

Introduction

Following the replacement of AACR2 by RDA, BIBFRAME is widely viewed as the replacement for MARC. Much like MARC, it was initiated by the Library of Congress. BIBFRAME is an abbreviation - not an acronym despite the capitalisation - for the BIBliographic FRAMEwork Initiative. It is also frequently written as "Bibframe" which this article will use from now on. Bibframe's initial remit was wider than establishing a technical standard¹, although that is the current major focus of its work and of this article. This article will try to assess what Bibframe's impact is likely to be, especially on cataloguing.

The Bibframe initiative was initially undertaken with the consultants Zepheira² - whose president Eric Miller was involved with the RDF specification itself - in partial response to RDA testing in 2011 which determined that MARC21 wasn't up to handling RDA properly:

"Most felt any benefits of RDA would be largely unrealized in a MARC environment. MARC may hinder the separation of elements and ability to use URIs in a linked data environment."³

For RDA to be adopted, the Committee suggested that the national libraries

"Demonstrate credible progress towards a replacement for MARC."

The Library of Congress decided in 2011 to use linked data as the basis of a replacement:

"The new bibliographic framework project will be focused on the Web environment, Linked Data principles and mechanisms, and the Resource Description Framework (RDF) as a basic data model."⁴

This was a highly significant decision, both for determining the shape of Bibframe and the context in which it is to be placed.

1. "Bibliographic framework is intended to indicate an environment rather than a "format"" – Library of Congress. *A Bibliographic Framework for the Digital Age*. 31 October 2011. <http://www.loc.gov/bibframe/news/framework-103111.html>
2. Zepheira. Zepheira homepage. <https://zepheira.com/>
3. U.S. RDA Test Coordinating Committee. Report and Recommendations of the U.S. RDA Test Coordinating Committee. Executive Summary. 2011. P. 8. http://www.nlm.nih.gov/tsd/cataloging/RDA_report_executive_summary.pdf
4. Library of Congress. *A Bibliographic Framework for the Digital Age*. 31 October 2011. <http://www.loc.gov/bibframe/news/framework-103111.html>

Linked Data

Linked data is not a tightly defined technical standard but an approach and set of technologies that aim to bring the benefits of the web to data, not just to documents. This is how Tim Berners-Lee described it in 2006:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs so that they can discover more things.⁵

URIs are basically URLs used as identifiers. In cataloguing we commonly use authorised strings as identifiers. These change when life events or rules change, but URIs do not: they are more like system numbers but are not confined to particular databases and are universal. HTTP URIs are simply URIs on the web. When a human looks up a URI they should get a textual document in HTML to read; when a computer looks up the same URI, it will prefer instead some data in RDF, which provides a way of representing data on the web. What made the web good was linking to other things, and following your nose to find out more. The same is also the case for linked data. In cataloguing terms, you don't have to provide all the information about an author in the record for a book if you can follow a link to an authority record or to a page about the author.

In short, linked data gives us a web of data rather than a web of documents, and it is RDF that gives linked data its basic shape.

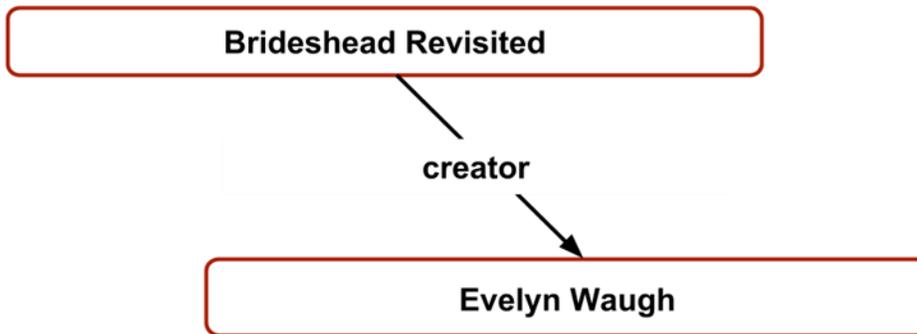
RDF

RDF is a data model based (as is FRBR) on the entity-relationship model. All facts or assertions are expressed as *triples*. An assertion such as:

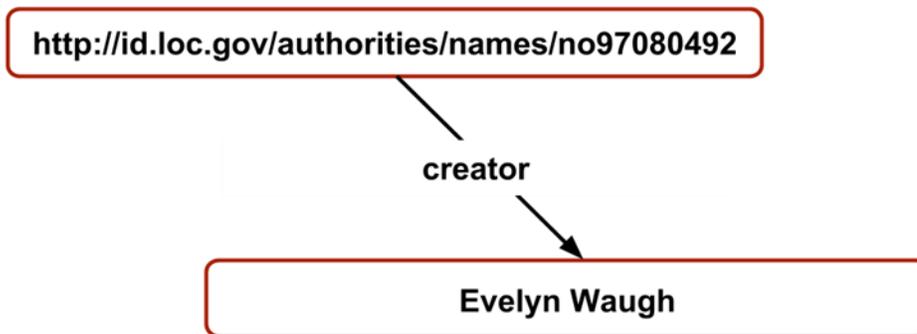
Brideshead revisited *was written by* **Evelyn Waugh**

can be broken down into two entities (“**Brideshead revisited**” and “**Evelyn Waugh**”) and a relationship (“*was written by*” or “*creator*”):

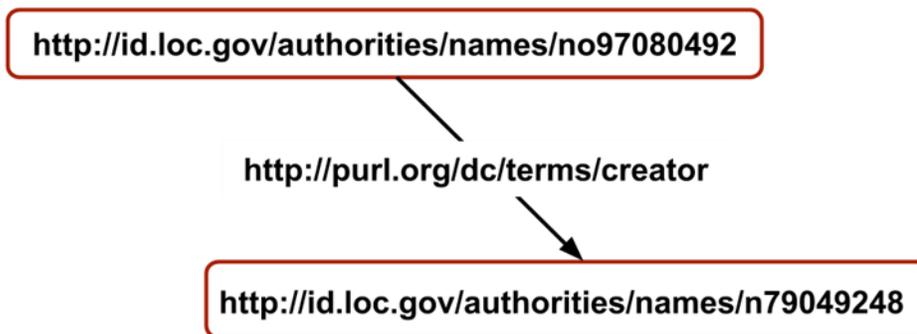
5. Berners-Lee, Tim. Linked Data: Design Issues. 2009. <http://www.w3.org/DesignIssues/LinkedData.html>



Rather than using text to identify the book in this case, we can use a URI to unambiguously identify which book we are talking about. The Library of Congress has published a URI for Brideshead Revisited:



Similarly, LC have a URI for Evelyn Waugh, and Dublin Core have a URI for the creator relationship itself:



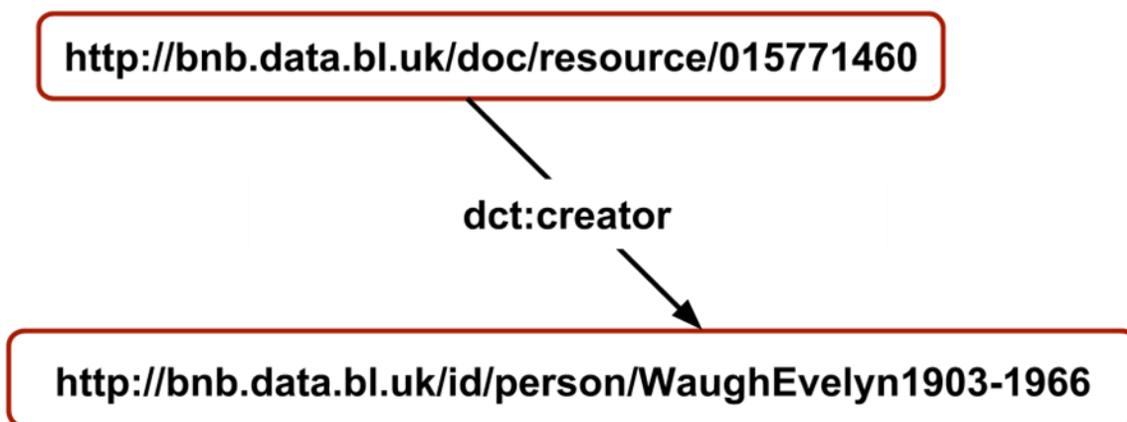
This is a triple expressed graphically. We could also express this as text, with some abbreviations to make it readable:

@prefix lcn: <http://id.loc.gov/authorities/names/> .

@prefix dcterms: <http://purl.org/dc/terms/> .

lcn:015771460 dcterms:creator lcn:n79049248 .

