executive Director, Data Services & WorldCat Quality at OCLC



ARE CURRENT BIBLIOGRAPHIC MODELS SUITABLE FOR INTEGRATION WITH THE WEB?



Libraries are traditionally seen as the gatekeepers to information. A defined process guides the selection of which information enters the library and the cataloging process creates the metadata necessary for the discovery of (non-digital) resources. The advent of the World Wide Web and full-text search has been a game changer in that online publications and resources are better incorporated into the major general-purpose search engines than is (non-electronic) library material.

To enable a better integration into modern, web-based workflows—be it the identification of a book for private reading or the construction of a bibliography for a PhD thesis—it is important that library (meta) data is not only available on the web, but really an integral part of it,¹ thus helping to build what Tim Berners-Lee calls the Giant Global Graph.² Given the structure and rich interlinking of this information, an obvious option to realize this is to publish it as linked data. »



The publication of library data as linked data not only helps search engines to improve the findability of library resources, it also makes library data (authorities and bibliographic information) more accessible to organizations outside of the library sphere.

The publication of library data as linked data not only helps search engines to improve the findability of library resources, it also makes library data (authorities and bibliographic information) more accessible to organizations outside of the library sphere. Following the lead of Kungliga biblioteket (the Swedish National Library),³ several libraries—e.g., Országos Széchényi Könyvtár (National Széchényi Library, Hungary),⁴ the Bibliothèque nationale de France (French National Library, BnF),⁵ the British Library,⁶ the Biblioteca Nacional de España (Spanish National Library),⁷ and the Deutsche Nationalbibliothek (German National Library),⁸—and library service centers (such as OCLC⁹ or the German library networks) have sparked projects and initiatives to include bibliographic information in the linked data cloud. The transformation of traditional, records-based bibliographic data to RDF¹⁰ made it necessary to deal with the actual semantics of the elements of a bibliographic description. Whereas the translation of some elements was fairly straightforward, other elements posed a major difficulty and revealed that we often are at odds with what bibliographic information actually *is* and that the bibliographic universe lacks an agreed-upon model. Such a model would have large advantages when it comes to explaining the structure and the value of this information to non-librarians and would also simplify interoperability with data adhering to other models. Currently, however, the main discussion in the library community seems to focus more on the formats (e.g., MARC 21¹¹) than on an underlying model that can be expressed/serialized in different ways. This focus on the format is insofar counter-productive in that it tends to encourage the use of literals (strings) without analyzing what the information is about and how it relates to other pieces of information (things)—within or outside of a specific bibliographic description. Further, the preoccupation with data in the context of a particular format tends to prevent real innovation, since it is more focused on carrying the existing data forward than on analyzing which data would be necessary for what operation. A shift to a more model-driven view on bibliographic information would increase the possibilities to interlink the individual parts of a bibliographic record to other entities outside of the library domain, particularly within the cultural heritage sector, but also in settings like academia and e-commerce.

The bibliographic data itself

The bibliographic world still very much mirrors the card catalogs. The problem is that the card display was not built around the concept of pivot points (e.g., authorities) but for sequential display organized according to certain criteria (title, headings). ISBD,¹² the format for sharing bibliographic information in a standard, human readable form, has an inner structure and groups the description elements into eight distinct areas composed of multiple elements. But it still focuses very much on the bibliographic record and does not build on an explicit model based on entities and their relations. Many linked data representations of bibliographic data—e.g., the recently published DINI-KIM recommendation for the RDF representation of bibliographic information¹³—still mimic the traditional record-based structure and are more an application profile aiming to provide an easy-to-implement bridge from the library world into the linked data domain than an actual bibliographic model.

There are currently several initiatives working on creating a recognized model for bibliographic information. The most well-known is probably IFLA's Functional Requirements for Bibliographic Records (FRBR)¹⁴ where the entities in the bibliographic universe are first separated into three groups (bibliographic, authority, and topic) and then within the first group into work, expression, manifestation, and item. FRBR is a well-recognized model that was developed from the user tasks of find, identify, select, and obtain. The model is not without problems and there is work underway in IFLA to improve it and also to harmonize it with the other members of the IFLA FR* family: Functional Requirements for Authority Data

(FRAD)¹⁵ and Functional Requirements for Subject Authority Data (FRSAD).¹⁶ Nonetheless, the FRBR approach to group elements and properties common to different versions of the same publication obviously struck a chord with the semantic web community as shown by the transformation of FRBR into RDF by Ian Davies and Richard Newman in 2005.¹⁷ The FRBR model was later adopted by the upcoming cataloging code RDA,¹⁸ and the European Commission's CESAR service¹⁹ uses FRBR concepts to model the publication of semantic assets in different revisions and formats. Further, research has shown that users intuitively relate specific abstractions of a bibliographic description to the appropriate FRBR group 1 entity.^{20,21} RDA is currently in the process of defining relations between the entities that go beyond what FRBR specifies and given that the archives' community is interested in adopting RDA standards, RDA has the potential to serve as a common foundation for data models in the cultural heritage communities.

Another major initiative for modeling cultural heritage data is CIDOC-CRM²² which is an event-based model originally designed for museum materials. There has been work undertaken to harmonize FRBR and CIDOC-CRM through FRBRoo, "a formal ontology intended to capture and represent the underlying semantics of bibliographic information and to facilitate the integration, mediation, and interchange of bibliographic and museum information".²³ Even if some institutions use CIDOC-CRM (e.g., WissKI^{24,25}) and FRBR (e.g., BnF) as models for their electronic services, it is important to bear in mind that both are *conceptual* models, and that it might not be intended to implement them verbatim. Instead we should look at them as what they are—models—and discuss what elements and relations are useful in which context, as in the Europeana Data Model (EDM)²⁶ used by Europeana²⁷ and serving as the basis for the data model of the German Digital Library (DDB),²⁸ and how we can encode the instances of our models in an interoperable fashion using widely agreed-upon exchange formats.

The conflation of model and exchange format becomes very visible in the work of the BIBFRAME initiative.²⁹ The primer declares that a "major focus of the initiative will be to determine a transition path for the MARC 21 exchange format to more Web based, Linked Data standards" and talks about the initiative as "Bibliographic Framework as a Linked Data Model".³⁰ In the introduction it is stated that the "goal of this initial draft is to provide a pattern for modeling both future resources and bibliographic assets traditionally encoded in MARC 21." Indeed the intention seems to be to create a complete replacement to MARC 21 as a format, both as an exchange format, as a cataloging format, and as the internal format of integrated library systems.³¹ Further, BIBFRAME is



In order to transport instance data adhering to other models, the plan is to create community profiles that map the community model to the BIBFRAME model.

intended to be both "rule agnostic" (i.e., not tied to a particular cataloging code) and "model agnostic" (i.e., flexible enough to accommodate both "flat" record-based as well as highly interlinked FRBRized data). But whereas the first of those two is relatively easily achieved, the de-coupling from any specific model is questionable. The BIBFRAME architects have chosen not to adopt FRBR (the reasons are not completely clear, but it seems that they consider FRBR too complex for the Semantic Web) but instead they have created their own model based on the entity types Creative Work, Instance, Authority, and Annotation. In order to transport instance data adhering to other models, the plan is to create community profiles that map the community model to the BIBFRAME model.³² A complete round-trip transformation of data between two models, however, is only possible if both models are equally granular and their entity types and relations have (approximately) the same semantics (in which case it is questionable why there are two different models in use). If not, there will be a loss of specificity when transforming in either direction. Within BIBFRAME, the focus seems to be equally on the format and on the model. This is not explicitly stated in the BIBFRAME documents, but the use of concrete XML syntax to illustrate core concepts and relations gives the format (syntax) an unnecessary emphasis that occasionally puts the actual model in the background. The use of XML instead of RDF serializations (e.g., RDF/XML³³ or Turtle³⁴), since "support for RDF is not yet as widespread as support for XML,"³⁵ is a valid argument when looking at actual implementations and data transfer. If the intention is to focus on the model, however, it would be preferable to have a graphic notation showing the entities and how they are connected and give examples for how this construct can be expressed in several serializations, including at least one RDF syntax.

CONTINUED »



In the LoC-NAF, John Lewis Burckhardt is both a name (*madsrdf:PersonalName*) and a *skos:Concept*. This is in line with cataloging tradition, but a non-librarian would be surprised that the *rdf:type* is not, for example, *foaf:Person*.

CONTINUED »

Another discussion of entity types and their relations is currently taking place within the scope of Schema.org bibliographic extension group.³⁶ In contrast to most library initiatives that model top-down, Schema.org takes a bottom-up approach when incorporating new resource types into their vocabulary. The discussion within the group focuses on what constitutes a specific entity type (e.g., eBook), what are its specific properties, and what properties does it have in common with other entity types so that they can be generalized to a higher level in the hierarchy. It has been argued that the schema.org ontology “is deep enough to create rich and subtle descriptions of many library resources and the events that impact them,”³⁷ which might be true or not, depending on whom the data is intended for: e.g., the bibliographic description necessary for a national bibliography is different from the one needed for a freshman course reading list.

Authorities

A case where library models sometimes differ from what customers might expect are the non-bibliographic items such as people, places, and things that bibliographic descriptions often rely on—the authorities. In the Anglo-American cataloging tradition, the role of the authority is to provide a unique name or a unique heading for an entity that can then be used consistently throughout the catalog. Since the advent of electronic cataloging, libraries, library networks, and (national) bibliographic agencies have collected authorities into authority files that were first distributed directly to interested parties (e.g., other libraries) on magnetic tapes and now increasingly are published in RDF in order to make the data reusable for parties outside of the library sphere.

Those RDF-based authority services are often a core part of a library’s linked data service since the authority data acts as a hub for all information relating to a specific entity (e.g.,

a person or a particular topic). One example is the Library of Congress’s service id.loc.gov,³⁸ where the LoC publishes a rich set of the commonly-used authority data and value vocabularies that it maintains. An inspection of the site and of some of the descriptions reveals several points where the model used differs from what the non-library community might expect. As an example we can look at the following piece of RDF about the publication *Travels in Nubia* by John Lewis Burckhardt:

```

dnb:956706967 a bibo:Book ;
  dc:title "Travels in Nubia"@en ;
  dct:creator lc-naf:n50045595 ;
  dct:subject lc-naf:n81103291 .
lc-naf:n50045595 rdfs:label "John Lewis Burckhardt"@en .
lc-naf:n81103291 rdfs:label "Nubia"@en .

```

Without deeper knowledge of the library domain, a developer would intuitively assume that `lc-naf:n50045595` identifies a person (books are written by people) and that `lc-naf:n81103291` identifies a place (in this case Nubia). The actual data, however, reveals another world-view:

```

lc-naf:n50045595 a madsrdf:PersonalName,
  madsrdf:Authority, skos:Concept ;
  madsrdf:authoritativeLabel "Burckhardt,
  John Lewis, 1784-1817"@en ;
  madsrdf:hasExactExternalAuthority
  <http://viaf.org/viaf/sourceID/
  LC%7Cn+50045595#skos:Concept> ;
  madsrdf:identifiesRWO [
    a madsrdf:RWO , foaf:Person .
  ] .

```

In the LoC-NAF, John Lewis Burckhardt is both a name (`madsrdf:PersonalName`) and a `skos:Concept`. This is in line with cataloging tradition, but a non-librarian would be surprised that the `rdf:type` is not, for example, `foaf:Person`. There is a hint in the description that the entity described identifies a RWO (real world object) of type `foaf:Person`, but in order to find the description of that person you need to follow the link to the external authority in VIAF.³⁹

```
<http://viaf.org/viaf/sourceID/
LC%7Cn++50045595#skos:Concept> a skos :Concept ;
skos:prefLabel "Burckhardt, John Lewis, 1784-1817" ;
foaf:focus <http://viaf.org/viaf/59176329> .
<http://viaf.org/viaf/59176329> a foaf:Person,
rdaGr1Entities:Person .
```

The description of Nubia in the LC NAF might be even more confusing to a non-librarian since the main `rdf:type` given is `madsrdf:Geographic` which suggests that the URI `lc-naf:n81103291` identifies a geographic area. Again, however, `lc-naf:n81103291` is a `madsrdf:Authority` and a `skos:Concept` and it is only through the link to VIAF that we can find out that it is linked to a `dbpedia:Place`.

