Cataloguing standards are evolving – still

Abstract

Just when you thought you had your cataloguing standards all under control, there's a new development lurking in the wings … BIBFRAME.

Two years ago, the cataloguing community embraced Resource Description and Access (RDA) as a new cataloguing standard. The dust is settling, workflows have evolved to accommodate RDA standards; Machine Readable Cataloguing (MARC) has been adapted to carry RDA data and library management systems have been updated.

Meanwhile, the Library of Congress has been working on the Bibliographic Framework Initiative, BIBFRAME. BIBFRAME is being developed to replace MARC as the standard to represent and exchange bibliographic data in the digital information environment. In conjunction with RDA, BIBFRAME will ensure that bibliographic data is able to integrate with and engage in the wider information community.

This paper will raise awareness of BIBFRAME, explaining its underpinning concepts and explore the benefits and implications for resource discovery systems and standards. BIBFRAME is still under development, but it up to everyone who is interested in resource discovery to be aware of what is happening. By being informed, all information workers can be a part of the evolving cataloguing and information discovery landscape.

Introduction

Cataloguing standards have been constantly evolving ever since the first librarian in ancient times decided that information needed to be consistently organised to facilitate access for users (Denton, 2007, p. 35). In the past 20 years this rate of evolution appears to have increased. New cataloguing models and standards have come into existence to meet the needs of an information world dominated by powerful information technologies and, of course, the ubiquitous World Wide Web.

Since 1998, the library world has been grappling with Functional Requirements for Bibliographic Records (FRBR) as a conceptual framework. It organises bibliographic entities (things, people, works, resources etc.) using entity-relationship modelling, based on user needs (find, identify, select and obtain) (Tillett, 2003, pp. 3-5). In 2013, the long-serving standard for bibliographic description, Anglo-American Cataloguing Rules, was replaced with Resource Description and Access (RDA)[1], which is underpinned by FRBR (Mering, 2014, p. xvii).

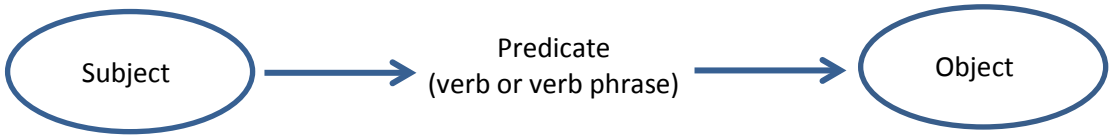
Bibliographic Framework (BIBFRAME) is the current initiative under development in the library world, led by the Library of Congress (LOC) [2]. It is a model for expressing and connecting bibliographic data, with a view to replacing Machine Readable Cataloguing (MARC) (Library of Congress, n.d., para. 1). These new standards have been developed in response to the need for more flexibility in the metadata created to describe information resources in the library. This flexibility allows use of library metadata in a wider range of applications, including the web.

Libraries and other memory organisations, such as archives and museums, have vast quantities of quality data and resources available for free and open use. These resources are curated by organisations which are credible and are trusted by the community. But potential users are often unaware that these resources exist. This data is often locked away in systems incompatible to where potential users undertake their research and discovery, a web search engine. This incompatibility is sidelining library catalogues, library databases and digital repositories as easily accessible, relevant sources of information.

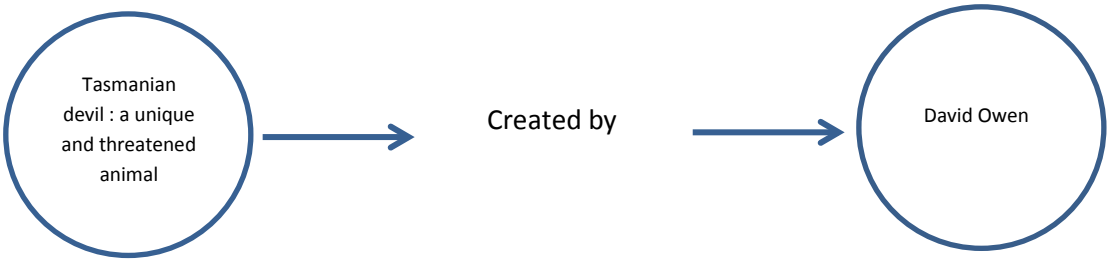
The web, as a data management and integration platform, is constantly evolving and developing more effective standards to enable information discovery. Developments in the semantic web, linked data and Resource Description Framework (RDF) are providing the opportunity for libraries to place their data where the users are. Linked data will enable library resources to be found when a user searches the web for information. BIBFRAME leverages the power of linked data to integrate library data with the web, and vice versa, integrate information on the web with library information.

Linked data and RDF

Underpinning the web are standards that enable users to move from one resource to another, primarily using hyperlinks from one Uniform Resource Locator (URL) to another. A URL is a unique identifier for a webpage. Linked data uses the same concept, but instead of linking from one webpage to another using URLs, linked data enables the linking of unique pieces of data or information using explicit relationships. Linked data is based on a very simple model of triples:



The predicate, or verb phrases, identifies the relationship between the subject and object. For example:



RDF is a standard model used for data interchange on the web (RDA Working Group, 2014, para. 1). RDF extends the linking structure by adding Uniform Resource Identifiers (URI) to all stages of the triple. Subjects and objects of the triple will have unique URIs to identify the specific entity, whether it be a person, place, thing or concept. RDF also clearly identifies relationships (the predicate) between entities using URIs. Unlike URLs, RDF URIs work with small unique granular pieces of data which can be uniquely identified. Below is an example of a linked triple using URIs.

