

Crossing the bridge from cataloguing to programming – is it essential for metadata specialists?

Renate Beilharz, Box Hill Institute

Preamble

I am a cataloguer; more precisely, I am a teacher of cataloguing. Ever since my studies as teacher-librarian in the 1980s, I have found cataloguing very comforting and enjoyable. I love all the standards and rules. Anglo-American Cataloguing Rules, 2nd Edition was my cataloguing ‘home’ for many years and I have come to appreciate the functionality of the Resource Description and Access toolkit. I enjoy the logic challenges that Dewey Decimal Classification poses when trying to create that specific number to reflect a resource’s content. Subject headings and thesaurus terms bring their own linguistic delights. And then to top it all off, catalogue records can be nicely bundled into a Machine Readable Cataloguing format for the computers. This makes the records accessible to all users through the public access search screen.

In the past four years I have been pushed out of my library cataloguing ‘comfort zone’. While learning about RDA in 2012-2013, I was introduced to the world of Semantic Web, Linked Data and entity-relationship database models. In 2015, BIBFRAME, the proposed replacement for MARC, gave me even more new terminology to understand: Resource Description Framework, URIs, data triples and relational databases. The more I read about library metadata and the World Wide Web, the more I came across concepts with which I am not comfortable. I am beginning to think I need to be familiar with web technologies to effectively function as a metadata specialist in this online world of Linked Data.

Librarians and cataloguers have been creating metadata for resources since the days of Callimachus in the 3rd century BCE. Over time, rules and standards for describing resources were developed, adapted and shared. These created a healthy foundation for effective sharing and retrieval of information. Library cataloguers have a strong tradition of meeting user needs through the use of consistent and robust cataloguing tools unique to the library industry. More recently, there is an emphasis on breaking down the barriers between library bibliographic repositories (library catalogues) and the World Wide Web (WWW). Discussion has focussed on using Linked Data, with its applicable data structures and metadata languages. It is timely to consider how important it is for cataloguers to become familiar with standards and structures outside the traditional library bibliographic universe and thereby be considered metadata specialists in an online, digital world.

The creation of catalogue or bibliographic data is not an end in itself; it has a purpose: to assist users in retrieving and accessing information to meet their needs. This has been a focus of library catalogues from the earliest days of cataloguing standards. Cutter, in his renowned publication *Rules for a library catalog*, states that the first object of the catalogue is “To enable a *person* [author emphasis] to find a book ...” (Cutter, 1904, p.12). Since then statements of cataloguing principles have periodically been published by international bodies, with the most recent iteration published by International Federation of Library Associations and Institutions (IFLA) updated in 2016 called *Statement of International Cataloguing Principles* (ICP). This document again reinforces the notion that the most important principle of any cataloguing code is the “convenience of the user” (Galeffi, Bertolini, Bothmann, Rodriguez, McGarry, 2017, p. 5). The ICP lists the objectives and functions of the catalogues in terms of user needs: find, identify, select, acquire or obtain, navigate and explore (Galeffi et al., 2017, p.10-11).

While the focus on the user has been a feature of cataloguing principles, the context has changed considerably in the 141 years between Cutter’s statement and the 2016 ICP. During this period the majority of the current cataloguing tools were developed: Anglo-American Cataloguing Rules and Anglo-American Cataloguing Rules Second edition (AACR2), International Standard Bibliographic Description, Library of Congress Subject Headings (LCSH) and Dewey Decimal Classification (DDC). All of these tools were first developed for the card catalogue and have been revised, updated and adapted to ensure that they continue to meet the needs of the users. With the advent of computers, Machine Readable Cataloguing (MARC) standards were developed in the 1960s. MARC is a very effective, rigorous data exchange format for bibliographic data. It allows sharing of records between libraries and effective user discovery via the online public access catalogues.

Until the early 21st century cataloguing standards have provided a stable base for library-based metadata creators, that is, cataloguers. Even so, the cataloguing environment has never been static. Incremental developments have been made within the context of well-known data structures, all underpinned by the standards originally developed for the analogue card catalogue which had served the library world well.