This way of writing RDF is called turtle⁶. There are several such ways to write RDF, the most common one being RDF/XML, although this is harder for people to read. The BNB (British National Bibliography) has released a large amount of linked data. Here is a real example based on the way the BNB would say the same thing, first as a graphic...



... and as turtle:

@prefix bnbr: <http://bnb.data.bl.uk/doc/resource/> .

@prefix bnbp: <http://bnb.data.bl.uk/id/person/> .

@prefix dcterms: <http://purl.org/dc/terms/> .

bnbr:015771460 dct:creator bnbp:WaughEvelyn1903-1966 .

6. Wikipedia. *Turtle (Syntax)*. http://en.wikipedia.org/wiki/Turtle_%28syntax%29

The Impact of Using Linked Data

Using linked data sends a number of powerful messages. First, Tim Berners-Lee thinks it is a good idea! Linked data builds upon the incredible success of the web and extends it from documents to data.

Second, it is not a cataloguing standard, nor even a library standard, but a web standard. RDF and HTTP for instance are W3C-maintained standards. This gives libraries and catalogues the opportunity to escape from many of the silos we have complained about being trapped in. By contrast, MARC is firmly a cataloguing standard and arguably not even a library-wide standard: libraries are full of repositories, archives, and databases using anything but MARC. MARC also requires highly specialised software to use it and makes it difficult to share our data or include others' data.

Third, linked data doesn't require but often assumes a notion of openness wholly missing from MARC. Most MARC records are essentially unlicensed where the assumption must be that they are copyrighted. By releasing data openly under a specific licence, people can re-use or re-purpose our data. It fits in really well with the open access, open software, and open data movements.

A number of libraries have already published linked bibliographic data. The following is merely a selection⁷:

2008 Swedish National Library⁸

2011 BNB⁹

Cambridge University Library¹⁰

Europeana¹¹

French National Library¹²

7. The list doesn't include a number of authority files such as VIAF (OCLC's Virtual Authority File), the Library of Congress (including its Name Authorities and LCSH), and the German National Library.
8. LIBRIS nationella bibliotekssystem. LIBRIS available as Linked Data. 3 December 2008. <http://librisbloggen.kb.se/2008/12/03/libris-available-as-linked-data/>
9. Wallis, Richard. Significant Bibliographic Linked Data Release From The British Library. 14 July 2011. <http://dataliberate.com/2011/07/significant-bibliographic-linked-data-release-from-the-british-library/>
10. European Library. Case Study: Cambridge University Library delivers linked open data and enrichments. 29 April 2014. <http://www.theeuropeanlibrary.org/tel4/newsitem/5450>
11. Europeana. EDM Case Study: Europeana LOD pilot. 2011? <http://pro.europeana.eu/europeana-lod-pilot-edm>
12. European Library. The BnF Transforms the Visibility of its Resources with Linked Data. 23 April 2014. <http://www.theeuropeanlibrary.org/tel4/newsitem/5350>

2012 OCLC Worldcat (using schema.org)¹³

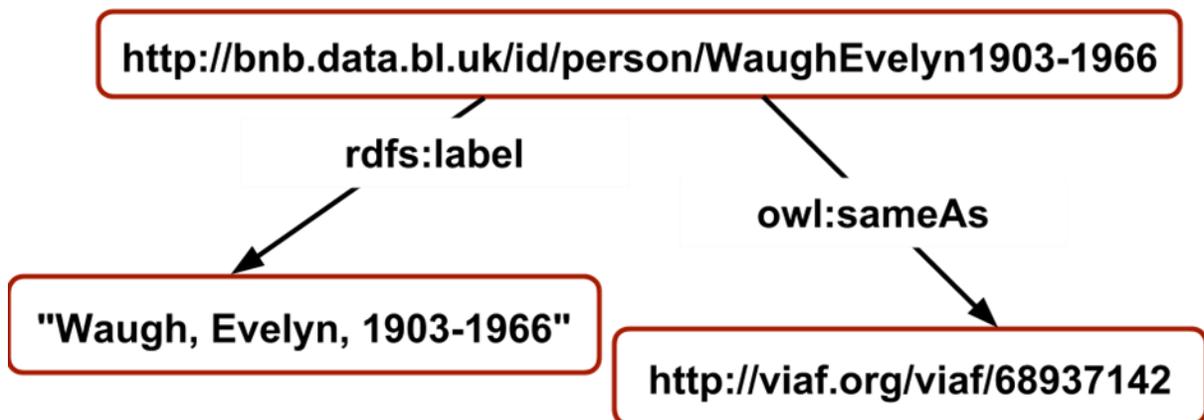
Spanish National Library¹⁴

2014 RLUK (as part of the European Library)¹⁵

These libraries have not only already established models but have actually published data. It is also worth noting that these models all differ, using different properties from different vocabularies.. This makes sense if the needs and priorities of these libraries are not homogenous.