RDF linked data enables the integration of data from different sources. It allows mixing, exposing and sharing of data across different applications. Application of linked data and RDF can already be seen in the web. In 2012, Google developed its knowledge graph, which gathers information from a wide variety of sources, including DBpedia (the linked data version of Wikipedia), and presents it to the user, along with search results. A search for Tasmanian devil in Google will show a list of search results as well as information gathered from various sources using linked data standards (Figure 1). The BBC Nature website uses linked data to automatically create pages of information drawn from the BBC's information repositories, including up-to-date news items (Figure 2).

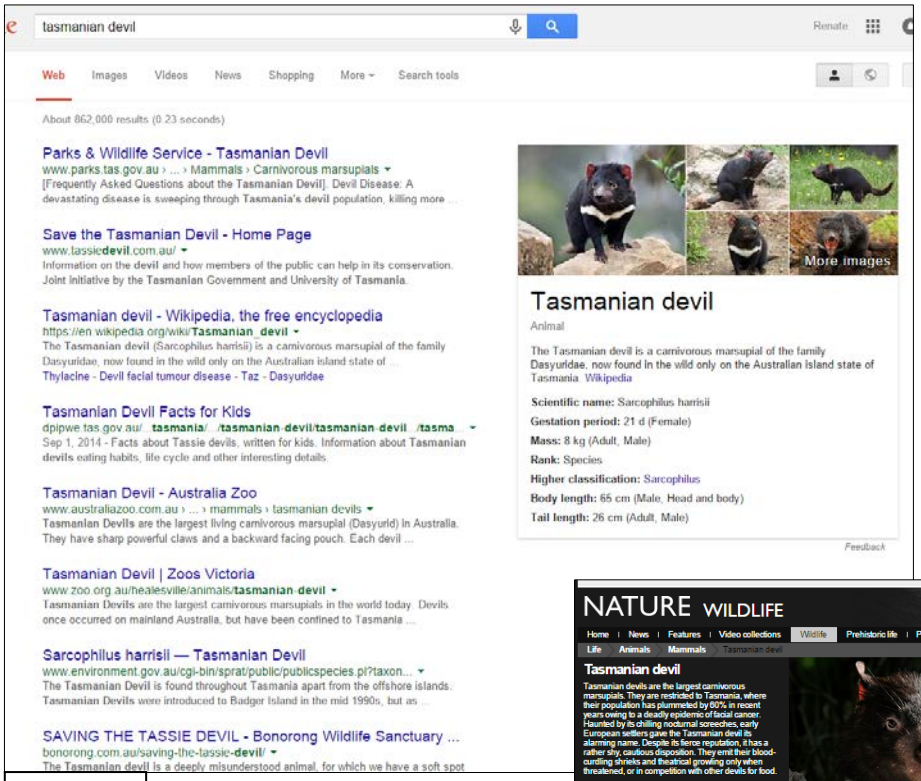


Figure 1 https://www.google.com.au/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=tasmanian%20devil

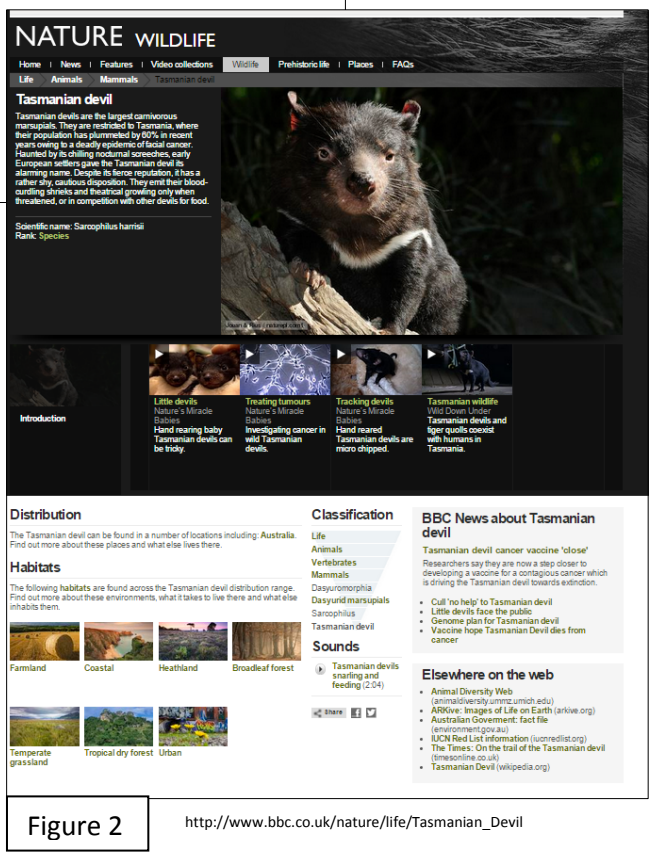


Figure 2 http://www.bbc.co.uk/nature/life/Tasmanian_Devil

Library bibliographic data lends itself to easily become linked data using RDF standards. URIs are being added to bibliographic entities, authorities and works across the world. Subject ontologies such as Schools Online Thesaurus[3] (Figure 3) and Library of Congress Subject Headings[4] (LCSH) (Figure 4 allocate URIs for subject headings. The Virtual International Authority File[5] (VIAF) is creating URIs for personal names, corporate names and geographic places (Figure 5).