Another approach was taken by the German National Library (Deutsche Nationalbibliothek, DNB) and the German library networks when in a cooperative project they revamped the authority files used in the German-speaking countries. Until April 2012, descriptions about persons, topics, geographic areas, corporate bodies, and work titles were kept in four separate authority files. When designing the new, common authority format for the Integrated Authority File (Gemeinsame Normdatei, GND),⁴⁰ one of the requirements was that the data model should be directly reusable in the DNB's linked data services in order to expose the information in the authority file better on the web and allow third parties to more easily reuse that information. The result of the design process was an entity-based model featuring seven different types: Corporate Body, Conference or Event, Topic, Work, Place or Geographic Name, Personal Name, and Person. A core feature of that model is that the URIs for the entities in the GND identify, as far as possible, the world objects (e.g., persons, places, or corporate bodies) and only for subject headings, for example, the authority is a concept instead of the RWO. Further, the model reduces redundancy in that there is only one record (and thus one URI) for each entity regardless of the number of roles it can occur in. As an example, we refer to the German author Hermann Hesse using the same URI both for Hesse as an author and for Hesse as the topic of a PhD thesis or a biography.

Using the GND data model, the representation of John Lewis Burckhardt makes it obvious that the authority record is about the real person and does not only model his name, which is more in line with the general expectation that if you dereference a URI pointing to a book's author, you will retrieve a representation of the actual person (or agent), not only a representation of its name:

```
gnd:118702203 a gndo:DifferentiatedPerson ;
owl:sameAs <http://viaf.org/viaf/59176329> ;
gndo:preferredNameForThePerson
"Burckhardt, Johann Ludwig" .
gndo:DifferentiatedPerson owl:subClassOf
gndo:Person .
gndo:Person owl:equivalentClass foaf:Person ,
rdaGr1Entities:Person .
```

BIBFRAME introduces a so-called lightweight abstraction layer for representing authorities in their model. Those authorities are local to a specific library's data but can be linked to other commonly used authority providers such as `id.loc.gov`, GND, or RAMEAU.⁴¹ The authorities in BIBFRAME's lightweight abstraction layer are identified by URIs but again the authority's URI does not denote the real thing; it only denotes the authority record, thus introducing an extra, non-intuitive level of indirection. Since BIBFRAME is still very much a work-in-progress, it cannot be anticipated if there are plans to switch to a different view on authorities. On the other hand, the model used by VIAF very nicely bridges the two views of how to represent real-world objects in library data. For people from outside the library domain, however, a common model would simplify the understanding.

Discussion

The library community needs to enter into a deeper discussion on the actual semantics of bibliographic descriptions. In order to create descriptions where the various parts can be reused outside of the library domain, we need an entity-centric model based on real-world objects (as in the GND) and not on traditional library authorities. In order to find a suitable model for the bibliographic information, more research is necessary. We can expect that different serialization formats with various levels of granularity will be necessary depending on the target application, but in order to retain interoperability a common

CONTINUED »



It is noticeable that the most problematic entity in the FRBR model—the expression—also is one that is core to the user task “find”: users often search for a specific text in a certain language and then in the next step pick the edition (manifestation) of their choice, be it hardcover, paperback, or e-book.

CONTINUED »

conceptual model will be necessary. It has been argued⁴² that we can make models interoperable by using vocabulary alignment and rdfs/owl reasoning but this requires that the semantics of the aligned elements are very similar in order to achieve true interoperability and then the question is if it would not be better to reuse the other vocabulary anyway. Further, the applicability of this approach has not been tested in a large-scale setting where data adhering to several different models is brought together.

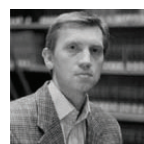
The BIBFRAME approach is so far very promising in that it thoroughly analyzes the existing data and builds its model from that. The discussion, however, seems too much focused on replacing MARC 21 as both a cataloging and an exchange format instead of analyzing the elements of bibliographic descriptions in the light of its constituting parts and entities. It is noticeable that the most problematic entity in the FRBR model—the expression—also is one that is core to the user task “find”: users often search for a specific text in a certain language and then in the next step pick the edition (manifestation) of their choice, be it hardcover, paperback, or e-book. Further, the FRBR Work level overlaps with work descriptions in authority data and can enhance the value and the reusability of those descriptions.

On the way to the future model, we will have to deal with some elements of bibliographic descriptions where the semantics are extremely fuzzy. The best example is the publication statement (e.g., London: Topographical Society, 1898), which merely is a transcription of information found on a publication’s title page and where the exact meaning of the parts is not clear. It can be argued that in the example above the string “London” refers to the real place London (the capital of the United Kingdom) and that this denotes the place of business of the Topographical Society. When confronted with a publication statement like “Berlin, Heidelberg, New York, NY, London, Paris, Tokyo, Hong Kong: Springer, 1990”⁴³ the question arises if the strings “Berlin,” “Heidelberg,” etc. really denote place names and what is their relation to the publisher “Springer.” A future bibliographic model needs to clarify what entities are involved in the publication statement and if this can be modeled using corporate bodies from a library authority file.

Conclusion

The publication of bibliographic information as linked data has left the laboratory and is increasingly entering a stage where it is part of everyday library operations. The work done so far clearly shows that there is no one-size-fits-all model for bibliographic information. In order to replace the current records-based model with one that allows library information to be reused in other settings and also allows libraries to make better use of data originating outside of the library domain, it is necessary to agree on a common model that reduces the complexity of that data integration. To build such a model, librarians—as the domain experts—need to cooperate with potential data consumers from industry and from other cultural heritage institutions.

IFE | doi: 10.3789/isqv25no4.2013.02



LARS G. SVENSSON (L.Svensson@dnb.de) is Advisor for Knowledge Networking at the German National Library (Deutsche Nationalbibliothek, DNB).

ORCID: <http://orcid.org/0000-0002-8714-9718>

Acknowledgements

The author is very much indebted to Reinhard Altenhöner, Christine Frodl, Julia Hauser, Reinhold Heuvelmann and Brigitte Wiechmann for their comments on this article.

REFERENCES

1. Kett, Jürgen, Sarah Beyer, Mathias Manecke, Yvonne Jahns, and Lars G. Svensson. *The Deutsche Nationalbibliografie as linked open data: applications and opportunities*. Presented at 78th IFLA General Conference and Assembly, Helsinki, Finland, August 11-17, 2012. <http://nbn-resolving.de/urn:nbn:de:101-2012052306>
2. Berners-Lee, Tim. *Giant Global Graph*. <http://dig.csail.mit.edu/breadcrumbs/node/215>
3. Malmsten, Martin. *Making a Library Catalogue Part of the Semantic Web*. DC-2008 - Berlin Proceedings. <http://dcpapers.dublincore.org/pubs/article/view/927>
4. National Széchényi Library (National Library of Hungary) on the semantic web http://nektar.oszk.hu/wiki/Semantic_web
5. Bibliothèque nationale de France (French National Library) semantic data <http://data.bnf.fr/>
6. British Library linked open data <http://www.bl.uk/bibliographic/datafree.html#lod>
7. Biblioteca Nacional de España (Spanish National Library) linked data <http://datos.bne.es/>
8. Deutsche Nationalbibliothek (German National Library) linked data service <http://www.dnb.de/EN/lids>
9. Ted Fons, Jeff Penka, and Richard Wallis. "OCLC's Linked Data Initiative: Using Schema.org to Make Library Data Relevant on the Web." *Information Standards Quarterly*, 2012 Spring/Summer, 24 (2/3): 29-33. <http://dx.doi.org/10.3789/isqv24n2-3.2012.05>
10. Resource Description Framework (RDF) http://www.w3.org/standards/techs/rdf#w3c_all
11. MARC 21 <http://www.loc.gov/marc/>
12. International Standard Bibliographic Description (ISBD) <http://www.ifla.org/publications/international-standard-bibliographic-description>
13. Empfehlungen zur RDF-Repräsentation bibliografischer Daten [Recommendation for the RDF representation of bibliographic data], v.1.0, September 2013. <http://edoc.hu-berlin.de/series/dini-schriften/2013-14/PDF/14.pdf>
14. Functional Requirements for Bibliographic Records (FRBR) <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>
15. Functional Requirements for Authority Data (FRAD) <http://www.ifla.org/publications/functional-requirements-for-authority-data>
16. Functional Requirements for Subject Authority Data (FRSAD) <http://www.ifla.org/news/functional-requirements-for-subject-authority-data-frsad-final-report-published>
17. Davis, Ian, and Richard Newman. *Expression of Core FRBR Concepts in RDF*. <http://vocab.org/frbr/>
18. RDA: Resource Description and Access <http://www.rda-jsc.org/rda.html>
19. Community of European Semantic Asset Repositories (CESAR) <https://joinup.ec.europa.eu/community/cesar/description>
20. Pisanski, Jan, and Maja Žumer. "Mental models of the bibliographic universe. Part 1: mental models of descriptions." *Journal of Documentation*, 2010, 66 (5): 643 - 667. <http://dx.doi.org/10.1108/00220411011066772>
21. Pisanski, Jan, and Maja Žumer. "User verification of the FRBR conceptual model." *Journal of Documentation*, 2012, 68 (4): 582 - 592. <http://dx.doi.org/10.1108/00220411211239129>
22. CIDOC Conceptual Reference Model (CRM) <http://www.cidoc-crm.org/>
23. FRBRoo [FRBR object-oriented] http://www.cidoc-crm.org/frbr_inro.html
24. WissKI <http://wiss-ki.eu>
25. Hohmann, Georg. "Die Anwendung des CIDOC CRM für die semantische Wissensrepräsentation in den Kulturwissenschaften [The use of CIDOC-CRM for semantic knowledge representation in the humanities]." In: Sieglerschmidt, Jörn and Peter H. Ohly, eds. *Wissensspeicher in digitalen Räumen: Nachhaltigkeit, Verfügbarkeit, semantische Interoperabilität. Proceedings der 11. Tagung der Deutschen Sektion der Internationalen Gesellschaft für Wissensorganisation Konstanz [Knowledge storage in digital space: persistence, availability, semantic interoperability]*. Würzburg: Ergon, 2010, pp. 210-222.
26. Europeana Data Model (EDM) Documentation <http://pro.europeana.eu/edm-documentation>
27. Europeana <http://europeana.eu/>
28. Deutsche Digitale Bibliothek <http://www.ddb.de/>
29. BIBFRAME: Bibliographic Framework Initiative. Library of Congress. <http://bibframe.org/>
30. *Bibliographic Framework as a Web of Data [BIBFRAME model primer]*. Library of Congress, November 21, 2012. <http://www.loc.gov/bibframe/pdf/marcl-d-report-11-21-2012.pdf>
31. See for example, the e-mail from Eric Miller on the BIBFRAME list where he agrees to the notions that the BIBFRAME format must have the capacity to function as an internal format, and that it must be a replacement for both carrier and content formats. <http://listserv.loc.gov/cgi-bin/wa?A2=ind1310&L=bibframe&T=O&P=2403>
32. Miller, Eric. *BIBFRAME Community Profiles [slides]*. Presented at LC Bibliographic Framework Initiative Update Forum, June 30, 2013. <http://de.slideshare.net/zepeiraorg/alaBIBFRAME-1c20130630>
33. *RDF/XML Syntax Specification (Revised)*. W3C Recommendation, February 10, 2004. <http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/>
34. *Turtle: Terse RDF Triple Language*. W3C Candidate Recommendation, February 19, 2013. <http://www.w3.org/TR/turtle/>
35. *BIBFRAME model primer*, op. cit. [no. 30], p. 25.
36. Schema Bib Extend Community Group <http://www.w3.org/community/schemabibex/>
37. Godby, Carol Jean. *The Relationship between BIBFRAME and OCLC's Linked-Data Model of Bibliographic Description: A Working Paper*. OCLC, June 2013. <http://www.oclc.org/content/dam/research/publications/library/2013/2013-05.pdf>
38. LC Linked Data Service: Authorities and Vocabularies <http://id.loc.gov>
39. VIAF: The Virtual International Authority File <http://viaf.org/>
40. Integrated Authority File (Gemeinsame Normdatei, GND) <http://www.dnb.de/EN/gnd>
41. RAMEAU: Répertoire d'autorité-matière encyclopédique et alphabétique unifié <http://rameau.bnf.fr/>
42. Dunsire, Gordon, Diane Hillmann, Jon Phipps, and Karen Coyle. *A Reconsideration of Mapping in a Semantic World*. DC-2011, The Hague proceedings. <http://dcevents.dublincore.org/index.php/IntConf/dc-2011/paper/view/52/6>
43. DNB catalog record for cited example: <http://d-nb.info/891684808>

OP

[OPINION]

A judgement formed about something;
a personal view, attitude, or appraisal



Paul Moss

PAUL MOSS

Replacing MARC: Where to Start

The MARC format for transmittal of bibliographic records has been an unparalleled success for interlibrary communication. This success, however, has also brought about a world that deals exclusively in MARC and is inherently bound by its limitations. Several efforts over the years have been made to break free of these perceived limitations, but these efforts often miss the crucial mis-step of MARC-like thinking: that a library interchange format should be the only way to ingest, expose, or build systems around bibliographic data. In order to begin a transition away from MARC, each function MARC serves should be examined independently and may be replaced by a different technology.