Bibframe as RDF

So what does Bibframe actually look like? Above we looked at an example from the BNB. Below we can see more statements made about the author by the BNB:



Note that three vocabularies have been used by the BNB for various properties here: Dublin Core Terms¹⁶ (dct), RDF Schema¹⁷ (rdfs), and OWL, the Web Ontology Language¹⁸ (owl).

13. OCLC. OCLC adds Linked Data to WorldCat.org. 30 June 2012. <http://www.oclc.org/news/releases/2012/201238.en.html>

14. European Library. Case Study: National Library of Spain transforms the user experience of Linked Open Data. 16 May 2014. <http://www.theeuropeanlibrary.org/tel4/newsitem/5800>

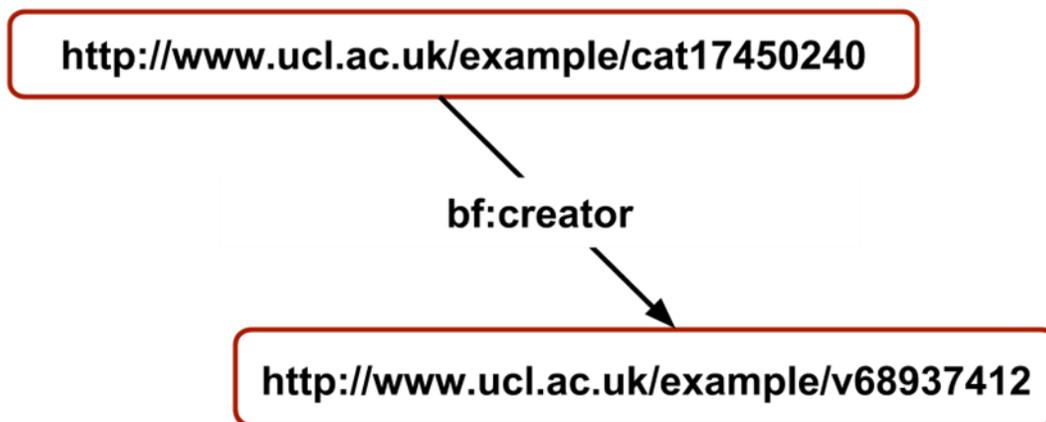
15. RLUK. RLUK Hack Day #RLUKhack. 2014. <http://www.rluk.ac.uk/events/rluk-hack-day/>

16. Dublin Core Metadata Initiative. DCMI Metadata Terms. 2014. <http://dublincore.org/documents/dcmi-terms/>

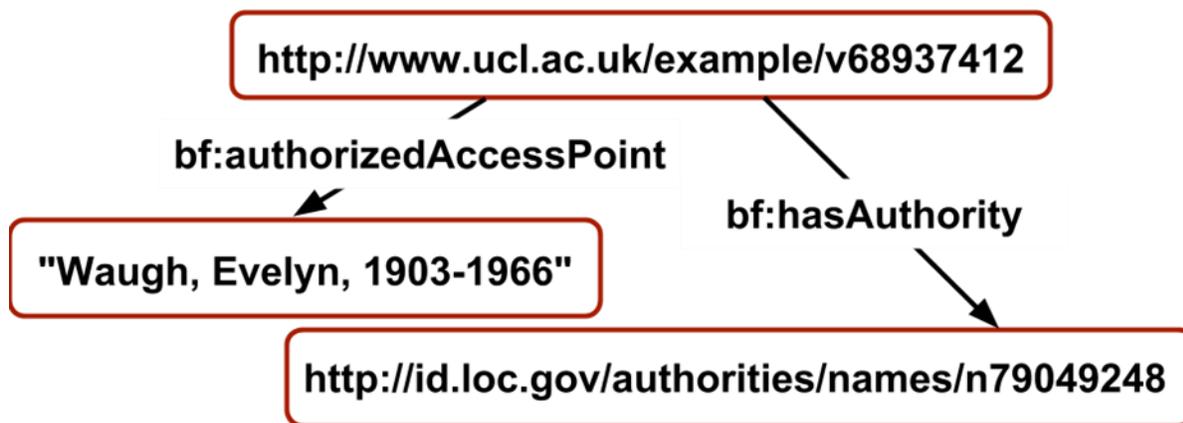
17. OWL. RDF Schema 1.1. 2014. <http://www.w3.org/TR/rdf-schema/>