The cataloguing context has changed dramatically in more recent decades, with the growth of what has been called the information society (Beniger, 1986). A plethora of physical and digital resource formats in a variety of information repositories are now being used to store the world’s data. Correspondingly, many new online information discovery systems and platforms have developed using ever more complex web technologies, including Semantic Web and Linked Data functionalities. This has resulted in the adaption and development of new processes for creation of metadata to assist with user retrieval and information access. Along with these advances, user needs, user skills and user expectations have adapted to respond to the ever expanding information available to consumers. In response to the changes in the

information environment, new cataloguing structures and codes have been developed which are transforming the cataloguing domain. The focus of bibliographic data is being directed away from the structures based on card catalogues to structures used in the online, digital space in the WWW.

The rate at which changes in cataloguing are taking place is increasing. A brief glance at Anne Welsh's visual timeline of cataloguing codes clearly demonstrates the increase in speed of change in the past decade (figure 1). Concurrently with these developments, information technology language and concepts are being used to explain and justify the changes. The scope of cataloguing terminology is growing to encompass computing vocabulary.

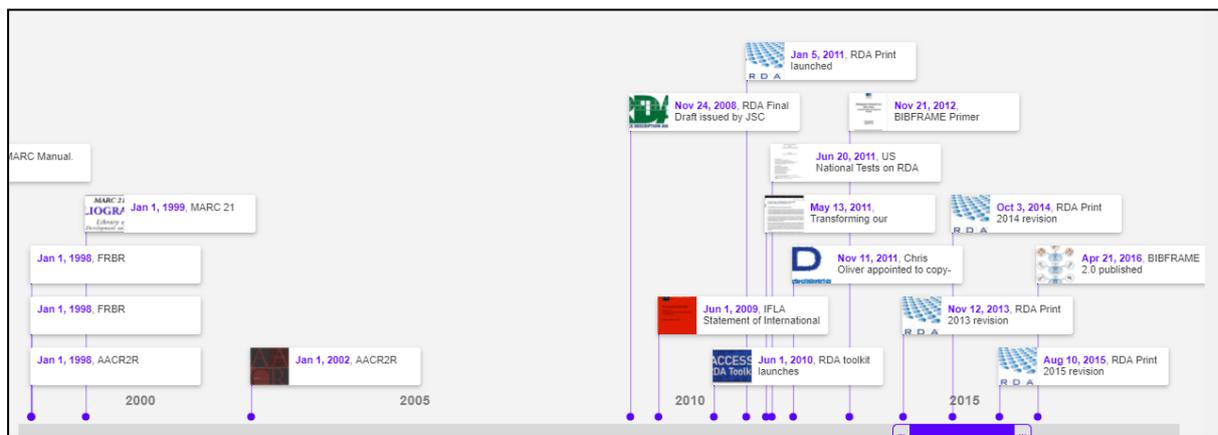


Figure 1 Anne Welsh's visual timeline of cataloguing codes <http://www.timetoast.com/timelines/9284>

The publication of Functional Requirements for Bibliographic Records (FRBR) in 1998 by IFLA, signalled the first key change in cataloguing direction and used database management concepts and language. FRBR, according Tillett, is “assisting us to review our traditions in cataloguing in light of today’s digital environment” (2003, p. 7). FRBR is a conceptual model of the bibliographic universe, based on entity-relationship models of relational database systems (IFLA, 2009b, p. 9). It heralded an expansion of the language used in cataloguing from library card-based terminology to using terminology from the database design community, such as entities, relationships and attributes (Coyle, 2015).

The introduction of relational database language in FRBR ensured that entity-relationship terminology permeated into subsequent cataloguing codes. The IFLA ICP 2009 states that any cataloguing code should take into account “the entities, attributes, and relationships as defined in conceptual models of the bibliographic universe”, referring to FRBR and its siblings FRAD and FRSAD (IFLA, 2009b, p. 2). RDA, the content standard for bibliographic data replacing AACR2, was released in 2013. RDA instruction 0.2 clearly states that the conceptual models underlying RDA are FRBR, FRAD and FRSAD (RDA toolkit, 2017).

Most recently, terminology and concepts surrounding the Semantic Web and Linked Data have been added to the conversation about the future of library bibliographic data. In 2012 the Library of Congress initiated the Bibliographic Framework Initiative (BIBFRAME) to “provide a foundation for the future of bibliographic description, both on the web, and in the broader networked worlds that is grounded in Linked Data techniques” (Library of Congress, n.d.). BIBFRAME aims to replace MARC, which has served the library community effectively for over 50 years (Library of Congress, n.d.). BIBFRAME uses the Resource Description Framework (RDF) model. This model encodes semantic relationships between items of data in the form of Linked Data triples, with the goal of enabling more effective data interchange between information repositories and the web.

