

## MARC21 to Bibframe: outcomes, possibilities and new directions

*Much of the documentation and literature about the Bibframe is focused on its impact on bibliographic description and semantic web representation. However Bibframe also has the potential to radically transform the way library and information services are discovered, accessed, used and managed.*

The cataloguing world is looking forward with much enthusiasm to the new development, known as Bibframe. Bibframe is a new bibliographic framework and a Library of Congress led initiative that will revolutionise the way library data is described, created, exchanged and used. Presently there are tremendous efforts underway to develop the semantic web aspects of Bibframe by Zepheira, a company well versed in open web standards and software. Some institutions are already experimenting and creating Bibframe based data, including the Library of Congress, Princeton University, the National Library of Medicine and the German National Library.

To envisage the future with Bibframe is a daunting task, so this article is an attempt to describe albeit briefly some of the impacts on library services and activities as we move headfast into the world of Bibframe and its associated semantic web world.

For many years we in our library world have used MARC (Machine Readable Cataloguing), AACR2 (Anglo-American Cataloguing Rules), and Library of Congress Authorities, and more recently RDA (Resource Description and Access). MARC based library catalogues have traditionally provided the maintenance activities of acquisitions, cataloguing, and circulation for both digital and non-digital materials. The OPAC (online public access catalogue) and federated search systems are at the heart for search/find and document delivery, and are integral to online library services.

It has long been acknowledged that the MARC based model is not sustainable Web wise for libraries. MARC does not fit into the Web "model" of data formats. MARC data is hidden in the 'deep Web' in information silos which cannot be found by search engines, unless heavily manipulated into Web friendly formats.

To enable a better fit for library data on the World Wide Web needs a change in the way bibliographic data is represented and a radical rethink of the technologies that support library services. This is what the Bibframe Initiative is working toward, to make a new future for libraries which is much more integrated on the Web a viable option.

At the heart of Bibframe are linked data technologies which are open standards for the Web. These comprise the Web technologies known as HTTP (Hypertext Transfer Protocol), URIs (Uniform Resource Identifiers) and RDF (Resource Description Framework). These standards fit the semantic web technology stack and are built and designed in a way that the Web understands, can interpret and enable data reuse.

Much of the literature and documentation about Bibframe currently focuses on it in terms of its impact on bibliographic description and semantic web representation. It appears to be a sort of alphabet soup. We see technical words and acronyms like: RDA, bf:identifiers, bf:scribe bf:scheme, RDF/XML, semantic web, linked data, Turtle, OWL, JSON, ontologies, SPARQL, URIs, triples and schema.org. However if this jargon is removed and Bibframe is seen for what it aims to achieve, what are the implications for libraries? How will Bibframe provide new opportunities for libraries to improve their work and create innovative new services and products? What will happen when library metadata is

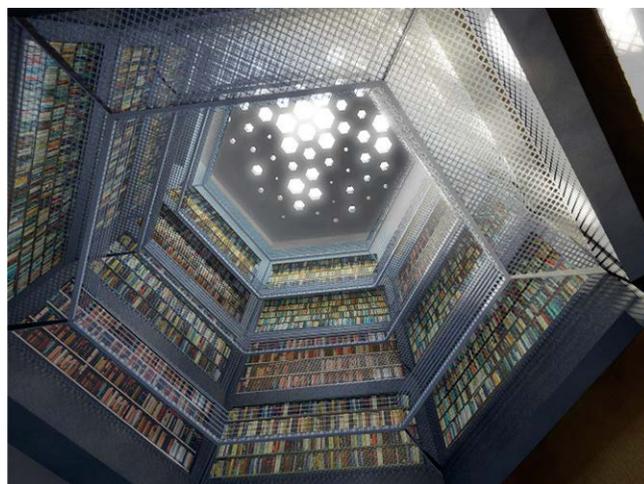
available in formats that are much more Web friendly?

Bibframe and its associated linked data developments have the potential to bring new opportunities and radically transform library and information services. With linked data technologies libraries can create, use and reuse data that is actually in a format the Web understands. Libraries which adopt Bibframe technologies, will be able to truly participate in the semantic web.