18. W3C. OWL Web Ontology Language Reference. 2014. <http://www.w3.org/TR/owl-ref/>

Here is the Bibframe equivalent of the triple, asserting Waugh to be the author of Brideshead Revisited. It uses an imaginary implementation at UCL:



There is a Bibframe-specific version of the creator property with a locally authorised author. Here is some more detail about that author:



The thing to note here is that both additional properties - `bf:authorizedAccessPoint` and `bf:hasAuthority` - are also Bibframe-specific. This is significant. To get a better appreciation of this, let's look at the description for a whole book, the equivalent of a MARC record:

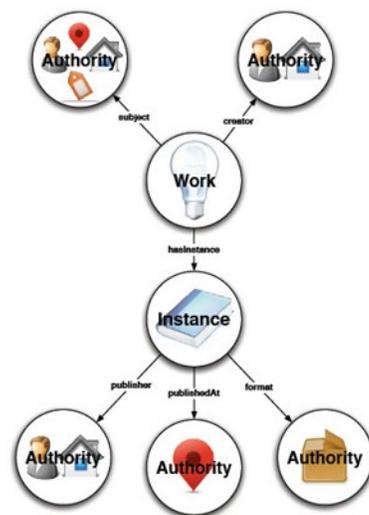
report from the Library of Congress said that:

“While the recommendation of a singular namespace is counter to several current Linked Data bibliographic efforts, it is crucial to clarify responsibility and authority behind the schematic framework of BIBFRAME in order to minimize confusion and reduce the complexity of the resulting data formats.”²³

Although it is hard to see that maintaining a single data format for all bibliographic purposes will necessarily reduce complexity!

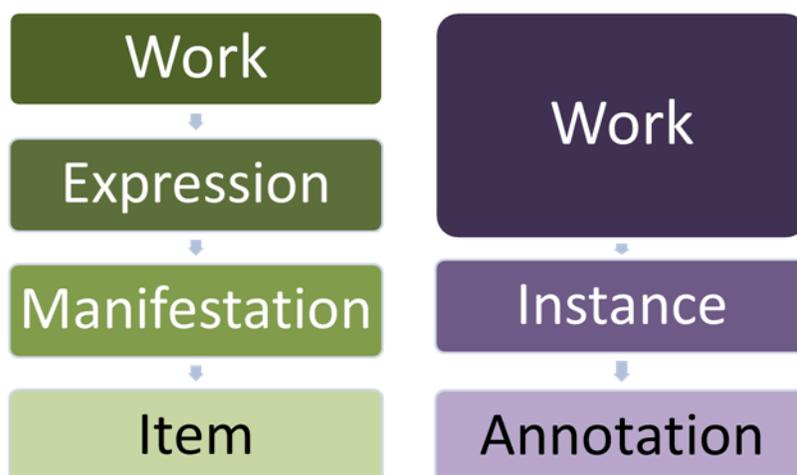
Relying on others is also potentially risky for stability. If libraries start using Dublin Core and it disappears, or the defined meanings change, then what? Arguably though, URIs still retain their meaning by consent even if there is no content on the web to back them up. The Bibframe approach also has the opposite flaw of putting all its eggs in one basket: if Bibframe fails then more would be lost than with a single smaller vocabulary. It is perhaps instructive to consider that many libraries used - and still use - UKMARC despite its official demise some years ago. The Bibframe plan is instead to stabilize the vocabulary then make mapping links as appropriate to other vocabularies.

Below is the common image showing the Bibframe model which demonstrates one of its most unsettling aspects for cataloguers: that it doesn't follow FRBR!