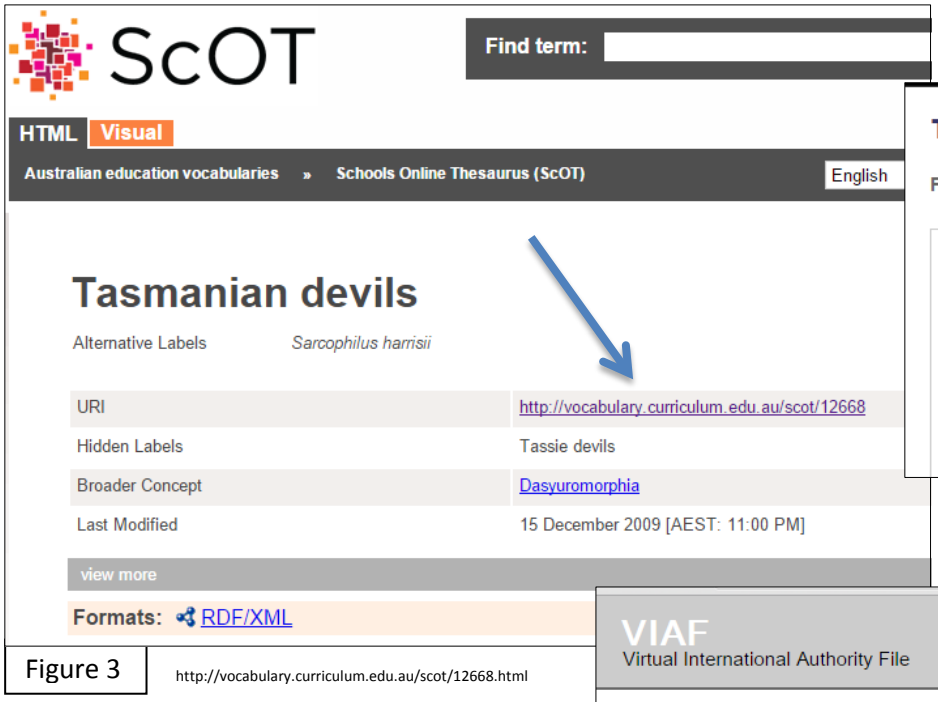


Figure 3 http://vocabulary.curriculum.edu.au/scot/12668.html

http://id.loc.gov/authorities/subjects

Figure 4

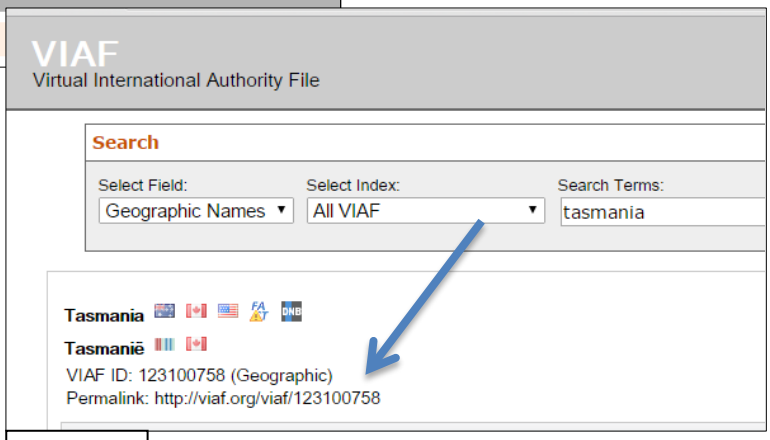


Figure 5 http://www.viaf.org/viaf/123100758/#Tasmania

In RDA, consistent terminology is used to uniquely identify relationships. RDA bibliographic data can have URIs allocated to them. For example the RDA relator terms for creators, contributors, publishers, owners etc. are can become URIs. Relationships such as 'sequel to', 'adaptation of' are clearly laid out in RDA and can be allocated unique URIs. This will enable linked data manipulation of library records.

By using RDF standards to organise bibliographic data, it is possible to place library data where users search. For example, a search on a book title in Google, could tell the searcher where to buy it (Figure 6), and what libraries nearby hold it.