With the Bibframe developments it will be easy for libraries to create datasets that can stand alone or be blended with other datasets. In the Linked Open Data cloud (LOD) there are many different types of datasets from many different types of organisations and services. Some examples of these are datasets for GeoNames covering eight million place names, the British National Bibliography, the Library of Congress Subject Headings and Flickr. More can be found at the datahub. Because the data is available as structured data, it provides the opportunity for any of these datasets to link to any other datasets. So when libraries join the LOD their datasets can be linked to any other type of dataset and not just those that are bibliographic. Libraries typically collaborate with museum and archives, but Bibframe has the potential to bring about many new innovative and creative connections.

Bibframe will make it possible to build new discovery, and reuse type applications over linked data to create new and imaginative services. Applications, mashups, widgets and similar toolsets built for linked data are typically shareable, and so with Bibframe based data libraries will be able to fully participate in the Web to manage and access data and generate new types of library services.

Bibframe has the potential to change all library activities and services from cataloging to systems work, to discovery to document delivery and library management. Similarly it will impact library education and is likely create new library roles and occupations. The following is an attempt to describe some of the key areas in library activities on which Bibframe has potential for impact.



A new future for libraries

## Bibframe and cataloguing

Cataloguing is the topic we hear most about in the context of Bibframe. The cataloguing community is very interested in Bibframe and preparing for it because it will impact how they do their work.

Literature about Bibframe, cataloguing and metadata is easily found. Bibframe is the solution for the new cataloguing code, RDA (Resource Description and Access) which has at its core the Functional Requirements for Bibliographic Records (FRBR). FRBR is based on an entity-relationship model and so highly compatible to the linked data technologies.

The elements vocabularies and relationship designators used for cataloguing and published in RDA are already available in the RDA Registry as linked data and are therefore ready to use in the semantic web.

New tools for cataloguers to describe and create Bibframe based linked data are being scoped and trialed. Zepheira already has developed a prototype tool known as bf:Bibframe Scribe which demonstrates cataloguing using Bibframe to create linked data. From the prototype it can be seen how cataloguing will change from creating MARC records to creating metadata with links. Eric Miller calls this a change from cataloguing to catalinking.

The cataloguing community are well on the way toward a semantic web.

## Bibframe and library technologies

Library systems experts now spend time managing MARC data - manipulating it, doing quality management activities, and mining it, and all the while keeping the MARC data moving through systems in client server architecture. Technologies for OPACS are primarily built upon the MARC record.

Library systems and technology will be affected by Bibframe developments. Moving past the current popularity of Software as a Service (SAAS) and cloud computing based system, library systems will need to support common and interoperable Web standards and technologies used for linked data.

With traditional relational MARC type database published as linked data, library systems experts will add more to their current repertoire of skills and understand the building blocks of the semantic web and have an understanding of linked data concepts and graph based databases. New and massive databases can bring different solutions for data management, such as NoSQL databases (Not only SQL).

Library systems experts and those with skills in semantic web technologies will be able to develop search and discovery tools beyond the OPAC. With the data available in an open standard and in Web friendly formats software developers will be able to build tools and semantic web applications to provide new types of reuse and discovery services. These applications could utilise datasets from multiple communities including the library community. Efforts are already underway at the Tutt Library at Colorado College to build APIs (Application Programming Interface) for data in linked data formats, which include applications for access and discovery and productivity applications.

MarcEdit, an application initially developed for MARC data, already has a tool which can inherit URI identifiers into MARC data and generate semantic web formats, such as RDF/XML, N-Triples and JSON. Data in these formats in the open Web standard format can be used to build a myriad of new services such

as Google-like knowledge graphs. Data can also be repurposed into open source search platforms such as Solr, which is a Lucene based search server, as described by Götz Hatop. So by exporting MARC data from the library catalogue it will be possible to repurpose it to create many and new innovative services.

It will be a challenge to vendors as they consider the need to upscale their products significantly to support Bibframe and linked data and potentially deal with very large datasets. Library software vendors will want to develop products and technologies to provide new ways to query and display data in more interactive ways such as faceting and visualisation. How will vendor products incorporate linked data and datasets into their products, and how will vendors cope as APIs are created to query linked data?

## Reference, research and discovery

Bibframe will influence reference, research and discovery services. With library data on the Web library users will find themselves in a completely different world where library data is easy to access and more integrated into the Web. When searching the Web library data should be as visible as Wikipedia results are in Google search results. Google's developing Knowledge Graphs use semantic technologies, and could also make use of library metadata and content to provide added value and content.