The MARC mindset

Over the years, usage of the MARC format has expanded into every facet of libraries and how they operate. For a library to ingest data from outside parties, it requests and even demands MARC records. When a library wishes to expose its collections whether it be in an exported file or via Z39.50 or other means, the basis for the exposed data is MARC. Many library application vendors have chosen to accept the limitations of MARC at the core of their applications by making it their fundamental data model. Everywhere you look in the library and its systems you can find some evidence of MARC data or cataloging rules applied to the data.

It is somewhat reasonable, given the expanse of uses of the MARC format

that any intended replacement of this format would assume that it must be a replacement suited to all of its use cases. This does not necessarily have to be the case though. Modern technologies very often espouse a very clear separation of concerns, such that each component may work together and even be separately improved without affecting the other.

Use cases

There are quite possibly too many different uses of MARC to cover them all in a brief article. I will focus on data exposure, core data model, and library data exchange.

➔ Data Exposure

For these purposes, I am using “exposure” to mean making library

data available to non-library services on the web. The goal of this kind of exposure is clear. Libraries want their users to find the research materials they seek wherever they are. It is a commonly accepted idea that many library users will go to Google or Wikipedia to begin their work. People will tweet links to interesting material to the world or share their research with colleagues on Facebook. This is the world of the web as it exists today and this is the world that the library must break into if it is going to be able to continue to offer services its users care about.

Interestingly, these services often have predefined ways of sharing metadata. Google has recently pushed its schema.org

initiative (along with Bing, Yandex, and Yahoo!). Twitter and Facebook have ways to create Cards, or small snippets of a page, that will be meaningful to users.

These are the de facto standards of the web. Data exposure to non-library services should follow these de facto standards. The library is not in a position to define its own standard for interoperability with those players, but rather should accept that the price of getting their materials in front of users is to do what is necessary to get where the users are. The systems that expose library data must include mechanisms to expose that data using these de facto web standards. Today it is schema.org; tomorrow it will be something else. Library data management and exposure systems must be prepared to follow the trends of the web.

➤ Core Data Model

The core data model seems to be largely where MARC replacement efforts are focused. The MARC record format is one intrinsically based on a model of collapsing all information pertaining to a particular book or other item into a single set of fields which make up a record. There are various reasons why this can be problematic. A study by Tom Delsey for the Library of Congress summarizes this challenge by saying:

In the past decade, the rapid evolution of digital information media and communications networks has posed significant challenges for the continued development and viability of the MARC format. Adapting the format to the demands of this new environment entails more than simple incremental enhancement to format specifications; it requires extensive re-examination of the underlying logical structure of the format and its application.

There is enough consensus in the industry that this must change, that adding my words to it feels like just piling on. Due to the prevalence of MARC formats inside of different facets of the library, making a wholesale change to the data model will be extremely difficult without separating the data model from the rest of the system(s) which use the data.

➤ Library Data Exchange

There is still a need to transmit data between libraries and/or library vendors. And there is still a need to improve upon the way that is done today. One of the problems here is that most providers of books and other materials to libraries do not use MARC as a fundamental data model. This presents problems for libraries to accept their data.

Take, for example, the recently developed KBART recommendations for interchange of electronic resource data. This set of recommendations can be loosely summed up as: Put your data in a spreadsheet and please use this set of column headers. It may be overly simplistic to describe the full richness of library cataloging, but it has a key feature: it does not, in any way, proscribe how to design the producer or the consumer applications. This benefit means that disparate systems, created for different purposes and with different technologies, may talk to one another.

A proposal

What would a standard for interchange of library data look like if it were *only* that? This, I think, is the proper purview of a MARC replacement at this stage. By removing the



People will tweet links to interesting material to the world or share their research with colleagues on Facebook. This is the world of the web as it exists today and this is the world that the library must break into if it is going to be able to continue to offer services its users care about.

CONTINUED »

requirement to be the future of bibliographic description for every purpose and focusing simply on the problem of moving metadata around, we may achieve a state which allows us to transition away from MARC as a representation of bibliographic data.

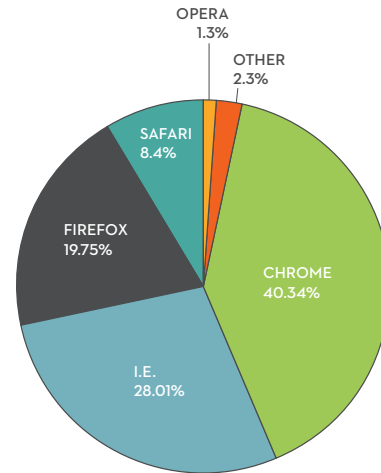
Consider a simple example: A list of books packed into a box and shipped to a single library. The current practice is for the library to obtain, either from the book vendor or third party service, a full MARC record describing these books at roughly the same time the box is received. This creates a coupling between the library system and the supplier of these records. If either party chooses to alter their end to support some alternate representation, then a translation between that format and MARC must occur. A small but very powerful change could be made to this transaction which breaks this coupling. If, instead of transmitting a MARC record, a simple list of identifiers (expressed as URIs) was passed, then the description of which books are in the box is no longer tied to the MARC format.¹ The identifiers may point back to a central service like WorldCat or to a service provided by the vendor, if available.

The difference between these two scenarios is subtle. By abstracting the format out of the equation for simple data interchange use cases, both parties may now be free to adjust their preferences for format in a semi-independent manner. Actually retrieving a usable format of a record or other carrier for including in a local catalog can be done through HTTP content negotiation or other mechanisms. (UnAPI is an example of a more complex mechanism.) This changes the expectations of each party from an agreed upon MARC requirement to one where each expects a range of different formats to be supported and a preferred one decided only at the time the record is required. This type of decoupling is very similar to what allows internet users to update their browsers on an irregular and, importantly, different schedule from the rest of the people browsing the internet.

Altering the mechanism for interchange of bibliographic data in this way could allow a new data model, such as proposed by the current state of BIBFRAME, to be adopted in parallel to existing models for transmitting MARC records. By decoupling the systems, the ecosystem of libraries and vendors and other parties can start to adopt new models alongside old without causing significant disruption.

Conclusion

It is a good thing that libraries are rethinking how we transmit data among ourselves. MARC is unquestionably an artifact of an earlier era. However, in replacing it libraries must understand that it isn't just the complexity of an old format which must be replaced but rather the reliance on a single data format for everything. We must accept that potential users of the library can easily get their needs met elsewhere and instead of fighting to get a library-specific standard supported by the Googles of the world, we should focus on making it so that they don't have to.



This type of decoupling is very similar to what allows internet users to update their browsers on an irregular and, importantly, different schedule from the rest of the people browsing the internet.

In the meantime, we can define our library specific data exchange format without the requirement of being the future representation of all bibliographic data everywhere by simply following the modern concept of a separation of concerns.

| OP | doi: 10.3789/isqv25n4.2013.03

PAUL MOSS (mosp@oclc.org) is the Product Manager of the WorldCat knowledge base with OCLC.

¹ This is a very specific description of one way to decouple these systems. It is, however, only an example. There are other possible implementations that may achieve this end.

Delsey, Tom. *Functional Analysis of the MARC 21 Bibliographic and Holdings Formats*. Library of Congress, January 4, 2002.

www.loc.gov/marc/marc-functional-analysis/source/analysis.pdf

Knowledge Base And Related Tools (KBART)

www.niso.org/workrooms/kbart

MARC

www.loc.gov/marc/

UnAPI

unapi.info/

WorldCat

www.worldcat.org

Z39.50 standard

www.niso.org/standards/z39-50-2003/



RELEVANT
LINKS

A Transformative Opportunity: BIBFRAME at the George Washington University, an Early Experimenter

JACKIE SHIEH

The century's old tradition of the library's mission continues to resonate in the information profession, even in today's fast-moving development of mobile technology.¹ The centrality of this mission is indisputably integral to the user's research experience.

In the last two decades, information professionals have been under pressure to remain relevant in the world of web data.² Information professionals, in particular those who provide bibliographic description, have had to rethink and retrain themselves in the face of a new data service model for the records that they create and curate.

Library communities initiated several projects that attempted to respond to the shifting information landscape and remain relevant to their mission.³ On May 13, 2011, the Library of Congress (LC) issued a statement on transforming the bibliographic framework.⁴ Zepheira⁵ was engaged to spearhead the process of rethinking bibliographic control beyond the MARC communication format in a way that could extend to a wider bibliographic framework—content agnostic,⁶ and able to support traditional bibliographic, authority, and holdings data, in addition to aligning them with services that go beyond traditional information

structures, both physical and virtual. For practitioners—in this case, cataloging professionals—to begin working in this new environment, a change in their understanding of the anatomy of a record must occur. A record consists of various components—author, title, publisher, physical description, etc. To think and work with each component as data instead of text strings is the basis of the revolution. Data can be recognized by machine methods, and connections between data can be made among any resources containing an identifier. These data can be organized or regarded as an assertion or a set of assertions about a resource. These assertions state a named relationship between resources.

BIBFRAME Snapshots

BIBFRAME (Bibliographic Framework) seeks to serve as the foundation for the future of bibliographic description.

CONTINUED »