23. Library of Congress. Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services. Washington, D.C.: Library of Congress, 2012. P. 15. <http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>

The Work looks like a FRBR Work; the Instance looks very like a FRBR Manifestation; the Authority looks like any of the group 2 or 3 entities: authors, subjects, etc.; but where is the Expression? In reality, a Bibframe work can be both a FRBR work and a FRBR Expression, depending on how it describes itself:



This though does make the point that Bibframe is designed to handle RDA but also lots of things that are not RDA. Most AACR2 records don't have Expression records either and there are other FRBR-like models (e.g. CIDOC CRM²⁴) to take account of if Bibframe is to move beyond purely accommodating library catalogue data.

Maintenance of Bibframe

Bibframe is an initiative of the Library of Congress. There is as yet no governing committee for Bibframe although I understand²⁵ that this will be widened out to a similar governance system to that of MARC.²⁶ This is arguably a good thing as it has at least got off the ground in the absence of another clear mechanism; it means Bibframe is not in commercial hands; and it should have a strong central direction. On the other hand, a single large national library, albeit in a pivotal role, is not necessarily representative of the libraries that will want to use it. Some of the partners and early experimenters involved in Bibframe- e.g. George Washington University, the German National Library, and OCLC²⁷– do have a wider field of interest, but do not have

24. International Council of Museums. The CIDOC Conceptual Reference Model. 2013. <http://www.cidoc-crm.org/>

25. Personal correspondence with Kevin Ford on Twitter. 12 August 2014. <https://twitter.com/3windmills/status/499198681506406400>

26. See: Library of Congress. MARC Advisory Committee. <http://www.loc.gov/marc/mac/advisory.html>

27. Library of Congress. BIBFRAME Frequently Asked Questions. 13. <http://www.loc.gov/bibframe/faqs/#q13>; Library of Congress. BIBFRAME Implementation Register. <http://www.loc.gov/bibframe/implementation/register.html>

ownership as such. Consultation via the mailing list²⁸ has been increasingly vibrant, especially in response to some specific criticisms, but decision making is still in the hands of the Library of Congress. Not involving commercial entities from the beginning is also potentially risky as their input will be vital to any scheme's adoption in practice, especially with the centralised standards environment we currently work in.

The Purpose of Bibframe

The library linked data efforts mentioned above - e.g. those of the BNB or Swedish National Library - were instigated by those libraries for their own purposes. They drew on others' work, but did them for their own business reasons, even if those reasons were experimental; Bibframe by contrast is interesting in being undertaken by one national library for the benefit of the community for the broadest of purposes:

“BIBFRAME provides a foundation for the future of bibliographic description, both on the web, and in the broader networked world.”²⁹

MARC, Bibframe's predecessor, was a very similar creation by the Library of Congress but with a much narrower scope, which was testing “the feasibility of distributing Library of Congress cataloguing in machine-readable form to a variety of users” although it was recognised that “devising a method of recording bibliographical information in machine-readable form was basic to the solution of other problems.”³⁰

Given the low degree of automation at the time, MARC's success was remarkable; given the massively entrenched infrastructure of thousands of libraries, differing LMS's, vendors, and cooperative sharing schemes in existence now, the task facing Bibframe is daunting from the opposite point of view: the Library of Congress, basically single-handedly, is seeking to change an entire ecosystem and market on behalf of that ecosystem and market!

Bibframe in a Wider Environment

Before RDA, the typical stack of cataloguing standards looked like this:

28. Archives of BIBFRAME@LISTSERVE.LOC.GOV. <http://listserv.loc.gov/listarch/bibframe.html>

29. Library of Congress. Bibliographic Framework Initiative. <http://www.loc.gov/bibframe/>

30. Avram, Henriette D. The MARC II format: a communications format for bibliographic data. Washington, D.C.: Information Systems Office, Library of Congress, 1968. P. 1.



ISBD informs the AACR2 rules which are encoded in MARC. With the introduction of RDA and the replacement of MARC by Bibframe, it is often assumed that the new stack will look like this:



There are however some problems with this approach. First, linked data is greater than Bibframe.

Linked data > BIBFRAME

As discussed above, it's already in use in an increasing number of libraries, in some cases for current workflow, e.g. Oslo Public Library or the Swedish National Library. There are already other ways of cataloguing in linked data. As well as the vocabularies referred to above and the models used by the BNB, TEL, BnF, and others, even the whole RDA element set is available as RDF. Furthermore, it is being used: the TEL uses it for core bibliographic data.