As can be seen from this brief overview of recent developments in cataloguing standards, change has been frequent. This situation will not alter, with the next modification to these ‘new’ standards imminent. FRBR, and its siblings FRAD and FRSAD, have been consolidated and updated, into a new model called the IFLA Library Reference Model (LRM). The LRM is waiting for endorsement before being implemented (IFLA, 2017). This has, in turn, led to RDA being reviewed to ensure compatibility with LRM (RDA Steering Committee, 2017). The RDA Toolkit Restructure and Redesign Project (Hennelly, 2017), encapsulating LRM, is expected to be completed in 2018. BIBFRAME is now in its second iteration and is undergoing testing in pilot programs which are expected to generate further changes. Cataloguing standards are no longer static and are under constant review to meet the challenges of emerging technologies.

FRBR and BIBFRAME have heralded a conceptual shift in how bibliographic data is, or will be, represented in the online environment. Nevertheless, for many cataloguers the process of cataloguing has not changed much as a result of these models. Most cataloguers work within the constraints of their library management systems (LMS) and few LMS have commercially released FRBR or FRBR-like functionality. A scan of Marshall Breeding’s Library Technology Industry reports indicate that only one commercially available system, VTLS, is FRBRised (2017). BIBFRAME is still in a developmental stage, being trialled by select organisations (Library of Congress, 2017). While RDA is organised according to FRBR concepts, as a data format standard it can be used in current MARC systems effectively without cataloguers having to come to grips with the complexities of FRBR language and structures.

It is possible for cataloguers to use relevant content standards such as RDA, DDC and LCSH without comprehending LMS data structures and system design. Cataloguers can use data entry interfaces, without having any knowledge of the underpinning MARC coding of the system (figure 2). BIBFRAME developers have created an editor tool, which allows entry of bibliographic data without reference to any Linked Data schema or structures (figure 3). The BIBFLOW Roadmap, a report on how libraries can transition to a Linked Data environment, explicitly states: “Catalogers do not need to have in-depth knowledge of BIBFRAME’s data model or BIBFRAME vocabularies to perform cataloging because the terms used by Linked Data workbenches are the same ones currently used by catalogers” (Smith, Stahmer, Li, Gonzales, 2017, p. 33). Just because an in-depth knowledge of data models and vocabularies is not required, does not mean that this desirable situation for cataloguers aspiring to be called metadata specialists.

Figure 2 Cataloguing data entry interface of the Liberty LMS

Figure 3 BIBFRAME Editor interface <http://bibframe.org/tools/editor/#>

Cataloguers are not, and should not be, unthinking data entry operators. They have always interpreted the standards, keeping in mind user needs. According to AACR2, rule 0.9, cataloguers are required to “apply such judgement consistently within a particular context” (Joint Steering Committee for Revision of AACR, 2005). RDA uses the phrase “if considered important for identification or access” throughout the instructions (RDA Toolkit, 2017), explicitly expecting the cataloguer to make a judgement. Cataloguers analyse the content of a resource and contextualise it in relation to the rest of the information environment (Edmunds, 2016). Edmunds eloquently wrote “We [that is, cataloguers] describe. We disambiguate. We contextualize” (2017). This entails the creation and allocation of consistent and controlled subject terms and name access points, and classification of the content to assist with retrieval.

Cataloguers are more than just transcribers of data from a resource; they are responsible for contextualising it in a manner that is effective within the wider information environment in which their users reside. To do this effectively, cataloguers must understand how the data they create is manipulated and used in the discovery systems. Cataloguers must be aware of how the information is discovered by users and presented to users. If they are to be considered metadata specialists, cataloguers must be mindful of the information technology context in which they work. Cataloguers, if they wish to participate in the discussion of the future of bibliographic data and the future of discovery, must be able to effectively participate in the discussion and debate.