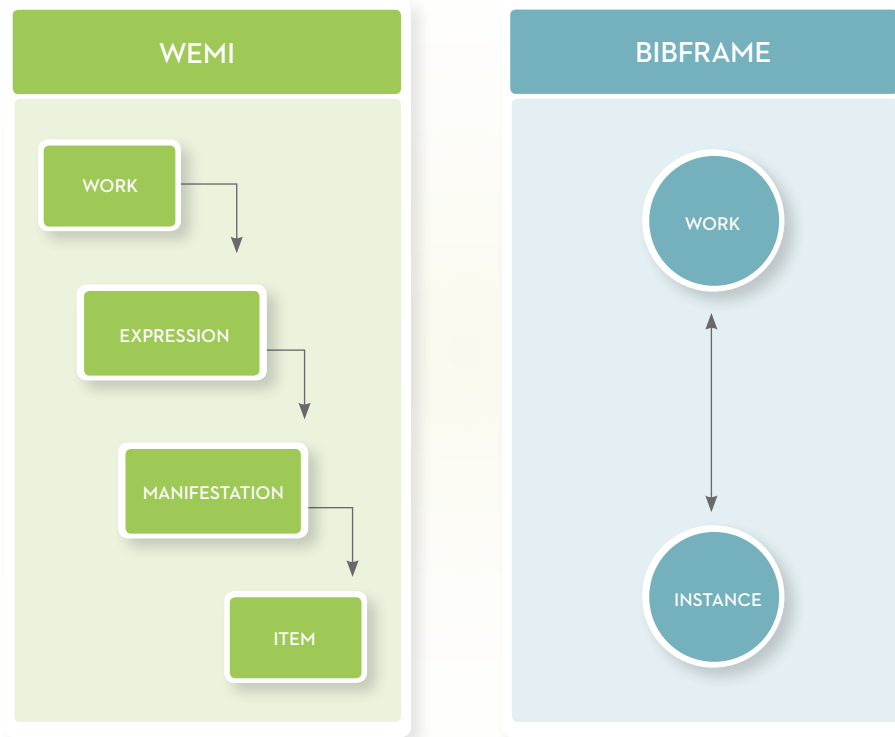


Figure 1: RDA Entities vs. BIBFRAME Classes

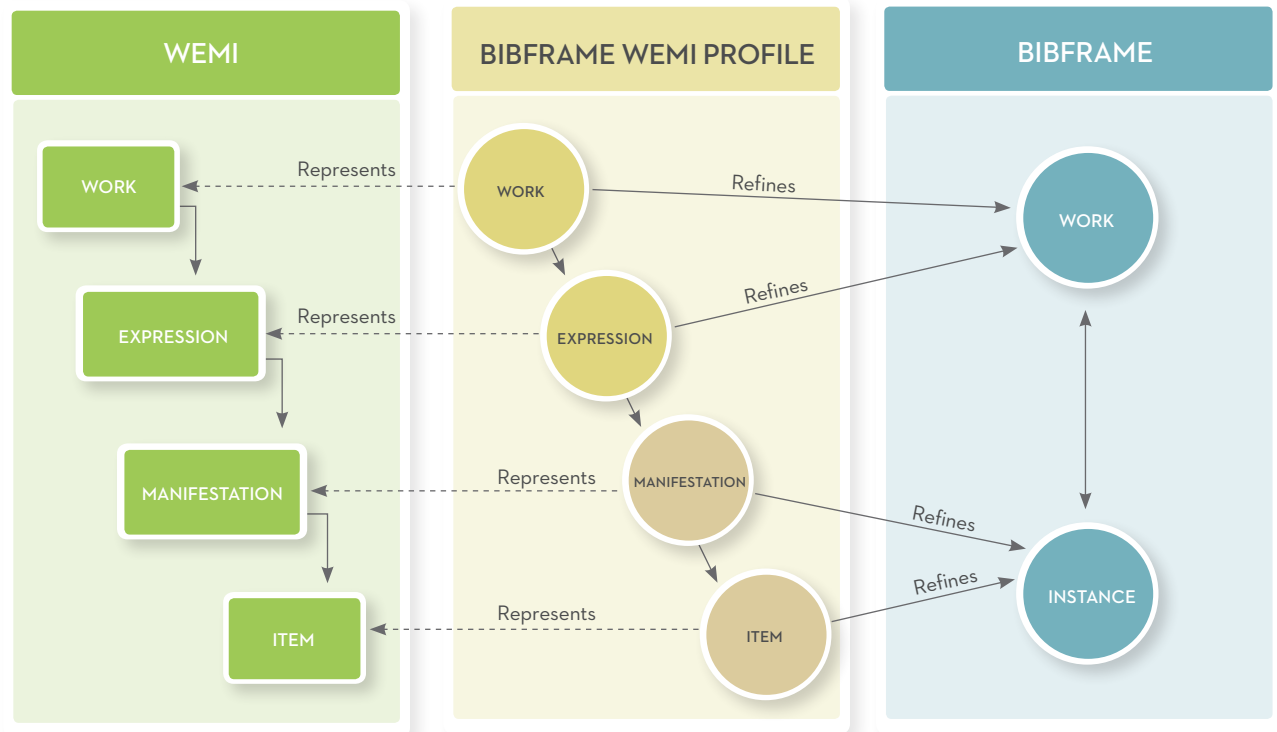


Figure 2: Refinement of WEMI Entities to BIBFRAME Classes through Profile

CONTINUED »

Because its approach to system platforms is agnostic, it hopes to integrate with the wider information community beyond libraries and organizations. The data model employs a linked data conceptual design and language (RDF/XML⁷) that is common for web architecture. A framework consisting of a web of data will leverage the web as an architecture that allows the assembly and reassembly of data defined in higher or granular levels. This model enables the integration of existing bibliographic standards and provides a roadmap toward the development of alternative approaches to information service. The structure organizes data in the following classes: Creative Work, Instance, Authority, and Annotation.

The relationship model is based on FRBR (Functional Requirement for Bibliographic Records)⁸ and RDA (Resource Description and Access)⁹ elements and is expressed as an entity with properties and attributes that show assertion(s) between two links of person, family, corporate body, concept, place, etc. The assertion relays a meaningful interpretation. Figures 1 and 2 illustrate, via an application profile, the concept of entity in the FRBR/RDA environment, and a possible alignment to BIBFRAME classes.¹⁰

Serialization of BIBFRAME RDF model is not locked in such a way that the modeling would impede communication and interoperability of the data. Several models were put in place for demonstration, such as RDF/XML, Turtle,¹¹ and N-Triple¹² in the hope that data points can connect seamlessly. Thus the model design is optimized and serves as a network central that advances data analytics and transforms research simply because it makes interconnectivities among things commonplace.

GW Environment

When the BIBFRAME initiative surfaced in 2012, its design characteristics struck a chord with the George Washington University Libraries (GW) administration: customization, openness, productivity, shareability, and resource development. They also recognized that GW staff could make an important contribution by participating in the initiative.¹³ By being an early experimenter (EE), GW Libraries had a unique opportunity to contribute and establish a new standard that would benefit researchers navigating the information sphere. An institutional commitment to be involved on this scale challenged both the lead participants and library staff members, who were called upon to contribute a portion of their skills and talents to the project. It was a journey for our small group that helped solidify our professional beliefs.

GW's data were created, contributed, and collected over a long period of time, and were migrated from various platforms. Given that situation, it would be unrealistic to expect data consistency throughout the lifecycle, and the possibility existed that these data might be erroneous. The analyses of GW's bibliographic data conducted in its consortial knowledge base, Voyager,¹⁴ validated this assumption.¹⁵

Project Process

The BIBFRAME Initiative established an ambitious roadmap¹⁶—the creation of a test set to be funneled through the Library of Congress and Zepheira pipelines in October 2012. A draft for local adaptive process was prepared by December 2012, and data modeling feedback occurred in January 2013.

CONTINUED »



A framework consisting of a web of data will leverage the web as an architecture that allows the assembly and reassembly of data defined in higher or granular levels.



THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON D.C.

When the BIBFRAME initiative surfaced in 2012, its design characteristics struck a chord with the George Washington University Libraries (GW) administration: customization, openness, productivity, shareability, and resource development.

Compared to other early experimenters, GW's smaller size allowed it to more easily get a team ready to meet the established benchmarks.¹⁷ However, it required a completely different mindset for catalogers, who view their work of describing, recording, and classifying a library item from a holistic angle, with an endpoint being the creation of a bibliographic record. Programmers, on the other hand, interpret a record differently: as data. BIBFRAME's approach is to dissect a record into data components, treating text as data, which can connect with other data in many different ways, on many different levels of granularity.

Learning to assess datasets, from analyses to selection and then transformation, was an excellent opportunity to build staff confidence. At that time, neither cataloging staff nor programmers at GW had needed to immerse themselves regularly in RDF/XML vocabulary and data structure. Possessing both a limited technical and programming skill set, and competing, existing library priorities, GW narrowed its data focus. Staff worked on transforming selective datasets, and examined the results with an eye both on the current "clinical" process and on using this data as building blocks for the future.

GW's modeling used simple bibliographic records of a monographic nature. Data contained mixed publication and creation date ranges, but excluded records describing multiple versions and complex holdings locations. Authority files were considered out-of-scope for this initial phase. Extracted data were placed in Washington Research Library Consortium (WRLC) servers for testing. Figure 3 illustrates the dissection of MARC data and its transformation to the proposed BIBFRAME vocabulary.

Figure 3: GW MARC to BF Transformation

MARC	BIBFRAME
LDR 00885cam a2200301 # 4540	Control
001 2792771	URI
005 20020509142301.0	URI
008 512418191 eng 000 1 eng	URI
010 1# 51052940	URI
019 1#29774405	URI
035 1#0001C0em15525100	URI
040 1#DLC 1#DGV	URI
100 1 #Austin, Jane, 1817?-1817.	URI
245 10 \$Title and prologue / \$Cline Austin ; with an introduc-	URI
650 0 \$Young women \$Fiction.	URI
650 0 \$Courtship \$Fiction.	URI
650 0 \$Sisters \$Fiction.	URI
651 0 \$England \$Fiction.	URI
652 0 \$Domestic fiction. \$2aaf	URI
653 7 \$Love stories. \$2aaf	URI
030 1#0279405429	URI
260 1#New York : \$Bantam / \$Distributed by Random House	URI
300 1#xiviii, 358 p. : \$22 cm.	URI
490 0 1#Everyman's library ; \$v1	URI
049 1#DGVW	URI
050 00 1#PR4034 1#P7 1991	URI
082 00 1#K23- / \$22	URI

Lessons Learned

Aligning tasks closely with existing skills and interests of library staff encouraged GW to envision what it would be like to transition to the BIBFRAME environment. The process recognized the value of building upon simple and less complicated scenarios first, reinforcing staff confidence in order to prepare them for the more complex endeavors ahead. Throughout the process, a learning environment was established, and new relationships among staff were forged and nurtured. Finally, collaborations with other early experimenters helped to discover and plan for skills improvements in addition to strengthening GW's commitment to service within its traditional confines and beyond, encompassing GW's faculty and students in an expanding circle of benefit as library staff continue their engagement in future collaborative projects.

BIBFRAME Next Steps


By Autumn 2013, early experimenters had completed drafting of more than a dozen point papers.¹⁸ Some topics have more than one draft available for public comment.¹⁹ Refinements to the initial pages of the Vocabulary Navigator²⁰ help to apply Work, Instance, Authorities, and Annotation relationships to MARC 21.²¹ Transformation tools that have been in place on the BIBFRAME website²² will become generally available. The group has also begun preparing use cases for public review.²³ Annotations are inserted into BIBFRAME classes to help the end user better understand the intended and potential usages, as shown in Figure 4.

GW invested a great deal of effort and resources in the BIBFRAME project. Library administration's attitude allowed that even if the result ended in an abrupt termination of the project, staff would have gained valuable lessons by participating in the process. In the overall scheme of things, the investment of resources—staff, equipment, time, and skills—will eventually pay off, if not in this direction, then in another venue. BIBFRAME opens up the library world in more ways than one could imagine. The information world, in particular the library world, has been transformed by information exploded out of the book into many formats, some of them as yet unimagined. GW staff, as one of the EEs, had a taste of this shifted change which prepared them to accommodate new approaches.

The proposed BIBFRAME vocabularies and data modeling were tested. Some appeared to have passed and validated the original goal. Stakeholders from diverse information communities actively participated in data

3.1. Books near me (Identify holdings of Instance)

Alex is writing a book report on the "Phantom Tollbooth". Unfortunately, his copy has been inadvertently left on the bus never to be seen again and the report is due tomorrow. In an attempt to solve this "problem", Alex's mom is looking for a copy of this book held by a nearby library.

NOTE: The following use case has been annotated to provide a prose description associated with the same query and response data. This key is included as a means of helping explain this and subsequent use cases. Annotations are reflected through the following icon .

3.1.1. Query Pattern

Alex's Mom clicks on her "Books near me" button to search for the "Phantom Tollbooth". The application knows her current location and has a default query pattern:

```
SELECT ?work ?inst ?lib
WHERE {
  ?work bf:title "Phantom Toll"
  ?inst bf:instanceof ?work
  ?holdings bf:holds ?inst
  ?holdings bf:holdsBy ?lib
  ?lib gs:nearestby(40,1563 83,0742 30) .
}
```

BIBFRAME Work
Search for BIBFRAME Works with the title of "Phantom Tollbooth". For all Works, return all of associated BIBFRAME Instances held by Libraries within 30 miles of my current location.

Figure 4: A BIBFRAME Use Case

modeling and refinements.²⁴ Its adaptability can be extended beyond MARC 21 to UKMARC, UNIMARC, etc. However, replacing MARC format completely as a feature cataloging system is monumental. Any replacement system, whether implemented in the current environment or deployed in a cloud, may take a few months or even years. Prediction is hard. BIBFRAME has made a good start. More awaits.

IP | doi:10.3789/isqv25n04.2013.04

JACKIE SHIEH (jshieh@gwu.edu) is Coordinator, Resource Description Group at the George Washington University Libraries.