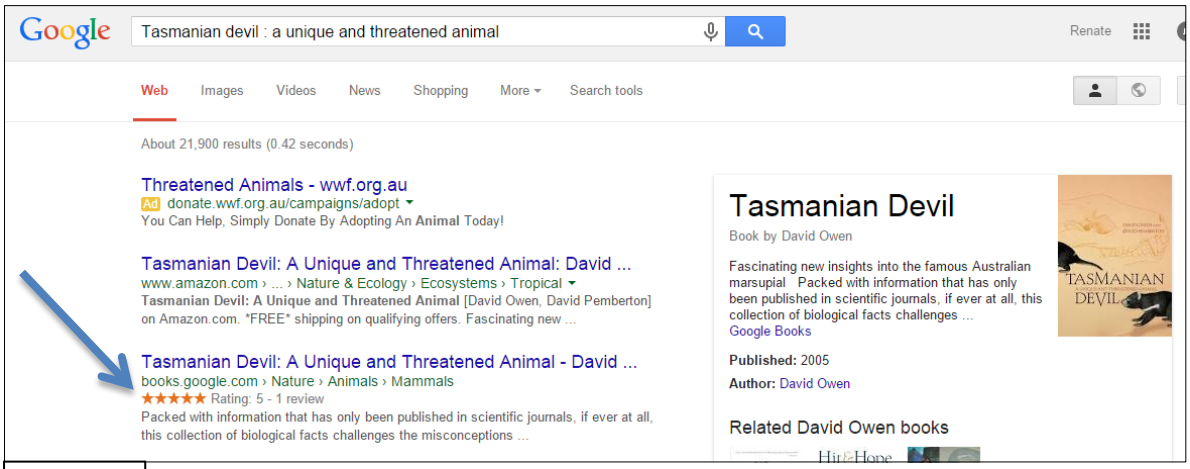


Figure 6 https://www.google.com.au/?gws_rd=ssl#q=tasmanian+devil:+a+unique+and+threatened+animal

For data to work with RDF, it has to be able to be pulled apart and re-assembled into different shapes for different purposes, just like Lego, a metaphor used by Eric Miller from Zepheira (2013, slide 18). The library community has great information assets and data. However, most of the current library data uses the MARC communication format which does not lend itself to the requirements of RDF.

MARC

In 2002, Roy Tennant wrote a popular article titled ‘MARC must die’. He argued that a more flexible encoding standard than MARC is required (Tennant, 2002). Thirteen years later, in February 2015, Lori Bowen Ayre wrote ‘MARC isn’t dying fast enough’. She claimed that library catalogues are disconnected from the web and from library users due to the unique nature of MARC records (Ayre, 2015).

MARC is solely used in library repositories. No other information systems can use the data in that format; MARC data can only be used by other MARC-based systems. Even with the introduction of MARCXML, MARC for the XML environment, MARC records are still difficult to share with non-MARC systems.

MARC lacks flexibility because it works in records and strings of data within the record. Its records are independent and do not rely on relationships with other records. Each MARC record aggregates information about the resource content and concepts as well as details about the physical carrier.

The Bibliographic Framework Initiative is focussing on new ways of presenting the bibliographic data that already exists in MARC records, so it can be used in more flexible ways to engage with the wider information community and the networked world.

BIBFRAME

Unlike MARC, which focuses on whole records for a resource, BIBFRAME focusses on granular data. In place of making a record for a resource, BIBFRAME creates relationships and links between entities/things such as works, people, concepts, places and languages. To enable bibliographic data from libraries and other information repositories to be part of the web, BIBFRAME has developed a simple model for recording unique pieces of data and relationships between them. BIBFRAME is a model for creating linked data triples using URIs for bibliographic data.

The BIBFRAME model is organised into core or high-level classes, each of which has a defined set of properties. These main classes are: Work, Instance, Authority and Annotation. The Bibliographic Framework Initiative model is described in Figure 7 (Library of Congress, 2012, pp. 8-9).

- ***Creative Work*** - a resource reflecting a conceptual essence of the cataloging item.
- ***Instance*** - a resource reflecting an individual, material embodiment of the Work.
- ***Authority*** - a resource reflecting key authority concepts that have defined relationships reflected in the Work and Instance. Examples of Authority Resources include People, Places, Topics, Organizations, etc.
- ***Annotation*** - a resource that decorates other BIBFRAME resources with additional information. Examples of such annotations include Library Holdings information, cover art and reviews.

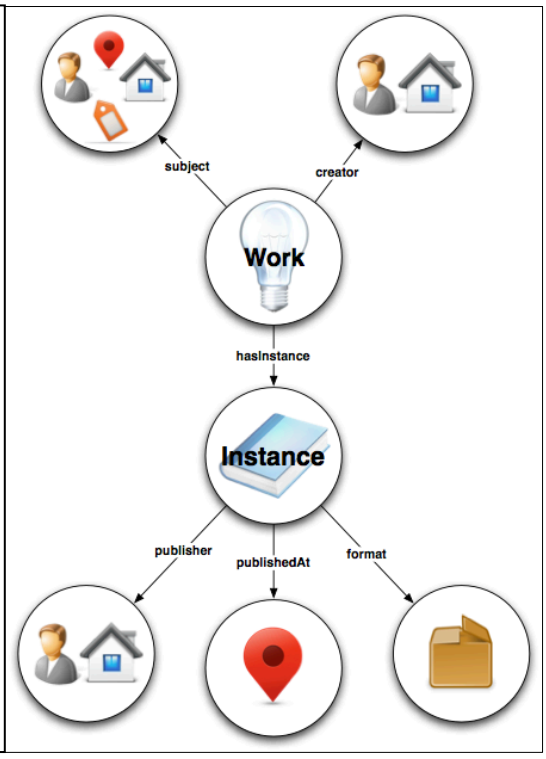


Figure 7

Reproduced from Library of Congress (2012). *Biographic Framework as a web of data: linked data model and supporting services*. Retrieved May 29, 2015 from http://www.loc.gov/bibframe/pdf/marcld-report-11-21-2012.pdf

The concept *work* in BIBFRAME has a very broad application; almost anything fits into the concept *work*. It is a point of commonality of content between *instances* of the *work* and includes relationships to *authorities* such as topic, person, place, organisation, meeting etc. *Instance* can be physical or digital in nature; it is a control point for describing the physical properties, as well as relationships to entities such as publisher. An *instance* is only related to one *work*.