Debate on the future of library metadata is ongoing and vigorous. Much debate is focussing on the use of Linked Data databases to break down the barriers between information databases and the WWW. The debate has become more vigorous with the advent of BIBFRAME as the proposed MARC replacement. The debate seems to be polarised between systems and software developers and cataloguers. Coyle, in a blog post responding to an argument ‘If it ain’t broke, don’t fix it’ in relation to MARC, wrote “This is more likely to be uttered by members of the cataloging community - those who create the bibliographic data that makes up library catalogs - than by those whose jobs entail systems design and maintenance” (2017, April 12). Suominen wrote “I have a somewhat different view on this universe. I come from a Semantic Web/Linked data background ...” in response to a cataloguer’s concern over the replacement of MARC (2017). Hammer, who calls himself a “small software developer”, writes on Coyle’s blog “I feel like lots of catalogers are curious about emerging technologies and keen to see their field evolve” (2017). There is a disconnection between how systems developers and cataloguers approach the same problem, however it is not an insurmountable gap.

To be able to effectively join the debate about emerging technologies and the role of metadata, cataloguers will need to be curious about more than the purely library-based metadata standards and structures, such as RDA and MARC. This is evident across the whole library industry; textbooks, journal articles, courses, conferences, online forums and debates refer to information and web technologies not within the traditional cataloguers' knowledge-base. Any recent text or course¹ on metadata, will refer to Semantic Web, Linked Data and RDF, and various schema and syntaxes used to organise data on the web such as Dublin Core, MODS and MARC XML. Articles and lectures on library Linked Data projects introduce the reader to serialization models such as RDF/XML, RDFa, JSON-LD and Turtle and to web standards such as OWL, SKOS, FOAF and Schema.org (Godby, Wang & Mixter, 2015, p. 9, Mitchell, 2016, p. 22). A basic understanding of languages such as HTML, CSS and JavaScript is also assumed when discussing how a LMS manipulates bibliographic data (B. Birchup, personal communication, July 12, 2017). A cataloguer's curiosity about emerging technologies can easily overwhelmed when confronted with all of this new terminology.

The online information space is complex, and web technologies used across the sector are convoluted. The vast nature of metadata standards that exist in the cultural heritage sector is portrayed in a visualisation of the metadata universe by Riley (figure 4). Library Congress Linked Data service provides 11 different Linked Data formats to present the data for one name authority file (figure 5). Each one of these links displays different marked-up versions of the authority file record, interpretation thereof requiring a sound understanding of mark-up syntaxes (figure 6 & 7). WorldCat uses Schema.org, another mark-up language, to structure its bibliographic Linked Data (figure 8), adding to the rich and complex tapestry that is the the Semantic Web.

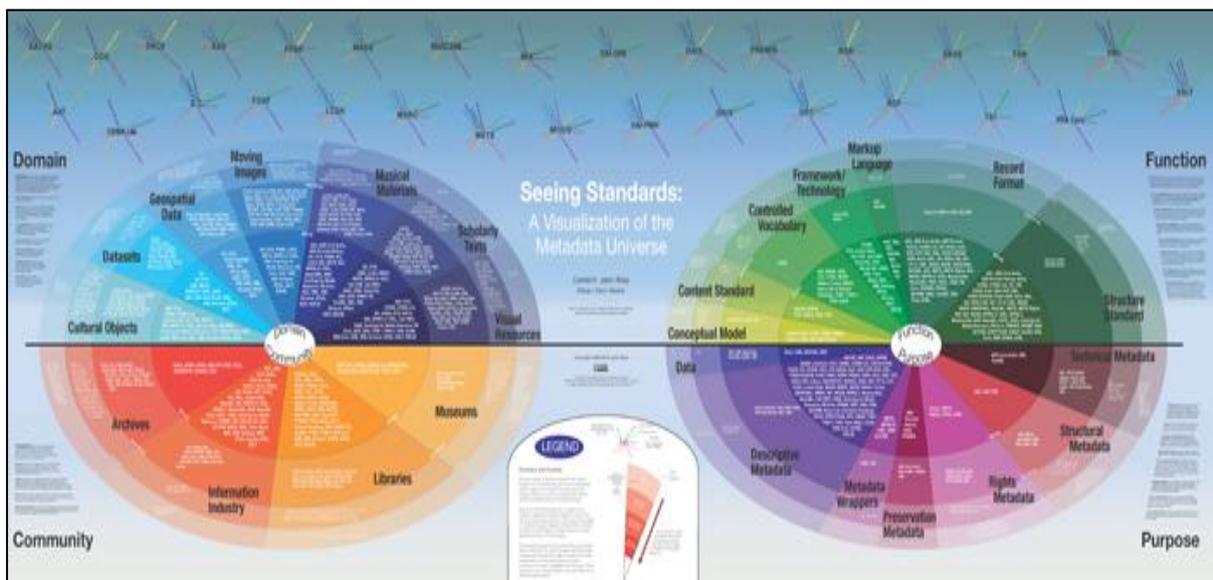


Figure 4 Seeing standards <http://jennriley.com/metadatamap/>

¹ E.g. Pomerantz, J. (2015). Metadata. Cambridge, Massachusetts : The MIT Press. and Zavalina, O. (2016) Introduction to Linked Data and Metadata eCourse. Chicago : American Library Association.