ORCID: <http://orcid.org/0000-0003-3214-8846>



The information world, in particular the library world, has been transformed by information exploded out of the book into many formats, some of them as yet unimagined.

REFERENCES & NOTES

- The mission was defined by Charles Cutter in the 19th century—enabling the user to find an item by its creator, title, or subjects in a library catalog having the ability to show its user all items that the library owns by any given author, title, subject, or type of literature. In addition, it guides users' choices by describing an item's edition and characteristics.
- Tim Berners-Lee's 1997 draft on metadata axioms and the conceptual principle of connecting resources on the net with an identifier associated with an actionable protocol laid down a paradigm shift for metadata architecture and standards. www.w3.org/DesignIssues/LinkLaw
- OCLC spearheaded several projects that rallied catalogers to explore potential solutions and attempts to organize the exploded web resources: InterCat (Internet Cataloging, 1994-1996), netFirst (1995), CORC (Cooperative Online Resource Catalog) 1998-2003, etc. In Nov. 2006, the Working Group on the Future of Bibliographic Control convened by the Library of Congress laid out strategies for the next decade, including a bibliographic framework beyond MARC.
- Transforming our Bibliographic Framework: A Statement from the Library of Congress*. www.loc.gov/bibframe/news/framework-051311.html
- Zepheira [website]. zepheira.com/
- Agnostic, in Greek means without knowledge and is commonly understood as a concept that refers to the design attributes and philosophies of a system, software, a framework, etc. It is generalized in such a way that it facilitates interoperability, is free of system, platform constraints, and enables compliance with widely accepted standards.
- RDF/XML Syntax Specification (Revised)*. W3C Recommendation, February 10, 2004. www.w3.org/TR/rdf-syntax-grammar/
- Functional Requirement for Bibliographic Records*. IFLA, 1998. www.ifla.org/publications/functional-requirements-for-bibliographic-records
- Resource Description and Access [website]. www.rda-jsc.org/rda.html
- Miller, Eric *BIBFRAME Community Profile*. Presented at LC Bibliographic Framework Initiative Update Forum, June 30, 2013. de.slideshare.net/zepheiraorg/alabibframe-lc20130630
- Turtle: Terse RDF Triple Language*. W3C Candidate Recommendation, February 19, 2013. www.w3.org/TR/turtle/
- N-Triples*. In: *RDF Test Cases*. W3C Recommendation, February 10, 2004. www.w3.org/TR/rdf-testcases/#ntriples
- In the early 2000s, the George Washington University (GW) Libraries, like many of its peers, began a series of reorganizations to maximize its resources for library service. In 2008, Technical Services began a final phase of personnel and resource realignments. Between 2009 and 2010, GW Libraries cataloging staff participated in the U.S. National RDA Test partners program. Collectively, staff spent 94 personal hours and 17 structured hours in the classroom and in workshops. This experience created a forward-looking, cooperative learning environment.
- Voyager Integrated Library System* [webpage]. www.exlibrisgroup.com/category/Voyager
- Reports from the Voyager database show almost 18% of the records contain at least one or more invalidated fields. This translates to over 300,000 records in a database containing 1.8 M bibliographic records.
- McCallum, Sally. *BIBFRAME: Proposed Roadmap*. Presented at LC Bibliographic Framework Initiative Update Forum, January 27, 2013. www.loc.gov/bibframe/pdf/ALAMw2013-bibframe-draft-roadmap_McCallum.pdf
- Besides GW, the other BIBFRAME Early Experimenters are: British Library, Deutsche Nationalbibliothek, Library of Congress, National Library of Medicine, OCLC, and Princeton University.
- The EEs prepared and shared a dozen point papers from March-May 2013. Those available to the public are posted at: bibframe.org/documentation/
- Bibliographic Framework Initiative. *BIBFRAME Documentation* [website]. Library of Congress. www.loc.gov/bibframe/
- The term was coined in the LC document announcing BIBFRAME [page 39]. www.loc.gov/bibframe/pdf/marcld-report-11-21-2012.pdf
- BIBFRAME Model webpage* [including core classes and vocabulary updates]. bibframe.org/vocab/
- BIBFRAME Tools* [webpage]. bibframe.org/tools/
- BIBFRAME Use Cases and Requirements* [webpage]. bibframe.org/documentation/bibframe-usecases/


 Gildas
Illien

Interview with Gildas Illien, Director, Bibliographic and Digital Information Department, Bibliothèque nationale de France (BnF)

Guest Content Editor, Ted Fons, presented Gildas Illien with a series of questions about the work that the Bibliothèque nationale de France is undertaking to transform bibliographic data exchange and to get his insight on the trends in the European library environment.



Can you summarize your opinion about the need for a new framework for bibliographic data exchange? Why is it necessary now? What is the biggest problem that we need to solve as metadata professionals?

There are many drivers for change in this area. Speaking from a national library perspective, I should start by stating that in a world where public policies and federal or national agencies are being constantly reevaluated, libraries need to demonstrate more output from their cataloging and metadata computing effort than they may have done in the past. The structured information they produce at high cost is expected to be used by more users from communities not restricted to libraries and linked to other data types in order to produce new knowledge and new services to people. The historical mission of national bibliographies remains valid in principle but must be radically revisited within this broader perspective of extended usage in the context of Linked Open Data.

I tend to look at the bibliographic transition we are undertaking mainly from a management angle—which is my role and contribution at the BnF [Bibliothèque nationale de

France]. In this position, my first concern is to evaluate whether our cataloging workforce is successful in serving what should remain its ultimate purpose: access and usage. In this respect, it has become commonplace to acknowledge radical shifts in information research and retrieval practices. Our end users are on the web. They are looking for relevant and trusted information more often than they are looking for specific documents. Fewer and fewer search bibliographic information specifically, nor within the particular boundaries, languages, and applications of library catalogs. Things, People, Places, and Dates need to be expressed in more generic terms and concepts matching web standards and practices. Moreover, when it comes to researchers or corporate organizations, we know their need is no longer about finding and reading documents only, but also about confronting and mining large (meta) datasets using new computing tools.

THINGS



PEOPLE



PLACES



DATES



Things, People, Places, and Dates need to be expressed in more generic terms and concepts matching web standards and practices. Moreover, when it comes to researchers or corporate organizations, we know their need is no longer about finding and reading documents only, but also about confronting and mining large (meta) datasets using new computing tools.

Many librarians feel they are competing with or being defeated by the web, while they should see this new environment and expectations as a great opportunity to promote the information they've been producing and managing for decades: unlike much of what one finds online, bibliographic information is standardized and worthy of trust. National libraries and bibliographies possess an amazing legacy of highly structured metadata that could make a difference in making the web smarter. Library data may look complex from the inside (and actually is) but this complexity, if properly used, could improve the search and discovery end-user experience. In my view, our first priority should thus be to make bibliographic information fully interoperable with the web standards and environment, especially those of the Semantic Web. Bibliographic data exchange transition must be envisioned within this global and digital context, which should certainly have a strong impact on the data exchange modeling and infrastructure we will choose.

Economic constraints (budget and staff cuts) and the continuous growth of the amount and types of publications (both analog and digital) libraries are mandated to handle bring a second driver for change. To summarize what many experience these days, libraries need to do more with less. They can no longer afford the luxury of duplicating efforts and have to rely on much more cooperation, with a variety

of stakeholders. While focusing their domestic production effort on their added value and unique or rare collections and references which will enrich the "long tail" of web contents, they will need to aggregate, confront, match, merge, or link an increasing amount of heterogeneous metadata from various provenances and of different status and quality level. As a result, institutions will have to organize many more data interactions and workflows involving other parties: interactions between libraries of course, but also between libraries and publishers along with other communities such as archives, museums, or research institutions. Many libraries may also consider giving a fresh eye at crowdsourcing in metadata, which will require managing direct interactions of end users with their bibliographic data or bridging their activities with those of powerful collaborative entities such as Wikipedia.

This means that from the original creation of records, metadata specialists will have to evolve as they will be handling more and more tasks designated to import, export, and transform metadata rather than creating it. This may imply outsourcing some of these tasks, sold as services by vendors, and participation in regional or global initiatives and knowledge bases maintained in the cloud for datasets that will not necessarily be made available for free. On the other hand, national libraries and bibliographic agencies will

CONTINUED »



Data.bnf.fr does not mean to replace the existing catalogs and other silos it exploits, but to provide some “glue” between them. In short, it aims at making our library data work better on the web, by delivering a service of information, with structured explicit data and permanent URIs.

need to remain worthy of trust and to maintain public, sustainable and free access to the databases they produce. Standards of quality, transparency, and publicity of the metadata they publish are crucial values they are certainly not ready to give up. This particular tension is to be taken into consideration as well in our vision of future metadata exchanges. There are and there will be even more players than today in the data arena, all with diverse, sometimes conflicting interests, missions, and business plans. In my view, the discussion of the business models capable of accommodating these various interests is also part of the picture we need to keep in mind while designing new data exchange infrastructure schemes.

I can see many technological opportunities to address these issues now, and to take action accordingly. The web of data quickly develops, offering potential solutions to some of these problems, provided professionals accept to move away from library-centric schemes and formats in order to seek better interoperability in a larger environment. It is *now* that libraries need to take position within the web of data if they want to be considered as significant players in this new environment—later might be too late. This is why we need to massively publish vocabularies and bibliographic data now, even if they aren’t as perfect as we would like them to be. From a metadata specialist perspective, I would say the biggest problem underlying all these issues may be: how much are we ready to give up, as libraries, from our added value, from our legacy, from our specificities in order to accommodate such interoperability needs? I believe we will certainly need to change most of our cataloging habits, standards, and tools—which certainly are crucial attributes to a cataloger’s culture and professional identity—but that losing the quality and granularity of the data itself should not be a requirement. What we need to do is to reformulate the information we manage in different terms. In the past 40 years, be it with MARC or other formats such as Dublin Core, we have experienced the limitations of trying to answer all functional and community requirements with a single format or implementation scheme. One size can’t fit all and doesn’t need to. The international community should rather consider developing strategies where various approaches may co-exist.

I would say we are ideally looking for a scenario where we could meet the joint requirements of:

- a internal metadata management, including the management of legacy data not only for descriptive purposes, but also for digitization, rights management, and long term preservation of collections;
- b rich bibliographic data exchange services with no loss of granularity in description; and
- c standard data exchange and exposure on the web the people and search engines use.

Q What has the BnF already done to transform the way you express your bibliographic data?

I think our first challenge in the past years has been to change our general vision and strategy as to the bibliographic transition and to adopt a more pragmatic, perhaps more relaxed attitude as well, finding the right balance between international interoperability dependencies and the need to demonstrate tangible

progress internally and at the national level. We felt the need for change but our initial vision to move forward was very linear. Initially, there was an assumption that, to do things properly, we first had to change the cataloging rules and standards, then envisage actual change of practices and tools for production. It was only at the end of this tunnel that we would eventually envisage how this long-term process impacting many people and involving considerable investments would practically make a difference to the end user. This was too stressful and too risky a process, also a very difficult roadmap to sell to our stakeholders and decision makers.

The BnF is investing heavily in the standardization effort and its best metadata specialists are still very much involved in ISBD, RDA,¹ and FRBR work, together with the national and international community. However, we are now looking at things the other way around. Our current priority is to work on the actual diffusion of our legacy data in order to achieve convincing and visible results in terms of web exposure and service. This has involved launching large data transformation campaigns of our catalogs, and supporting innovation efforts through various channels, always following the FRBR principles. Launching proofs of concepts, evaluating them, analyzing usage and community feedback, then scaling and industrializing them if relevant is currently our preferred method for organizing the transition. We learn and decide by doing and according to opportunities we discover step by step, while trying to take consistent options in the long run. When the benefits of change will become obvious to the majority, we will be able to change the production methods and infrastructure.