Authorities in BIBFRAME represent a concept or thing, such as person, family, meeting, place and topic. *Authorities* provide control points that help navigate *works* and *instances*. BIBFRAME *authorities* will use existing web-based authority standards, such as VIAF and LCSH.

Annotations are additional information which could be considered subjective, so it is important to know who has provided this information, for example, reading levels, ratings and reviews. Different authorship will not change the data in *works* and *instances*, while authorship will influence content of *annotations* (McCallum, 2012, slide 26). Holdings details are also annotations, they change depending on the authorship (Figure 8).

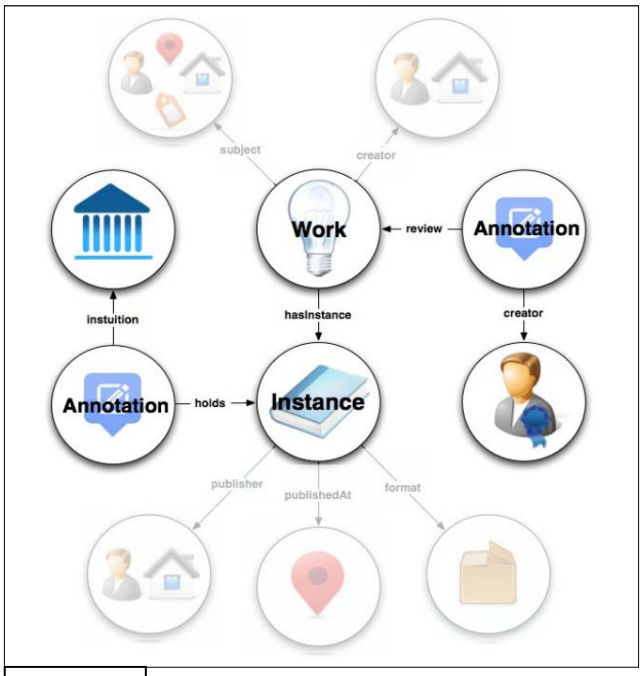
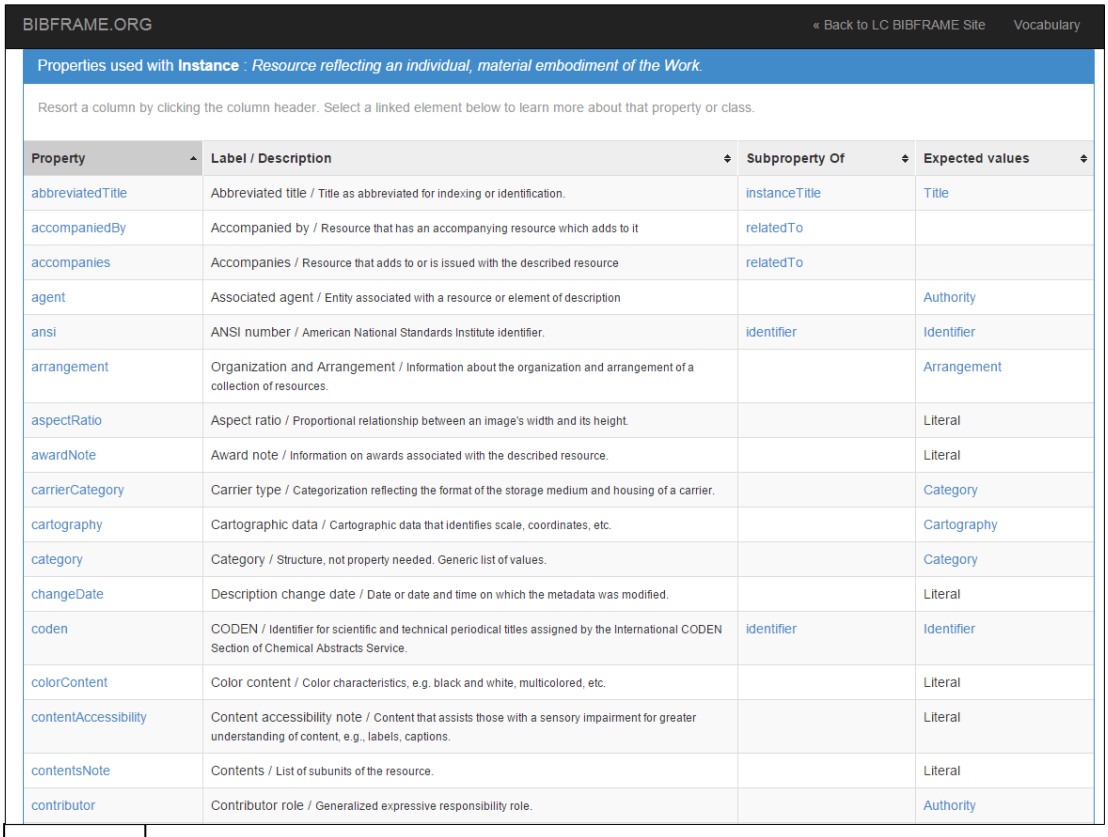


Figure 8

Reproduced from McCallum, S. (2012). *Bibliographic Framework Initiative approach for MARC data as linked data.* Retrieved June 18, 2015 from http://www.loc.gov/bibframe/pdf/BFI-IGeLU-ppt2003.pdf

The BIBFRAME model is simpler than RDA's (i.e., FRBR's) groups of entities, though it easily maps to them. RDA works and expressions are BIBFRAME *works*; RDA manifestations are BIBFRAME *instances*; and RDA items are one type of BIBFRAME *annotation*. The simplicity of the BIBFRAME model ensures that bibliographic data is understood outside the bibliographic community, including in the web community.