Building structures for the management, storage, manipulation and discovery of content involves the use of many different knowledge sets and skills. Consideration must be given to how much the library-based metadata specialist needs to understand. An analogy from the housing industry can be used here. When a house is constructed no single individual has all the specialised skills to undertake the tasks necessary to create a house from scratch, but they come together to create a structure to meet the user's needs. To build a house you need suppliers of the materials, architects to look at the overall structure and design and builders who put it together to make it work. Suppliers, builders and architects have their specialised skills, but each is aware of how their expertise fits into the bigger picture, all in the interested of meeting the needs of the user, the future resident of the building.

Similarly, building effective online discovery systems for users requires different sets of expertise while, at the same time, understanding the big picture. There are the suppliers of the metadata, cataloguers, who use consistent standards and mark-up languages to prepare the content for the system and, if relevant, the WWW. Then there are the architects of the system, who consider the functional requirements of systems. Information architects and systems analysts look at how the data can be structured for effective retrieval, creating the data flow diagrams, entity-relationship charts and mark-up languages for the data. Thirdly, there are the builders of the systems themselves, the developers of the programmes. As in the building industry, cataloguers, system architects and programmers all have interwoven roles to play in developing effective discovery systems for users. They must understand each other's roles to be able to collaborate and work together effectively to create online discovery systems to meet user needs.

Many library-based metadata specialists already operate effectively in this online information space, with skill sets and knowledge that allow them work collaboratively across the information technologies spectrum. At the VALA Techcamp (Melbourne, 12-14 July 2017), many presenters' biographies indicated that they are cataloguers who are skilled at data manipulation. They have knowledge of data structures, mark-up languages and programming concepts. Many camp participants have a cataloguing background and at the same time have the skills to extract and manipulate data at the mark-up level. They are cross-walking data to and from different repositories and using data scraping techniques to extract data from one program to another. These cataloguers are creating application programming interfaces to extract and mash data from various repositories for different purposes. These metadata specialists have skill sets that enable them to meet the specified needs of their users.

Cataloguers no longer own the information ecosystem (Riley, 2015, slide 12), so they must develop a broader understanding to effectively join in the debate about the future. “The change from MARC to ?? will come and it will be forced upon us through technology and economics. We can jump to a new technology blindly, in a panic, or we can plan ahead” wrote Coyle (2017, April 6). Edmunds, a cataloguer decrying the move to Linked Data, put a different spin on the future: “It would be a shame ... if our extinction were self-induced, if we vanished ... like lemmings rushing mindlessly into a Linked Data sea” (2017). Concluding a presentation on future cataloguers, Riley listed ‘good technical instincts’ as one of the required skills, along with creativity, judgment and ability to see patterns (2015, slide 15). It is clear, whether cataloguers agree with the directions for bibliographic data or not, they cannot be bystanders in the discussion and debates about their future.

Cataloguers must be able to participate actively in the debate on the future of bibliographic data. They must become metadata specialists in the broad sense of the word. They need enough knowledge to understand systems designs, the web and programming languages. This will enable them to debate the continuing changing and developing options intelligently and logically. How much in-depth knowledge of different aspects of system design, web technologies and mark-up languages is required, depends on to what level the individual wishes to actively participate in shaping the future of information discovery. Nevertheless, it is essential for cataloguers to at least become familiar with standards and structures outside the traditional library bibliographic universe and so be considered metadata specialists in an online, digital world.

Postscript

Undertaking research for this paper, has convinced me that I need to learn more about web and information technologies. I don’t need to become a programmer or have expert skills in marking-up Linked Data statements. But I do need to understand how it all works together, so see the big picture, so I can participate in the debate and follow the conversations taking place in the metadata ecosystem. I am now looking for professional development opportunities to fill the gaps in my knowledge to be able to ensure that I do not “jump blindly into new technologies” like “a lemming”. I aim to feel confident in my role as a teacher of future cataloguing-based metadata specialists.