The main visible manifestation of this approach is the data.bnf.fr project.² This application was designed to be usable by individual, human-driven browsers, navigating through the various pages of a website. It generates web pages providing standardized information, references, and links about authors, works, or subjects. The service is also intended to be used by machines and search engines in particular. [Data.bnf.fr](http://data.bnf.fr) groups and exposes online data in RDF form coming from heterogeneous sources which can be easily indexed by search engines and densely linked to other resources, either internal to the BnF (its MARC and EAD main catalogs, the digital library Gallica, etc.) or external (the

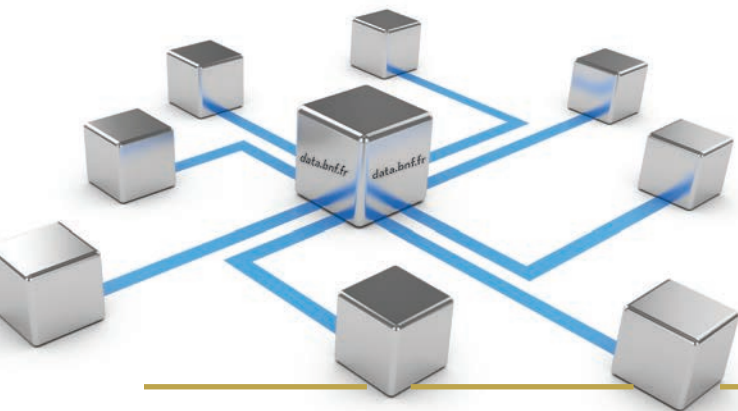
Union catalog for French Academic libraries SUDOC, the French Union catalog CCFR, WorldCat, VIAF, Wikipedia, etc.). The whole process requires the transformation of MARC or EAD formatted metadata into the information hub, based on modeling techniques in RDF and on standard vocabularies (DC, SKOS, RDA, and FOAF). The modeling activity has a direct link with aligning and enriching the data that have to be extracted and processed. Contents, links, and services are brought together in compliance with information concepts based on the FRBR bibliographical entities or groups of entities: those are integrated within a publication architecture designed both to build the HTML pages and to display raw data dumps in RDF and JSON. The data gathered from various datasets is brought together at the right level, so that works and expressions can be found in a way that complies with the new bibliographic description requirements. [Data.bnf.fr](http://data.bnf.fr) does not mean to replace the existing catalogs and other silos it exploits, but to provide some “glue” between them. In short, it aims at making our library data work better on the web, by delivering a service of information, with structured explicit data and permanent URIs—a bibliographic information hub constitutive of a trusted environment made of reliable data. In order to facilitate data dissemination and reuse, all raw datasets are made available for free download under an ODC-BY and CC-BY compliant public open license recommended by the French Government Open Data mission Etalab.

Launched in 2011, this project demonstrates encouraging results. With over 5.6 million links to bibliographic records from the BnF main catalog, covering 200,000 authors, 92,500 works, and 171,000 subjects or themes, it is now estimated to cover 40% of the references from the BnF source catalogs. We target to reach 80 to 90% of the total by the end of 2015. At the end of 2012, for its first full year in operation, data.bnf.fr cumulated 637,650 unique visitors and 1.2 million page hits. On a monthly basis, we currently observe an average of 50,000 unique visitors per month. 80.6% of the visits come from a web search engine. This is an encouraging figure, which shows that most people using data.bnf.fr find it via a search engine, demonstrating success as to web exposure. The conversion rate is 70%, which means that 70% of the visits to data.bnf.fr lead to a visit of another BnF application (catalogs, Gallica, etc.). This is a good figure as well, as it

CONTINUED »

¹ To follow the discussions and ongoing work on RDA implementation in France (strategy, standardization, education, dissemination...), see this dedicated website (in French) “RDA en France”: <http://rda-en-france.enssib.fr/>

² [Data.bnf.fr](http://data.bnf.fr) won the 2013 Stanford Prize for innovation in research and national libraries. The text supporting BnF’s application for this prize provides a comprehensive presentation in English of the project goals and outcomes: <http://library.stanford.edu/projects/stanford-prize-innovation-research-libraries-spir/2013-spir-winners>



The conversion rate is 70%, which means that 70% of the visits to data.bnf.fr lead to a visit of another BnF application (catalogs, Gallica, etc.). This is a good figure as well, as it shows that data.bnf.fr is fully playing its role as an information hub (rather than a substitute), driving new traffic towards other BnF resources and applications.

shows that data.bnf.fr is fully playing its role as an information hub (rather than a substitute), driving new traffic towards other BnF resources and applications. The BnF cataloging staff has shown great interest in the development of this project. It is indeed a very concrete use case for professionals to see the data they produce in MARC presented in FRBR mode. The project development leads to the discussion of priorities and processes in the bibliographic transition: which datasets should be exposed next in data.bnf.fr and along which quality or content criteria? To which external data should the BnF link its own data to? Should current data transformation processes and algorithms influence the existing metadata models and production practices? Conducting such conversations and encouraging collective decision-making on the basis of this project has considerably improved the general perception of metadata issues at the library.

Although this project is used as a powerful vehicle for internal and external communications, it is only the visible part of the BnF bibliographic iceberg. Behind and beside data.bnf.fr and the continuation of our long-term effort in standardization work on ISBD, RDA, and FRBR, we have identified some “building blocks” that we believe will be key requirements to sustain the library’s future presence and architecture in the Semantic Web. One of these building blocks has been the implementation of a comprehensive approach for the management of persistent identifiers. This started several years ago by assigning ARK identifiers to all objects and records from the library. Our current priority in this field is the implementation of ISNI for public identities. We are convinced that the management of authorities should be a strong focus to prepare the future, which explains BnF’s strong institutional involvement both in the VIAF council

and in the ISNI Agency. Last summer, we managed to ingest 1.3 million ISNI identifiers in the BnF catalog. We are now planning their dissemination via our bibliographic services and are hoping this will ultimately answer some of the expectations of French publishers, academic, or rights management organizations, which are all in need of a global identifier to manage information databases about creators. In this context, we are getting closer than in the past to French publisher organizations, seeking more interoperability solutions and envisioning new workflows between their publishing industries and the library within the legal deposit framework, notably its extension to e-books. This involves working on ONIX/INTERMARC conversions and exploring various scenarios where the BnF could derive more metadata from the publishers just like we now derive many more records from WorldCat for our foreign acquisitions.

As to metadata exchange, our observation so far is that the data model designed for data.bnf.fr seems to be an acceptable compromise between generic web usage and exposure and basic bibliographic exchange needs: it is poorer than MARC but richer than schema.org, which we use in data.bnf.fr but consider more like a sitemap for webmasters and search engines than a data model. However, the way we serve our metadata in data.bnf.fr is not rich enough for high quality bibliographic exchange. This is why we are now looking into the possibility of expressing the full granularity of our INTERMARC format in RDF, the goal being to offer triple stores (via SPARQL endpoints) where people could just pick and choose what they need.

Q Can you summarize the focus of European libraries in the past five years? What has been the main focus of effort under the topic of metadata management?

The current discussions developing in North America and within the broader Anglo-American cataloging community regarding bibliographic data exchange models raise a mix of excitement and confusion in Europe. In the view of many European national libraries and bibliographic agencies, the invention and consolidation, within the framework of IFLA, of the FRBR model (and its later extensions to authority records and subjects with FRAD and FRSAD) is seen as the conceptual starting point of what we now call the bibliographic revolution. It is a strong view in Europe that the vision underlying this model remains valid and should be the main driver for bibliographic change, as FRBR is being consolidated by the IFLA international principles of cataloging while allowing for innovation and adjustments to the digital Age. European libraries invested a lot in FRBR theory and data modeling and still do, as shows, for instance, current developments with FRBRoo and PRESSoo and other models deriving from FRBR.

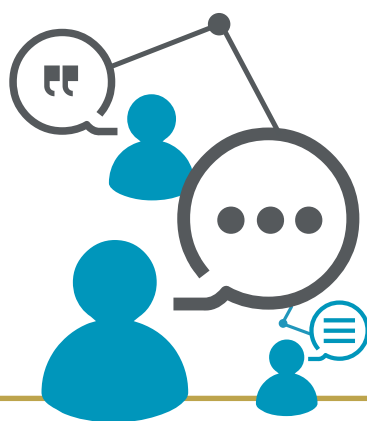
The development of the RDA cataloging rules and the beginning of their actual implementation in several major libraries is seen as a very significant and positive step to implement the FRBR model and make it happen in real life. Several European libraries, mainly from the AACR2 and MARC 21 tradition, have started translating and implementing it or are planning to do so in the coming years. Others, coming from different bibliographic traditions—mostly ISBD and UNIMARC—still see some limitations in RDA and aren't eager to adopt it as it is, mainly because it doesn't fit some of their practices and still requires some improvements in terms of internationalization or full compliancy with the FRBR model. From that perspective, considering the cost of change, there is a notion that if they should invest in such radical change, it should be for ambitious implementation scenarios which best fit the promise of FRBR. These institutions have put much effort in understanding the rationale of RDA and proposing adjustments where they needed them. The European RDA Interest Group (EURIG) was formed two years ago to provide a forum for European bibliographic organizations to collectively discuss and propose adaptations to the RDA code in order to address these issues. To date, this process and the subsequent interactions with the RDA Joint Steering Committee have been judged a constructive one, where all parties are given a voice. Although the whole process can be too slow, we know international standardization in the bibliographic field is one of the most complex types and that consensus cannot be achieved in one day in such matters. All in all, the dynamics of moving from the FRBR model to the RDA rules and their actual implementation following principles of international cooperation are regarded as a very encouraging process in Europe. Most European libraries seem ready to make compromises in order to reach some agreement so that institutional roadmaps may converge in the same directions for the benefit of international interoperability and future metadata exchanges. This is the exciting part.

The more confusing part has to do with recent developments regarding data exchange models in North America. Several European libraries perceive a contradiction between the collaborative effort which helped in designing FRBR and RDA over time and the way the question of data infrastructure is presently being addressed. While both FRBR and RDA are supposed to be agnostic as to



In the view of many European national libraries and bibliographic agencies, the invention and consolidation, within the framework of IFLA, of the FRBR model (and its later extensions to authority records and subjects with FRAD and FRSAD) is seen as the conceptual starting point of what we now call the bibliographic revolution.

CONTINUED »



I personally believe that we need to restore the conversation within the international bibliographic community and to encourage better communications between metadata standard specialists and linked data architects. This could help clarify things and avoid some misunderstandings.

technical implementation, there is an overall feeling—which might be more of a misunderstanding about what the BIBFRAME initiative is actually trying to achieve—that important decisions and standards may be defined overseas without sufficient discussion with European libraries nor in compliancy with the initial vision and objectives that led to the definition of the FRBR model. At this very stage, I would say that this situation is a source of confusion for many, especially in the context where libraries feel the urge of demonstrating tangible results in metadata transformation and in developing new services fitting the Linked Open Data legal and technical requirements. Some libraries have started making their data open, but the data isn't linked. Others have started linking their data, but it's not open. Nobody really knows if the data exposed in RDF is being reused or has found proper metrics to evaluate this. FRBRization experimentations are being conducted in catalogs, at various levels of ambition, and through various channels (whether encouraged by ILS vendors or run internally via specific projects).

Q What should be the focus of the new metadata initiatives in the next two years? Are there any gaps in the current efforts that could be filled in the near term?

All in all, there is currently a bit of confusion on how various institutional, national, regional, and global initiatives may converge as it seems to me that there is no proper framework to share best practices and confront technical implementation with standards requirements. This is all the more critical since within institutions it is often not the same teams who are involved in bibliographic standardization and in linked data projects, which makes it rather challenging to identify institutional policies or strategic roadmaps. This is an issue each institution should try to address internally.

In the meantime, it seems to me that while we had a rather clear focus and collective framework on the basis of FRBR and RDA in the past years, the urge of action has lead North American as well as European organizations either to act individually or to adopt a “wait and see” attitude which sometimes paralyzes them, especially when they are short of resources. I personally believe that we need to restore the conversation within the international bibliographic community and to encourage better communications between metadata standard specialists and linked data architects. This could help clarify things and avoid some misunderstandings. Typically, many people (especially at management level) tend to mix up models (e.g., FRBR), cataloging rules (e.g., RDA), formats and languages (e.g., MARC or RDF), and technical implementation solutions while these concepts operate at different levels, in different timeframes, and have different impacts.