The BIBFRAME vocabulary is a key component of the model. It defines the metadata used to describe *works*, *instances*, *authorities* and *annotations*. The vocabulary is comprised of RDF properties. Most importantly the BIBFRAME vocabulary defines relationship between and among the properties. RDA elements are used in the BIBFRAME vocabulary, incorporating both RDA relationship designators and attributes. Figure 9 is a small extract from the BIBFRAME vocabulary[6] relating to the properties used in BIBFRAME *instance* illustrating relationship and attribute elements.



BIBFRAME.ORG « Back to LC BIBFRAME Site Vocabulary

Properties used with **Instance** : *Resource reflecting an individual, material embodiment of the Work.*

Resort a column by clicking the column header. Select a linked element below to learn more about that property or class.

Property	Label / Description	Subproperty Of	Expected values
abbreviatedTitle	Abbreviated title / Title as abbreviated for indexing or identification.	instanceTitle	Title
accompaniedBy	Accompanied by / Resource that has an accompanying resource which adds to it	relatedTo	
accompanies	Accompanies / Resource that adds to or is issued with the described resource	relatedTo	
agent	Associated agent / Entity associated with a resource or element of description		Authority
ansi	ANSI number / American National Standards Institute identifier.	identifier	Identifier
arrangement	Organization and Arrangement / Information about the organization and arrangement of a collection of resources.		Arrangement
aspectRatio	Aspect ratio / Proportional relationship between an image's width and its height.		Literal
awardNote	Award note / Information on awards associated with the described resource.		Literal
carrierCategory	Carrier type / Categorization reflecting the format of the storage medium and housing of a carrier.		Category
cartography	Cartographic data / Cartographic data that identifies scale, coordinates, etc.		Cartography
category	Category / Structure, not property needed. Generic list of values.		Category
changeDate	Description change date / Date or date and time on which the metadata was modified.		Literal
coden	CODEN / Identifier for scientific and technical periodical titles assigned by the International CODEN Section of Chemical Abstracts Service.	identifier	Identifier
colorContent	Color content / Color characteristics, e.g. black and white, multicolored, etc.		Literal
contentAccessibility	Content accessibility note / Content that assists those with a sensory impairment for greater understanding of content, e.g., labels, captions.		Literal
contentsNote	Contents / List of subunits of the resource.		Literal
contributor	Contributor role / Generalized expressive responsibility role.		Authority

Figure 9 http://bibframe.org/vocab/Instance.html

The BIBFRAME vocabulary uses RDF structures to enable deconstruction and reconstruction of the information assets that comprise MARC. It makes the data more flexible and able to be manipulated in a myriad of ways in the linked data environment of the web, using linked data triples and URIs.

Implications for cataloguing

Like MARC, BIBFRAME is a communication standard. Both are containers for the bibliographic content created by cataloguers using RDA. RDA is the descriptive standard, providing instructions for how to record the content of the records and how to create access points and define relationships. BIBFRAME has been developed with RDA in mind and, at the same time, is flexible enough to be used with other content standards.

MARC is very familiar to cataloguers, but it only works in library-based systems. BIBFRAME works in the wider linked data world. In the future BIBFRAME will replace MARC as the communication standard. Cataloguers will not need to learn RDF language; rather they will create records using templates which are even simpler than current MARC editors.

BIBFRAME developers have created a number of tools to enable demonstration and testing of the BIBFRAME model and vocabulary[7]: BIBFRAME Editor[7], BIBFRAME Profile Editor and MARC to BIBFRAME transformation tools. The Comparison Service enables transformation of a LOC MARC record into a record using BIBFRAME resources (Figure 10)