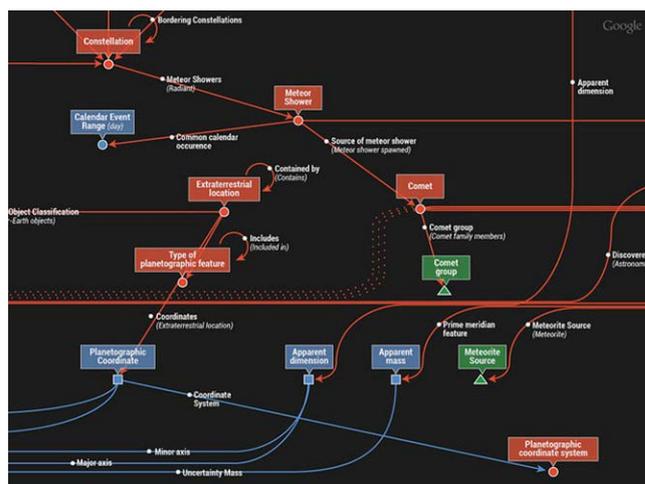


Figure 2. Concept of Google Knowledge Graphs  
(Francisco José López Villa, CC-BY-SA)

Alternate options to the traditional library OPAC will be a possibility. With library data as linked data and powered by semantic web technologies new types of Web interfaces can be developed. These Web interfaces can pull information on the fly from a variety of sources and present information and content in context. Libraries need not be limited to an interface built only for library OPACs, but could repurpose and reuse other open source API interfaces built to interact with structured linked data. The APIs need not be limited to search but instead support multiple scenarios, including those we may see in a library OPAC, plus provide context for queries, and enable mashups and reuse. Firefox already has an extension called Piggy Bank that can manipulate data to RDF, a linked data specification and render it in the Piggy Bank browser to be " browsed, searched, sorted, and organized" together, and then repurposed and shared. The data sources used by Piggy Bank can originate from multiple Web sites and datasets.

Smarter search interfaces that can access the structured data could replace the federated search interfaces we now use with search interfaces. It is possible to combine library data with other datasets which can then be queried and repurposed. Some services are now using these new smarter semantic web browsers such as Europeana. Some other semantic web browsers can be found at dbpedia. Topic and domain based services are available such as MusicBrainz providing information about artists and music and GeoSpecies providing information on biological orders, families, species, occurrence records and related data. Freebase is a huge collection of structured data covering millions of topics. Library data as linked data could be incorporated into these types of semantic web services, and then be provided in context with related information.

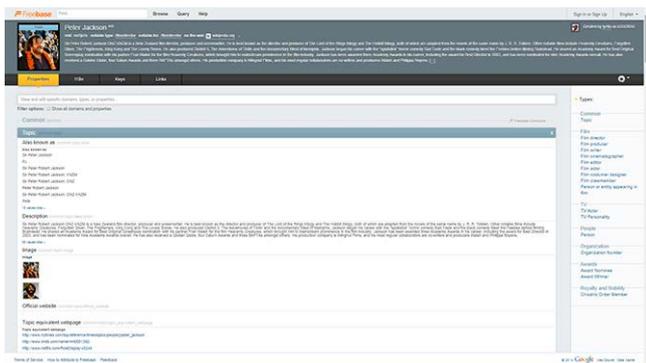


Figure 3. Freebase representation of Peter Jackson

## Bibframe and collection management activities

Much of the development work in Bibframe is concentrated on metadata management, discovery and reuse. The impact of Bibframe on collection management activities for acquisitions and circulation is one of the more difficult areas to predict. Acquisitions and circulation activities tend to have complex workflows and differ for library size and type. However with library metadata in the open standard format we will probably see new types of Web applications and toolsets built for specific niches and requirements. The Tutt Library at Colorado College has developed what they have named “productivity applications”, and include, a library hours app, an orders app, a call number app and are working toward developing a union catalogue for lending and borrowing to integrate into their library catalogue. Through the sharing and reuse of these types of applications there will be potential to create productivity gains and new opportunities for libraries.

## Strategic developments and new relationships

Because library linked data is in an open standard and can “talk” to any other type of data in any other domains there is potential for

organisations to get together and forge new strategic relationships beyond the traditional industry sector of museums, archive, galleries and libraries. Data sets in the LOD could be combined in new and innovative ways that are only limited by our imagination. Maybe there is some potential to join together the bibliographic universe into one large linked data dataset.

