

ARTICLE
EXCERPTED
FROM:

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

ANNOTATION

AUTHOR

NISO
How the information world
CONNECTS

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 »

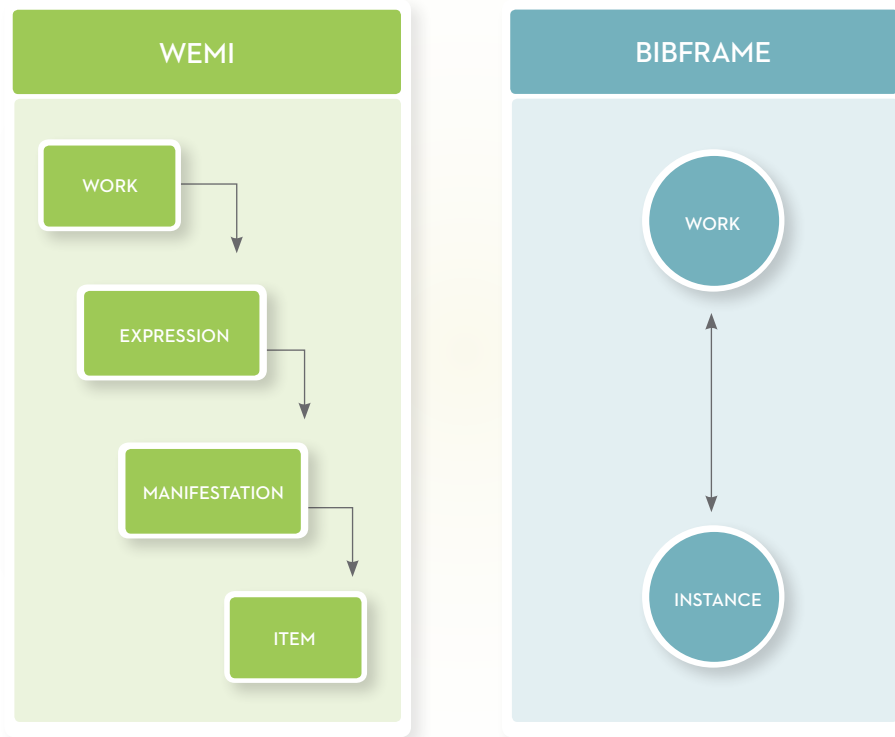


Figure 1: RDA Entities vs. BIBFRAME Classes

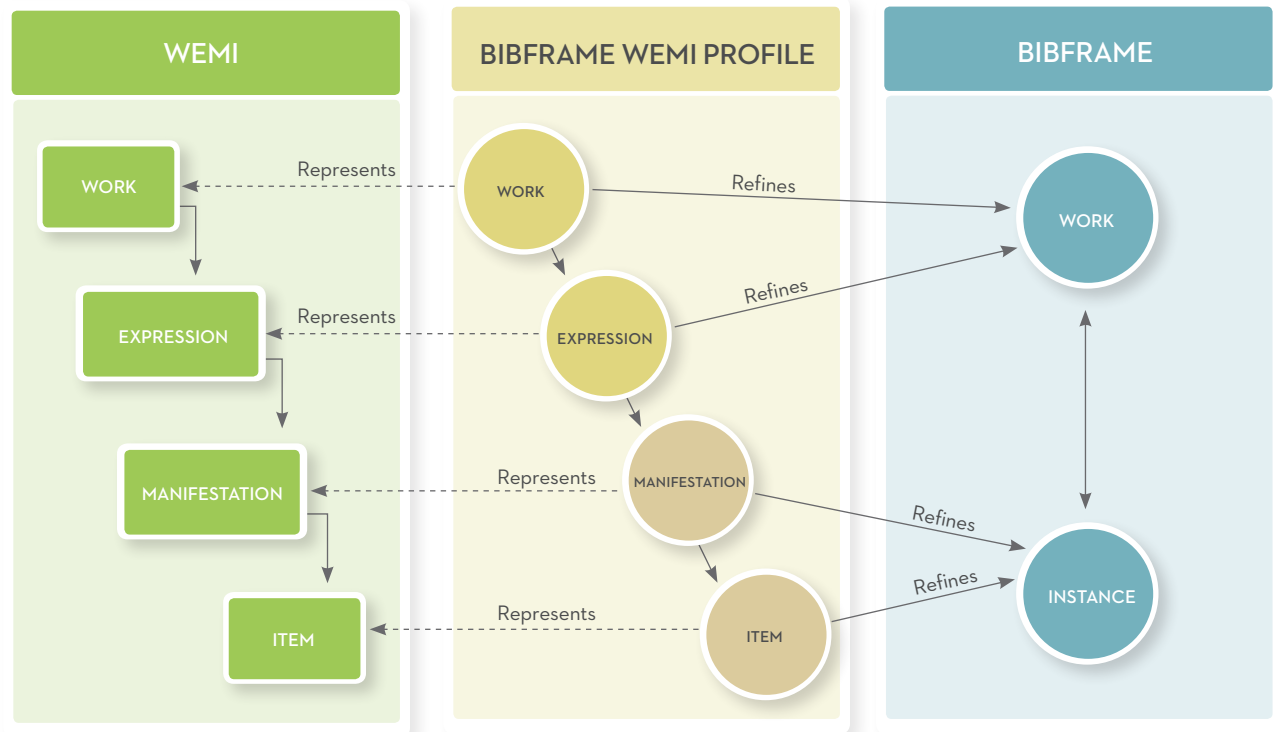


Figure 2: Refinement of WEMI Entities to BIBFRAME Classes through Profile

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

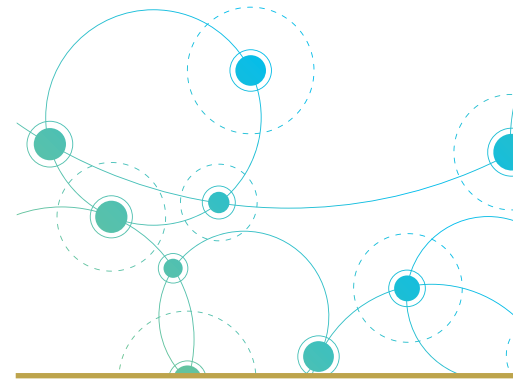
When the BIBFRAME initiative surfaced in 2012, its design characteristics struck a chord with the George Washington University Libraries (GW) administration: customization, openness, productivity, shareability, and resource development. They also recognized that GW staff could make an important contribution 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 3792771	Identifier
005 20020509142301.0	Identifier
008 512418191 eng 000 1 eng	Language
010 1# 51052940	Language
019 1#29774405	Language
035 1#0001C0em15525100	Language
040 1#DLC 1#DOW	Language
100 1 #Austin, Jane, 1817?-1817.	Author
245 10 \$Title and prologue / \$Cline Austin ; with an introduc-	Title
650 0 \$Young women \$Fiction.	Subject
650 0 \$Courtship \$Fiction.	Subject
650 0 \$Sisters \$Fiction.	Subject
651 0 \$England \$Fiction.	Subject
652 0 \$Domestic fiction. \$2aaf	Subject
653 7 \$Love stories. \$2aaf	Subject
030 1#027945429	Language
260 1#New York : \$Bantam / \$Distributed by Random House	Publication
300 1#xiviii, 358 p. : \$22 cm.	Publication
490 0 1#Everyman's library ; \$vi	Publication
049 1#DOW	Language
050 00 1#PR4034 1#P7 1991	Language
082 00 1#K23- / \$22	Language

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 many 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/isqv25no4.2013.04

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/