A number of non-library organisations are adopting linked data, such as the BBC³¹, DBpedia (a linked data version of Wikipedia)³², the UK Government³³, and Ordnance Survey³⁴; Schema.org³⁵, used by OCLC on its

31. Bartlett, Oliver. Linked Data: Connecting together the BBC's Online Content. 19 February 2013. <http://www.bbc.co.uk/blogs/internet/posts/Linked-Data-Connecting-together-the-BBCs-Online-Content>

32. DBpedia. <http://dbpedia.org/About>

33. Data.gov.uk. Overview of linked data. <http://data.gov.uk/linked-data/overview-of-linked-data>

34. Ordnance Survey. Ordnance Survey Linked Data Platform. <http://data.ordnancesurvey.co.uk/>

35. Schema.org. What is Schema.org? <https://schema.org/>

Worldcat service³⁶, is an initiative started by search engines to improve results and embed metadata within web pages. Libraries can benefit by creating and exploiting links with these efforts and enriching our own data, allowing services to, for example, crosswalk from Wikipedia to catalogue searches via OCLC's VIAF service³⁷. Both libraries and the wider world are to some extent already occupying the same space and performing the tasks that Bibframe seeks to do. Even if successful, how long will Bibframe take to reach maturity? It has gone from a recommendation in 2011 to a more stable but still very experimental vocabulary in 2014 but it is still some way off from being settled. The practical effects of this are clearly laid out by the Oslo Public Library:

“So why don't we wait for the library standards for linked data cataloguing that are bound to come sooner or later? Well, first of all there is the suspicion that «sooner» might be slightly less likely than «later». The new Oslo Public Library is opening in only four years, and we simply don't have the time to wait and see what happens in the meantime.”³⁸

Bibframe's success depends heavily on the software it is used on, and the software culture it will exist within. It requires complete re-writing of library software if we continue to use the classic LMS. There is little in the way of clear commitment by LMS vendors to this linked data generally, and Bibframe in particular. Indeed, one concern would be an LMS that handles Bibframe but not linked data generally, whereas linked data does give us the chance to break free of library-specific software for our metadata.

Metadata > Cataloguing

Archival and repository metadata, along with museum and commercial metadata all use various standards - by no means all of them will ever use RDA - but they will all benefit from being linked to each other and to a wider web of data. Whether Bibframe suits the needs of all traditional catalogues is one thing; its adoption for all bibliographic purposes is quite another. Linked data at least offers a way of bridging between bibliographic (and non-bibliographic) efforts across the web. Ultimately, a future involving library linked cataloguing data might look more like this:

36. OCLC. OCLC Adds Linked Data to WorldCat.org. 20 June 2012. <https://oclc.org/news/releases/2012/201238.en.html>

37. Meehan, Thomas. Bookmarklet for Searching Catalogues from Wikipedia. 25 March 2013. <http://www.aurochs.org/aurlog/2013/03/25/bookmarklet-for-searching-catalogues-from-wikipedia/>

38. Rekkavi, Asgeir. RDF Linked data cataloguing at Oslo Public Library. 6 July 2014. <http://digital.deichman.no/blog/2014/07/06/rdf-linked-data-cataloguing-at-oslo-public-library/>

ISBD/FRBR/ISBRBR



AACR2/RDA/AACRDA



[BIBFRAME/schema.org/bibo/dct/foaf/owl/rdf/rdfs/madsrdf/edm/rdau/rdaw/rdab/rdam/rdaa/rdac/rdar/rdad/rdaw/rdax/rdaz](http://BIBFRAME/schema.org/bibo/dct/foaf/owl/rdf/rdfs/madsrdf/edm/rdau/rdaw/rdab/rdam/rdaa/rdac/rdar/rdad/rdaw/rdax/rdaz/rdab/rdac/rdar/rdad/rdaw/rdax/rdaz)

There is a variety not only of vocabularies and models, but of rules and initiatives feeding the data. Whatever works, whatever fits libraries' various business cases, and fulfils the varied needs of our varied users will win out. Is this likely to be one scheme under the ultimate direction of one library? Bibframe, like RDA, will certainly be very influential because of its pedigree, who is doing it, and the strong need it is trying to fulfil, but those three things could also hold it back if it is not agile enough, does not join in properly the with the rest of the web of data, and does not look to satisfying newer and broader aims beyond the traditional catalogue.