There is scope for libraries to combine datasets to create new virtual libraries for domain or similar library types, for example health, law, the geosector, or the business sector. These datasets could be integrated with value added information from other linked data sets.

Linked data technologies also have the potential to be the underlying technology for the New Zealand National Union Catalogue (NUC). One of the primary functions of the National Union Catalogue is as a source of New Zealand library holdings and this is possible because Bibframe as linked data supports holdings at various levels of complexity. Datasets could be incorporated such as those from the Open Library and Amazon into the National Union Catalogue. To support metadata creation on the NUC, open source cataloguing applications could be added and the vocabularies from the Library of Congress Linked Data Service included as a dataset.

## Library education

Lastly but not least the changes ahead will impact library education in New Zealand, with a strong refocus toward new system technologies. It is likely that new types of occupations will emerge within the industry, such as library developers, who will amongst other things will understand the semantic web, taxonomies, ontologies and be able to work easily in the semantic web to create open source APIs, toolsets and datasets.

## Conclusion

Looking toward the future Bibframe is continuing its semantic developments with the Library of Congress’s recent Request for Quote (RFQ) solicitation to develop software for search and display and the development of Bibframe Profiles. Linked data developments continue both architecturally within the World Wide Web Consortium (W3C), an expanding Linked Data Cloud and projects such as the Linked Data for Libraries (LD4L) to create a Scholarly Resource Semantic Information Store (SRSIS) model. There is little doubt that the library world is fast heading toward a new future that is very much part of the semantic web and consequently Web-based. Bibframe has the potential to be a disruptive innovation that will completely transform the way library content is discovered, accessed, used and managed.

**Karen Rollitt**

Applications Specialist, Department of Internal Affairs  
[karen.rollitt@dia.govt.nz](mailto:karen.rollitt@dia.govt.nz)

## References

- Bibliographic Framework Initiative. <http://www.loc.gov/bibframe/>  
 Photo courtesy of Tyler Julian Hohnson. Helsinki Library.  
<http://www.tyler-johnson.com/Helsinki-Library>.  
 Linked data : connect distributed data across  
 the Web. <http://linkeddata.org/home>  
 Linked Open Data Graph. <http://inkdroid.org/lod-graph/>

- GeoNames.the GeoNames geographical database covers all countries and contains over eight million placenames that are available for download free of charge. <http://www.geonames.org/>  
 Yahoo! Labs: science driven innovation. [yahooslabs.tumblr.com/post/89783581601/](http://yahooslabs.tumblr.com/post/89783581601/).  
 Datahub: the easiest way to get use and share data.  
<http://datahub.io/dataset/hyperlink-graph>