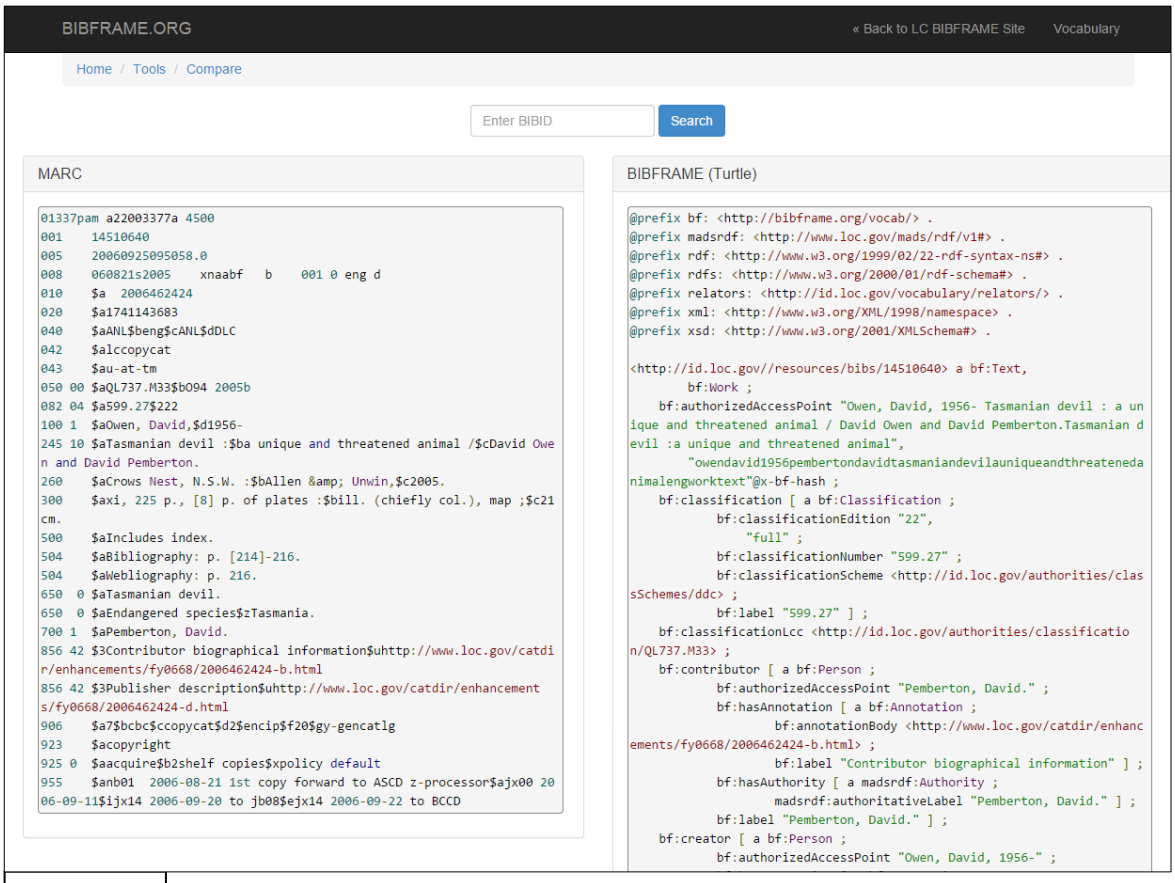


Figure 10 http://bibframe.org/tools/compare/bibid/1451064

The demonstration version of BIBFRAME Editor (BFE) enables recording of bibliographic description using the BIBFRAME vocabulary. At this stage the editor does not store input data for retrieval. This tool demonstrates how cataloguers can input BIBFRAME data without knowledge of RDF protocols. In the BFE there are two profiles: Kitchen Sink Profile, which includes all BIBFRAME properties, and Simple Monograph. Figure 11 is a sample of the BFE template for a new *instance* using the Simple Monograph profile. Cataloguing institutions can create their own profile using the BIBFRAME Profile Editor. This tool enables configuration of profiles to suit institutional needs.