It is obvious that different strategies will develop around the world as to the bibliographic transition, depending on institutional priorities, legacies, dependencies, and resources—especially in the context of the Semantic Web, which precisely allows for a diversity of approaches. But it could be helpful to define core areas of cooperation and implementation. Among those “building blocks” for the future that may require more international cooperation and reciprocal benchmarking (using existing forums such as IFLA or DCMI, or creating new,

dedicated platforms), I believe we should list: data exchange models, licensing and legal issues, publication and alignments of vocabularies, and global identifiers. At this stage of the bibliographic transition dynamics, we would benefit from a shared vision on these issues, which would help institutions planning their actions with a better notion of the areas where strong interoperability aspects are to be considered (and consensus searched, by means of collaborative discussions on standards) and other areas where they should feel more comfortable doing what they want to do depending on their specific needs and mandates. | IP | doi: 10.3789/isqv25n04.2013.05

Bibliothèque nationale de France - ISNI 0000 0000 7409 4530

It is obvious that different strategies will develop around the world as to the bibliographic transition, depending on institutional priorities, legacies, dependencies, and resources—especially in the context of the Semantic Web, which precisely allows for a diversity of approaches.

Bibliothèque nationale de France

www.bnf.fr/en/tools/a.welcome_to_the_bnf.html

BnF Data Project

data.bnf.fr

Dublin Core Metadata Initiative (DCMI)

www.dublincore.org

Dublin Core Metadata Élément Set

www.niso.org/standards/z39-84-2005/

Etalab

www.etalab.gouv.fr/pages/Licence_ouverte_Open_licence-5899923.html

European RDA Interest Group (EURIG)

www.slainte.org.uk/eurig/

FOAF Vocabulary Specification

xmlns.com/foaf/spec/

FRBR Object-Oriented Definition and Mapping (FRBRoo)

www.cidoc-crm.org/docs/frbr_oo//frbr_docs/FRBRoo_V2.0_draft_2013May.pdf

Functional Requirements for Authority Data (FRAD)

www.ifla.org/publications/functional-requirements-for-authority-data

Functional Requirements for Bibliographic Records (FRBR)

www.ifla.org/publications/functional-requirements-for-bibliographic-records

Functional Requirements for Subject Authority Data (FRSAD)

www.ifla.org/node/5849

INTERMARC

www.bnf.fr/fr/professionnels/format_intermarc/s.intermarc_presentation.html

International Federation of Library Associations and Institutions (IFLA)

www.ifla.org/

International Standard Bibliographic Description (ISBD)

www.ifla.org/publications/international-standard-bibliographic-description

International Standard Name Identifier (ISNI) International Authority

www.isni.org

JavaScript Object Notation (JSON)

www.json.org

MARC 21

www.loc.gov/marc/

ONIX

www.editeur.org/8/ONIX/

PRESSoo (a periodicals object-oriented ontology)

www.issn.org/files/issn/technicals/PRESSoo_01.pdf

Resource Description and Access (RDA)

www.rda-jsc.org/rda.html

Resource Description Framework (RDF)

www.w3.org/standards/techs/rdf#w3c_all

Simple Knowledge Organization System (SKOS)

www.w3.org/2004/02/skos/

SPARQL

www.w3.org/standards/techs/sparql#w3c_all

UNIMARC

www.ifla.org/publications/unimarc-formats-and-related-documentation

Virtual International Authority File (VIAF)

viaf.org

WorldCat

www.worldcat.org/



**RELEVANT
LINKS**

Richard
Wallis

Schema Bib Extend

RICHARD WALLIS

In the evolving world of the web, bibliographic resources have gained a reputation for being difficult to discover. Search engines are on a mission to identify *things* on the web, as against just indexing texts about those things. Their initiatives could help solve some of the lack of visibility and discoverability issues in the bibliographic domain—a domain where describing things in text, as opposed to data, is the centuries old, *modus operandi*.

To take best advantage of such progress, you need to be part of, or at least be well represented in, the evolution of the standards and practices that are building the *things* based view of the world. This is where Schema Bib Extend fits, an influencer recognizing the concerns, experience, knowledge, and ambitions of the bibliographic corner of the web. A corner with much to offer that could be undervalued if we do not speak up and get involved.

What is Schema Bib Extend?

Schema Bib Extend is a W3C Community Group focused on establishing a consensus within the bibliographic community around proposals to submit to the WebSchemas Group

for extending the Schema.org vocabulary to enhance its capabilities in describing bibliographic resources.

That statement needs unpacking:

A W3C Community Group is an open forum, without fees, where web developers and other stakeholders develop specifications, hold discussions, develop test suites, and connect with W3C's international community of web experts. The Schema Bib Extend group was formed as a Community Group to take advantage of the open forum for stakeholders the W3C provides.

The Schema.org vocabulary was developed through cooperation between Google, Bing, Yahoo! and Yandex, and released in June 2011. The purpose is to provide a vocabulary for marking up structured data on the web that will be recognized by the major search engines. The process for commenting upon and

proposing extensions and enhancements to the Schema.org vocabulary is also handled through a W3C Community Group—*WebSchemas*—with its associated *public-vocabs* mailing list.

In October 2012, I established and became chair of the Schema Bib Extend Community Group (SchemaBibEx). It has attained a membership in excess of 80 people, acting as individuals and/or representing organizations with interests in the bibliographic domain. Organizations represented include several national libraries, library system vendors, publishers, W3C, universities, cooperatives, and consortia. The group meets regularly by conference call and, via the community wiki, has already formed and submitted several proposals on topics such as Collections, Citations, and AudioBooks to the *WebSchemas* Group.

Formation of SchemaBibEx

I formed the group following many conversations that were stimulated by the release of open linked data in OCLC's WorldCat, using Schema.org as the vocabulary for data description.

By adding Schema.org-described metadata to the WorldCat pages, using the RDFa formatting technique, OCLC made available linked data descriptions of the over 300 million resources referenced in WorldCat. Schema.org was chosen as the vocabulary because of its general acceptance across the web and the fact that major search engines would recognize it. In the process of preparing these descriptions, it became clear that Schema.org did not cover certain concepts and format types. The OCLC developers created a prototype library vocabulary to supplement Schema.

In discussions, it was clear to me that there was a potential consensus that Schema.org could form the basis for describing bibliographic resources on the web, but it would need some enhancement to realize that possibility.

Following the lead of those behind Schema.org, the open group was formed, with the help of the W3C, believing that a proposal from a group of interested parties could carry more weight than those from individuals alone. Also such a group could bring informed discussion and use cases to bear on the proposals in their formation.

A change in thinking

In the early months of the group's discussions, it became clear that proposing extensions to an established general-purpose vocabulary is very different than creating and maintaining a vocabulary/standard focused on a single domain such as libraries.

Our experience and practice over many years has conditioned us to be a bit too deep and too bibliographic specific. The initial effect of this was to suggest that there was to be a significant amount of effort to identify many bibliographic vocabulary terms not present in Schema.org.

A change in approach evolved. Issues were addressed and explored by taking the Schema.org vocabulary as is and using it to describe resources, and their relationships, in the bibliographic domain. In this process, example webpages for bibliographic resources were examined to see what Schema.org markup would be appropriate. The outcome of this approach was to realize how good Schema.org was already for describing our resources, and to identify specific gaps in coverage—it had no Audiobook class for instance.

In a few cases, where the initial presumption was that new classes/properties would be required, it became clear that advice, documentation, or examples would be sufficient. In other cases, proposed tweaks to the descriptions in Schema.org documentation would be all that is needed.



In this process, example webpages for bibliographic resources were examined to see what Schema.org markup would be appropriate. The outcome of this approach was to realize how good Schema.org was already for describing our resources, and to identify specific gaps in coverage—it had no Audiobook class for instance.

CONTINUED »

An approach for Holdings

A good example of all the above is the work the group is currently engaged with to describe library holdings. This would enable libraries to describe, using Schema.org, the availability of items to loan or access in other ways.

Initial thoughts could have resulted in proposals for library-specific classes and properties. However the use of the Schema.org *Offer* class—with some adjustments to its documentation descriptions to take into account that offers can be to loan and share, as well as to sell—will go a long way to satisfying the library, available to access, use case. What then remained was some finer detailed work on which, and if any, new properties could be used to describe library-specific things such as shelf marks, call numbers, etc.

A Group with a short future

When setting up the Group, I expressed the ambition for it to have a lifetime measured in months not years. The reasoning behind this being that it was being set up to guide and inform the wider web community, served by Schema.org, on how to improve its representation of bibliographic resources—not to become a group emulating and duplicating metadata standards.

Although there is much to do, it could be possible for the majority of the issues to be addressed before the completion of the Group's second year.

What will the SchemaBibEx legacy be?

As a group representing many voices in the bibliographic domain, it has already become one looked to and referenced in broader discussions on the main, *public-vocabs*, Schema.org mailing list. Several group members are active on that list as individuals, participating in discussions some of which overlap with those in the SchemaBibEx Group.

Obviously if the Group achieves its goal, Schema.org will be better suited for the general representation of bibliographic resources, and hence such resources should be better represented in the web of data and easier to discover.

The documentation and examples that the group produces as part of its discussions could provide guidance to those wishing to describe bibliographic resources on ways to approach the issue. This should help deliver some consistency of output across the domain.

It is also apparent that through the activities of the group, system developers have been encouraged to look to using this approach to describing library resources on the web. For example, in addition to OCLC's WorldCat developments, open source library systems such as Evergreen and Koha have implemented code to expose Schema.org in their user interfaces.



If the Group achieves its goal, Schema.org will be better suited for the general representation of bibliographic resources, and hence such resources should be better represented in the web of data and easier to discover.

In summary, the SchemBibEx Group and its proposals as adopted should result in bibliographic resources being more consistently and more often represented in the web of data, and hence more discoverable. | IP | doi: 10.3789/isqv25no4.2013.06

RICHARD WALLIS (richard.wallis@oclc.org) is Technology Evangelist with OCLC in the UK.

Evergreen ILS

evergreen-ils.org

Koha ILS

www.koha.org

public-vocabs W3C mailing list

lists.w3.org/Archives/Public/public-vocabs/

RDFa

www.w3.org/standards/techs/rdfa#w3c_all

Schema Bib Extend

www.w3.org/community/schemabibex/

Schema.org

schema.org

WebSchemas wiki

www.w3.org/wiki/WebSchemas

WorldCat

www.worldcat.org



RELEVANT
LINKS



Todd
Carpenter

TODD CARPENTER

Charting a Course through a New Exchange Environment: The NISO Bibliographic Roadmap Initiative



The current bibliographic exchange environment is nearly 50 years old and it is showing its age. Since its pioneering development in the 1960s, the MARC record system has proven exceptionally adept at providing a convenient lingua franca of exchanging bibliographic content between libraries and library systems. Momentum has slowly been gathering over the past five years toward replacing that system, and development work is now underway within the Library of Congress (LC) to develop a new Bibliographic Framework (or BIBFRAME as it is often described). A great deal of energy and discussion has circulated around the initiative and effort by LC.

However, the BIBFRAME initiative is focused only on one aspect of the ecosystem surrounding bibliographic information exchange, namely the data model describing resources. While this is a critical component of managing and exchanging bibliographic information, it is by no means the only node in the web of exchange. In 2011, The Andrew W. Mellon Foundation generously funded the National Information Standards Organization (NISO) to undertake a community effort to describe those other elements and put forward a roadmap of additional work that might need to be undertaken before a transition to a new environment can be executed.

The NISO *Bibliographic Roadmap* initiative has the goal of identifying what are critical gaps in the foundation that need to be eliminated in order for the bibliographic exchange environment

to transition to a new method of data exchange. Beyond the data model, there need to be interchange methodologies, trust and provenance frameworks, financial incentives to invest in new systems development or purchase, staff planning and training, as well as legal and operational policies to govern this new environment. While it is too early in some cases to advance work on these larger issues, it is certainly not too soon to advance the discussions about what may be necessary and to begin laying the groundwork for the future development of those research initiatives, best practice developments, business policy frameworks, or exchange standards.

NISO began this initiative by hosting a public discussion about the goals of the initiative. Along with some online discussion sessions, the outline of a two-day discussion symposium was

developed. That meeting, held in April 2013 in Baltimore, MD, attracted 36 people from around the world and was also live streamed via the Internet to a larger and broader community than just those in the room. As an “unconference,” the session’s content was driven by the interests of the attendees. Several attendees gave short lightning talks about their respective institution’s work in fostering or developing linked data tools, exposing their data, utilizing linked data in their services. This primed the thinking of the participants in their brainstorming of what issues need to be addressed to advance a new ecosystem of bibliographic data exchange.