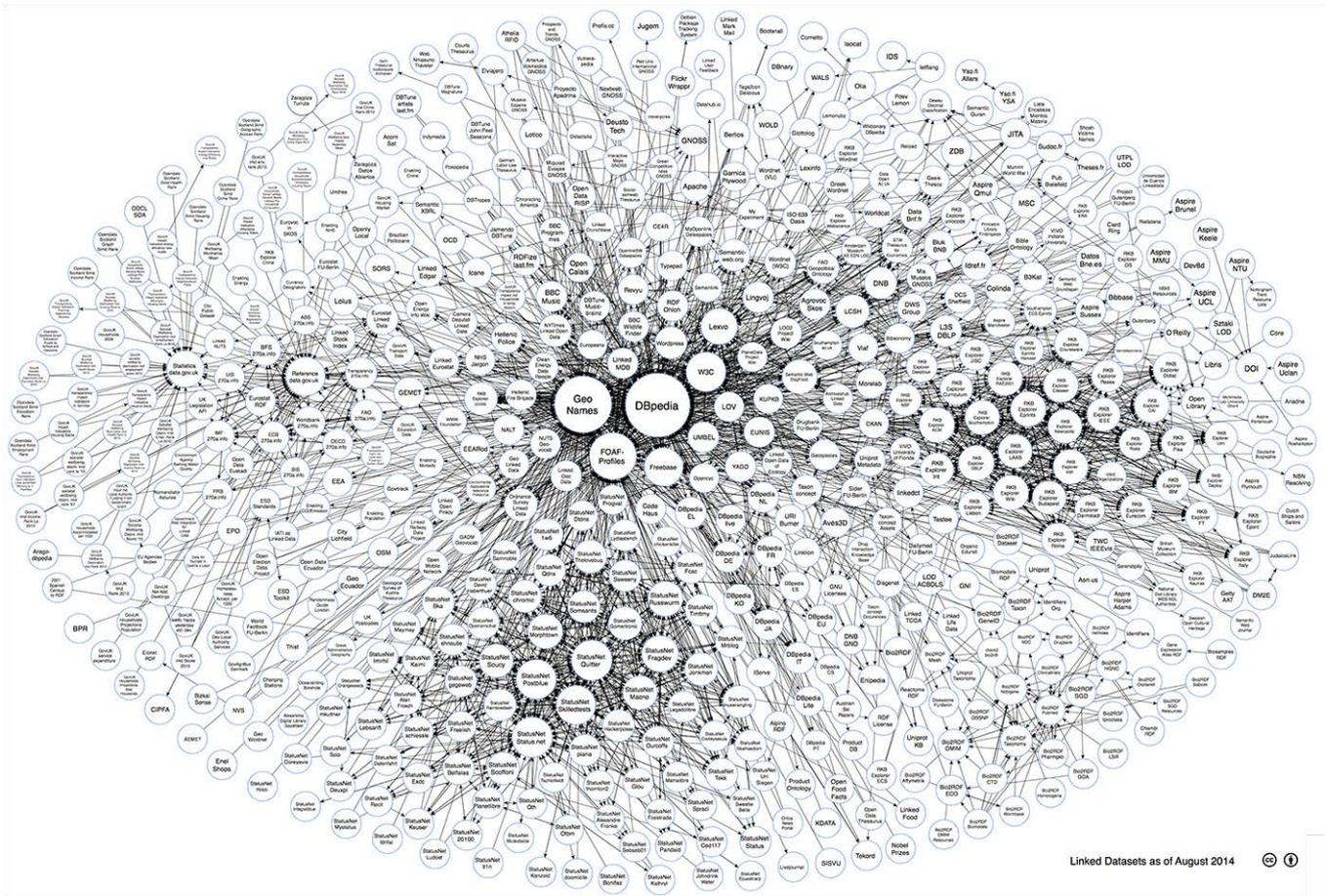


Figure 4. Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. <http://lod-cloud.net/>

<http://www.loc.gov/cds/downloads/FRBR.PDF>

<http://www.rdaregistry.info/>

<http://editor.bibframe.zepheira.com/static/>

From the Library of Congress in Washington DC. <http://www.loc.gov/today/cyberlc/transcripts/2013/130127lis1030.txt>

Linked-data-api : API and formats to simplify use of linked data by web-developers <http://code.google.com/p/linked-data-api/>

Terry's worklog: on my work (programming, digital libraries, cataloging) and other stuff that perks my interest (family, cycling, etc.). <http://blog.reset.net/archives/1368>

Hatop, Gotz. Integrating Linked Data into Discovery. Code4Lib journal, Issue 21, 2013. <http://journal.code4lib.org/issues/issues/issue21>

Piggy Bank: Piggy Bank is a Firefox extension that turns your browser into a mashup platform by allowing you to extract data from different web sites and mix them together. <http://hdl.handle.net/1721.1/29466>

Huynh, D., Mazzocchi, S. and Karger, D. Piggy Bank: experience the Semantic Web inside your web browser. Lecture Notes in Computer Science, Volume 3729, Oct 2005, Pages 413 - 430. <http://simile.mit.edu/papers/iswc05.pdf> P. 2

Europeana: think culture. <http://www.europeana.eu/>

DBpedia. <http://wiki.dbpedia.org/Interlinking>

Musicbrainz. <https://musicbrainz.org/>

GeoSpecies Knowledge Base. <http://datahub.io/dataset/geospecies>

Freebase: a community-curated database of well-known people, places, and things. <https://www.freebase.com/>

<https://wiki.duraspace.org/display/ld4l/Working+Documents>

<http://journal.code4lib.org/articles/7349>

Bibliographic Data: Library of Congress Planning to Develop Web-Based BIBFRAME Profile Creation and Search Tools. Infodocket library journal. September 22 2014. <http://www.infodocket.com/2014/09/03/bibliographic-data-library-of-congress-plans-to-develop-web-based-bibframe-profile-creation-and-search-tools/>

Duraspace: linked date for libraries. <https://wiki.duraspace.org/pages/viewpage.action?pageId=41354028>

[http://en.wikipedia.org/wiki/Disruptive\\_innovation](http://en.wikipedia.org/wiki/Disruptive_innovation)