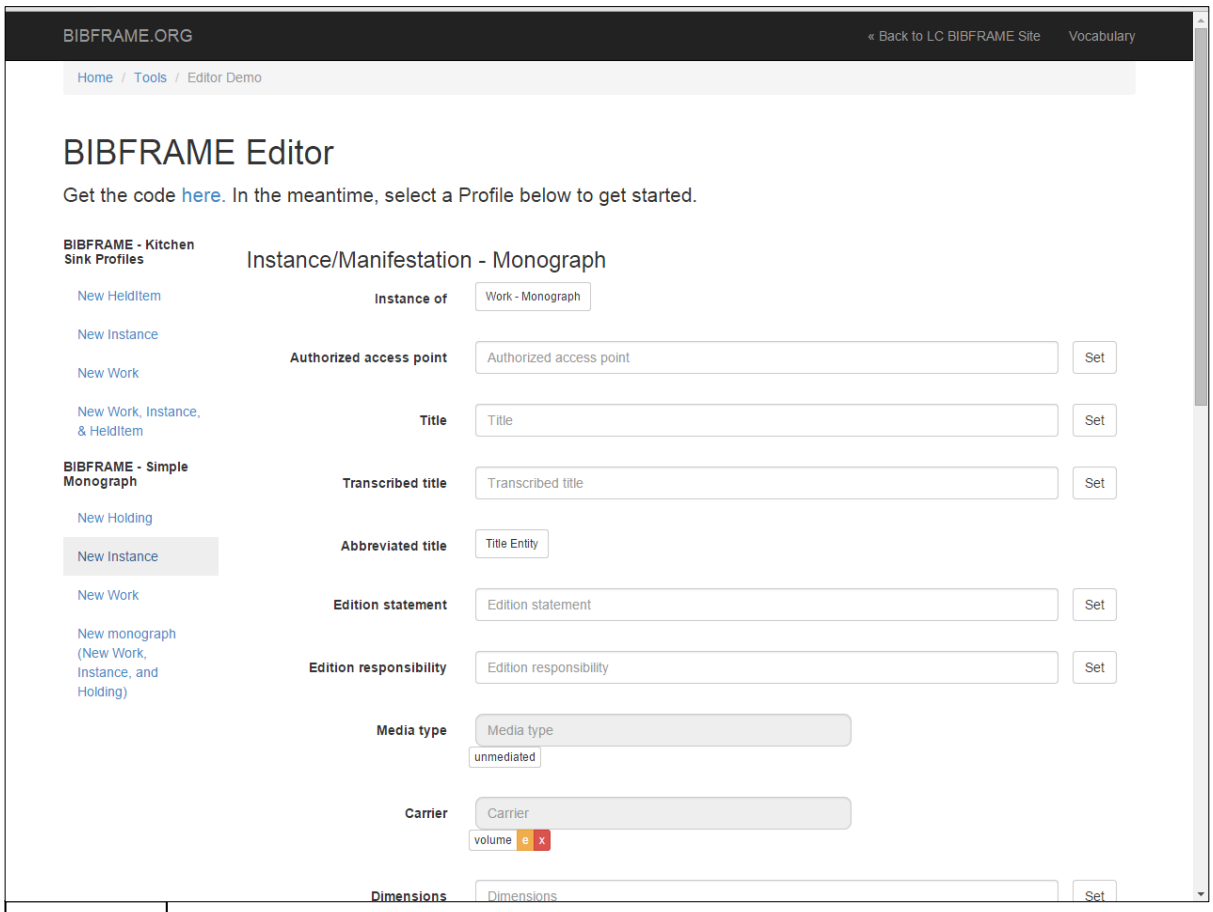


Figure 11 http://bibframe.org/tools/editor/#

Now what?

The Bibliographic Framework Initiative, spearheaded by LOC, is a 'work in progress'. The BIBFRAME model, vocabularies and tools are still being tested and developed. MARC will be replaced by BIBFRAME in the future, though MARC as a legacy communication format will continue to play a role in bibliographic information management.

BIBFRAME is focussing on the background data, it is not creating library management systems or online public access catalogues. BIBFRAME focusses on ensuring bibliographic data works in a wide variety of information and discovery systems. It is up to library systems developers and vendors to create information management systems and discovery tools suited to user needs.

Library system vendors are getting involved in implementing BIBFRAME. The most recent LOC BIBFRAME Update Forum, June 2015, incorporated presentations by vendors on how they will incorporate BIFRAME tools into their system development strategies. Vendors also spoke about how they plan to take advantage of the linkages that are increasingly going to be included in BIBFRAME data.

BIBFRAME is still under development, but it up to everyone who is interested in resource discovery to be aware of what is happening. By being informed, all information workers can be a part of the evolving cataloguing and information discovery landscape.

Final words

“BIBFRAME [is] the format that we expect to replace MARC in the future … that could be two years, five years, but definitely is going to lead us into the linked data world.” Beacher J.E. Wiggins (Director for Acquisitions and Bibliographic Access at the Library of Congress) (Library of Congress, 2014).

“So what is BIBFRAME all about? It's not just recognizing that we are in a network world, but fully embracing that fact. It's about not just being on the web, but being of the web … [It’s about] getting our resources in the hands of our patrons. We want the search engines to work to our advantage.” Kevin Ford (Digital project coordinator in the Network Development and MARC Standards Office at the Library of Congress) (Library of Congress 2014).

“So let’s agitate for library systems that leave MARC behind ... Let’s make sure the process to replace MARC doesn’t take another decade, and that, when it is done, we end up with something that really will help us do our jobs and participate with the rest of the online community” Lori Bowen Ayre (Head of Galecia Group) (2015, para. 16).

References

Ayre, L.B. (2015). MARC isn't dying fast enough. *Collaborative Librarianship 6*(4), February 25. Retrieved from http://galecia.com/presentations/2015/marc-isnt-dying-fast-enough

Denton, W. (2007). FRBR and the history of cataloguing. In Taylor, A.G. (ed.), *Understanding FRBR: what it is and how it will affect our retrieval tools.* Westport, Conn.: Libraries Unlimited.

Library of Congress (2012). *Biographic Framework as a web of data: linked data model and supporting services*. Retrieved May 29, 2015 from http://www.loc.gov/bibframe/pdf/marcld-report-11-21-2012.pdf

Library of Congress (2014). *BIBFRAME: update & practical applications* [webcast]. Retrieved June 19, 2015 http://www.loc.gov/today/cyberlc/feature_wdesc.php?rec=6401

Library of Congress (n.d.). *Bibliographic Framework Initiative.* Retrieved June 2, 2015 from http://www.loc.gov/bibframe/

McCallum, S. (2012). *Bibliographic Framework Initiative approach for MARC data as linked data.* Retrieved June 18, 2015 from http://www.loc.gov/bibframe/pdf/BFI-IGeLU-ppt2003.pdf

Mering, M. (Ed.). (2014). *The RDA workbook: learning the basics of resource description and access.* Santa Barbara, California: Libraries Unlimited.

Miller, E. (2013). *BIBRAME and the World Wide Web.* Retrieved June 18, 2015 from http://www.loc.gov/bibframe/pdf/ALAmw2013-bibframe-lcupdate-20130127_Miller.pdf

RDF Working Group (2014). *Resource Description Framework (RDF).* Retrieved June 5, 2015 from http://www.w3.org/RDF/

Tennant, R. (2002). MARC must die. *Library Journal*, October 15. Retrieved from http://lj.libraryjournal.com/2002/10/ljarchives/marc-must-die/#_

Tillett, B. (2003). *What is FRBR? A conceptual model for the bibliographic universe.* Retrieved June 15, 2015 from http://www.loc.gov/cds/downloads/FRBR.PDF

Webliography

1. RDA Toolkit http://www.rdatoolkit.org/
2. Bibliographic Framework Initiative http://www.loc.gov/bibframe/
3. Schools Online Thesaurus http://scot.curriculum.edu.au
4. LC Linked Data Service http://id.loc.gov
5. Virtual International Authority File http://viaf.org
6. BIBFRAME Vocabulary http://bibframe.org/vocab/

7. BIBFRAME Tools & Downloads http://www.loc.gov/bibframe/tools/