Ideas from that meeting have been posted into an online community feedback system that allows comment and voting; this allows an even broader community to provide input and feedback as well as

CONTINUED »

identify priorities of these potential work topics. Interested parties can provide feedback through the end of January 2014. A series of small discussion groups with invited experts will also be used to help frame and refine the potential work areas for identified workstreams.

The areas identified during the in-person meeting ranged from focusing attention on adoption barriers and business models to prototyping and developing interaction standards. Not surprisingly, the list of potential ideas and topics was diverse, wide-ranging, and in some cases, not fully defined. Some of this uncertainty is a result of the somewhat mutable state of what we mean when we say that library data exchange needs to be transformed. Some take that to mean linked data and the Semantic Web. Some conjure a system upon which MARC records are extended in some way using FRBR principles. And others envision a completely different environment with data systems sharing and linking from a variety of sources to craft a discovery and management environment. When even this basic understanding of the future we are moving toward is still so unsettled, precisely defining what next steps are necessary for this “foggy” future is complicated. That said, regardless of the structure of such a new bibliographic exchange ecosystem, several themes did emerge from the discussions that are worthy of additional exploration.

These themes could be grouped broadly into a few categories:

- » value proposition and business models;
- » data exchange and interoperability;
- » data quality and authority; and
- » education and communication.

Data exchange is reliant, of course, on a clear understanding of what is being transmitted and how. One idea was to foster prototype development, once progress on a data model has been achieved. Exploring the methods and rules for how data will be shared and exchanged efficiently and reliably was another suggested area of focus. Such efficiency and reliability requires some measures of authority and trust to be built into the system. Other identified areas of focus centered on the value that these new approaches would bring to both the users of the data and the libraries and content creators providing the data. This value proposition leads directly to the business models and economics surrounding development, deployment, and adoption of new library systems, which need further study and consideration. Along the way, barriers to development, exchange, and adoption would benefit from analysis and forethought on how to overcome them. Finally, communicating these identified benefits of the new solutions or tools was identified as a critical success factor to the overarching effort.

There were a great deal more specific ideas, both large and small, which surfaced during the meeting. All of the notes from the various discussion groups, as well as recorded videos of most of the meeting are available on the NISO Bibliographic Roadmap project webpage. The initiative will continue to

gather feedback from the community until the final report is prepared for the Mellon Foundation in early 2014. A follow-up webinar occurred in early December and a discussion meeting will be held at ALA Midwinter in Philadelphia in January 2014. If anyone has any input to share, comments are welcome and should be directed to the NISO office or added to the online feedback forum. We expect that the final report will be distributed to the community around April 2014.

The true shape of this new environment is nearly impossible to predict precisely, because, as previously noted, the community is not yet settled on a single approach to bibliographic data sharing. The various approaches will have their pros and cons and their advocates and opponents. At least in the short term, it is likely that there will be multiple approaches, with a period of both experimentation and overlap until a consensus approach is defined. Even well after that, some community members will continue using legacy systems for an extended period.

It is NISO’s hope that by fostering these conversations now about the larger environment of data sharing, adoption of these new systems, once they are ready for the market, will occur more rapidly. It also makes sense to begin laying the groundwork for what social or technological structures we will need in place once the new bibliographic ecosystem begins to develop. There is a very diverse community that will be impacted by a new bibliographic data ecosystem, one that goes far beyond the traditional library community of users. By identifying and considering the needs of a variety of affected stakeholders and defining the roadmap to meet these needs, our new bibliographic framework and supporting systems and infrastructure will have a much greater change of common acceptance and adoption. And ultimately, bibliographic data sharing will be fully integrated into the greater web of knowledge and information. | *NR* | doi: 10.3789/isqv25no4.2013.07

TODD CARPENTER (tcarpenter@niso.org) is Executive Director of the National Information Standards Organization.

NISO Bibliographic Roadmap project webpage
www.niso.org/topics/tl/BibliographicRoadmap/

NISO Online Bib Roadmap Feedback Forum
bit.ly/1hxr16P

Library of Congress Bibliographic Framework Initiative
www.loc.gov/bibframe/

FRBR: Functional Requirements for Bibliographic Records
www.ifla.org/publications/functional-requirements-for-bibliographic-records



**RELEVANT
LINKS**



The *PIRUS Code of Practice* provides the specifications and tools that will allow COUNTER-compliant publishers, repositories, and other organizations to record and report usage statistics at the individual article level that are credible, compatible, and consistent.

COUNTER Releases First PIRUS Code of Practice for Recording and Reporting Usage at the Individual Article Level

The *PIRUS Code of Practice, Release 1*, provides specifications for the recording and reporting of usage at the individual article level that are based on and are consistent with the *COUNTER Code of Practice for e-Resources*. The *PIRUS Code of Practice* provides the specifications and tools that will allow COUNTER-compliant publishers, repositories, and other organizations to record and report usage statistics at the individual article level that are credible, compatible, and consistent. (Non-COUNTER-compliant organizations may use the Secondary Clearing House services described in Section 1.10.)

This Code of Practice contains the following features:

- » A list of definitions and other terms that are relevant to recording and reporting usage of individual items
- » A methodology for the recording and reporting of usage at the individual article level, including specifications for the metadata to be recorded, the content types, and the versions whose usage may be counted
- » Specifications for the PIRUS Article Reports
- » Data processing rules to ensure that the usage data reported are credible, consistent, and compatible

- » Specifications for the independent auditing of the PIRUS reports
- » A description of the role of a Central Clearing House (CCH) in the calculation and consolidation of PIRUS usage data for articles, as well as other Clearing Houses in relation to the CCH

Unlike the standard COUNTER usage reports, which vendors must update monthly for all products covered, the PIRUS usage reports do not have to be provided monthly for every article they cover (but should be broken down by month when reported). Rather, the vendor must have the capability to produce the PIRUS reports for all the journal articles they host on an annual basis, as a minimum requirement.

COUNTER will be responsible for the ongoing management and implementation of PIRUS. ■

ⓧ About PIRUS: www.projectcounter.org/pirus.html

PIRUS Code of Practice: www.projectcounter.org/documents/Pirus_cop_OCT2013.pdf

AAP Publishes EPUB 3 Implementation White Paper

In July 2013, the Digital Issues Working Group of the Association of American Publishers (AAP) launched an EPUB 3 Implementation Project “to bring together a group of people who could provide perspectives from a variety of publishers, reading system developers, retailers, service providers, and the accessibility community to jointly articulate priorities for the implementation of EPUB 3 features by reading systems and best practices for the creation of EPUBs, with a special emphasis on enabling accessibility.”

Many publishers are planning to issue and distribute EPUB 3 files in 2014 but have concerns over the inability or inconsistency of reading systems to support various features in the EPUB 3 format. While the long-term goal is to have the EPUB 3 standard fully supported, the AAP project was developed to identify priorities to enable greater use of the new format in the near-term. Among the supporters of the AAP initiative were the International Digital Publishing Forum (IDPF), the Radium Foundation, the Book Industry Study Group, EDItEUR, Benetech, the American Printing House for the Blind, the National Federation of the Blind, and the DAISY Consortium.

A white paper published in October 2013 summarizes the results of the initiative’s workshop held in New York on September 10. Priorities of features were assessed in four major workstreams: general features, accessibility, metadata, and use cases.

The most critical issues identified were:

- » HTML5 structural elements
- » Implementation of manifest and HTML5 fallbacks

- » Consistent navigation
- » Improvement and use of validation and preflight testing mechanisms
- » Inclusion of image descriptions
- » Proper use of and rendering of fundamental HTML5/CSS3 features such as HTML tables, lists, MathML, SVG, fonts, asides, floats, and image sizing and positioning
- » Support for media using standard HTML5/CSS features for audio and video with proper fallbacks and EPUB 3’s Media Overlays

“It was clear to all participants in this initiative that improvements in both reading system feature implementation and practices for creating EPUBs on the part of publishers are not just important, they are urgent.” The white paper concluded that “while the ecosystem will never be perfect—both the EPUB 3 standard and the reading systems that implement it will continue to evolve—the prospect of a well-functioning EPUB 3 ecosystem is actually quite close: an ecosystem in which a great many fundamental and important features can be used consistently by publishers with the expectation that they will be implemented in a wide range of reading systems and platforms.” ■

Ⓢ AAP Digital Issues Working Group:
www.publishers.org/committees/22/

AAP EPUB 3 Implementation Project
White Paper: publishers.org/press/117/



ANSI Launches Online Portal for Standards Incorporated by Reference

The American National Standards Institute (ANSI) has launched an online tool, the Incorporated by Reference (IBR) Portal, to provide read-only access to standards that are incorporated by reference into federal laws and regulations.

The U.S. Federal government has in the last few years issued requirements for such standards to be “reasonably available” to those affected by the legislation. This has created problems with people violating copyrights on these standards and others have suggested that such copyrights be declared invalid. ANSI has been leading the effort to educate legislators and the public about why these standards shouldn’t automatically be free and the importance of the revenue from the standards in supporting the work or even existence of the relevant standards development organizations (SDOs).

The IBR Portal is a solution that makes standards incorporated by reference available in a read-only mode (no printing, downloading, transferral, or even screenshots), thus providing access to the information but still safeguarding the SDO’s intellectual property. Thirteen SDOs are offering their IBR standards through the ANSI portal and seven others are allowing links to the IBR standards on their own website. Additional SDOs are expected to participate following Phase 1 of the portal’s implementation. ■

Ⓢ ANSI Press Release: www.ansi.org/news_publications/news_story.aspx?menuid=7&articleid=3771&source=whatsnew102813

Incorporated by Reference (IBR) Portal: ibr.ansi.org



SD [STANDARDS IN DEVELOPMENT: *November 15, 2013*]

Listed below are the NISO working groups that are currently developing new or revised standards, recommended practices, or reports. Refer to the NISO website (www.niso.org/workrooms/) and the *Newsline* quarterly supplements, *Working Group Connection* (www.niso.org/publications/newsline/), for updates on the working group activities.

Note: DSFTU stands for Draft Standard for Trial Use.

WORKING GROUP	STATUS
Demand Driven Acquisition of Monographs Co-chairs: Michael Levine-Clark, Barbara Kawecki	Recommended Practice (NISO RP-20-201x,) in development.
Digital Bookmarking and Annotation Sharing Co-chairs: Ken Haase, Dan Whaley	Standard (NISO Z39.97-201x) in development.
Journal Article Versions (JAV) Addendum Chair: Michael Dellert	Revised Recommended Practice (NISO RP-9-201x) in development.
Knowledge Base and Related Tools (KBART) Phase II <i>Joint project with UKSG.</i> Co-chairs: Magaly Bascones, Chad Hutchens	Phase II Revised Recommended Practice (NISO RP-9-201x) being finalized for publication following the public comment period.
Open Access Metadata and Indicators Co-chairs: Cameron Neylon, Ed Pentz, Greg Tananbaum	Recommended Practice (NISO RP-22-201x) in development.
Open Discovery Initiative Co-chairs: Marshall Breeding, Jenny Walker	Recommended Practice (NISO RP-19-201x) being finalized for publication following the public comment period.
Protocol for Exchanging Serial Content (PESC) Co-chairs: Leslie Johnston, Kimberly Tryka	Recommended Practice (NISO RP-23-201x) in development.
Resource Synchronization Co-chairs: Herbert Van de Sompel, Todd Carpenter	Standard (NISO Z39.99-201x) being finalized for publication following the comments from the beta draft.
Standard Interchange Protocol (SIP) Co-chairs: John Bodfish, Ted Koppel	Standard (NISO Z39.100-201x) in development.
SUSHI Server Working Group Chair: Oliver Pesch	Recommended Practice NISO RP-13-201x, <i>Providing a Test Mode for SUSHI Servers</i> being finalized for publication following a draft for trial use.

ISDI doi: 10.3789/isqv25no4.2013.09



NISO

NISO helps you and your organization connect with the information world.

NISO crosses over the library, publishing, and information systems communities, providing an excellent forum for these groups to agree on solutions.



JOIN NISO
and play an active role in standards solutions!

www.niso.org