References

- Beniger, J.R. (1986). *The Control Revolution: Technological and Economic Origins of the Information Society*. Cambridge, Mass. : Harvard University Press.
- Breeding, M. (2017) Library technology guide : document repository. Retrieved 2017, July 15 from <https://librarytechnology.org/industryreports/>
- Coyle, K. (2015, December 17). *Mistakes have been made* [Video file]. Retrieved from <https://www.youtube.com/watch?v=d0CMuxZsAIY&feature=youtu.be&list=PL7fMsenbLiQ0eKJtpz3NCv0937HPwbWqV>
- Coyle, K. (2017, April 6). Precipitating forward : our legacy, our mistake [Web log post]. Retrieved from <http://kcoyle.blogspot.com.au/2017/04/>
- Coyle, K. (2017, April 12). If it ain't broke [Web log post]. Retrieved from <http://kcoyle.blogspot.com.au/2017/04/if-it-aint-broke.html>
- Cutter, C.A. (1904). Rules for a Dictionary Catalog. Retrieved from <https://digital.library.unt.edu/ark:/67531/metadc1048/>
- Edmunds, J. (2016, March 16). *Mommy, where does metadata come from?* [Video file]. Retrieved from <https://www.youtube.com/watch?v=ue-TUOtQgBOo>
- Edmunds, J. (2017). Roadmap to nowhere: BIBFLOW, BIBFRAME and Linked Data for Libraries [PDF]. Retrieved 20 May, 2017 from <https://scholarsphere.psu.edu/downloads/fj67314220>
- Galeffi, A., Bertolini, M.V., Bothmann, R.L., Rodriguez, E.E., & McGarry, D. (2017). Statement of International cataloguing principles (ICP) [PDF]. Retrieved 1 July, 2017, from https://www.ifla.org/files/assets/cataloguing/icp/icp_2016-en.pdf
- Godby, C.J., Wang, S., & Mixter, J.K. (2015). *Library Linked Data in the cloud*. Williston, VT : Morgan & Claypool Publishers.
- Hammer, S. (2017, April 4). Re: If it ain't broke [Web log comment]. Retrieved from <http://kcoyle.blogspot.com.au/2017/04/if-it-aint-broke.html>
- Hennelly, J. (2017). 3R project status report #1. Retrieved 5 July, 2017 from <http://www.rdatoolkit.org/3Rproject/SR1>
- IFLA (2009a), Functional requirements for bibliographic records [PDF]. Retrieved 1 July, 2017 from https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf
- IFLA (2009b), Statement of international cataloguing principles [PDF]. Retrieved 1 July, 2017 from https://www.ifla.org/files/assets/cataloguing/icp/icp_2009-en.pdf

- IFLA (2017), IFLA LRM. Retrieved 6 July, 2017 from <https://www.ifla.org/publications/node/11412>
- Joint Steering Committee for Revision of AACR (2005). *Anglo-American Cataloguing Rules, Second edition 2002 revision*. Retrieved 16 July, 2017 from <http://access.rdatoolkit.org>
- Library of Congress (n.d.). Bibliographic framework initiative. Retrieved 6 July, 2017 from <https://www.loc.gov/bibframe/>
- Library of Congress (2017). BIBFRAME 2.0 implementation register. Retrieved 21 July, 2017 from <https://www.loc.gov/bibframe/implementation/register.html>
- Mitchell, E.T. (2016). *Library Linked Data : early activity & development*. Chicago : ALA TechSource.
- RDA Steering Committee (2017). Implementation of LRM in RDA. Retrieved 4 July, 2017 from <http://www.rda-rsc.org/ImplementationLRMinRDA>
- RDA toolkit (2017). Retrieved 5 July, 2017 from <http://access.rdatoolkit.org>
- Riley, J. (2015). *The future of cataloguing? Future cataloguers!* [Slide show]. Retrieved from <https://www.slideshare.net/jenrile/the-future-of-cataloguing-future-cataloguers>
- Smith, M., Stahmer, C.G., Li, X., Gonzales, G. (2017). BIBFLOW : a roadmap for library Linked Data transition. Retrieved 2 July, 2017 from https://bibflow.library.ucdavis.edu/wp-content/uploads/2017/03/bibflow_roadmap_revised_3_14_2017.pdf
- Suominen, O. (2017, March 31). Life after MARC: the future of discovery [Online discussion list]. Message posted to BIBFRAME@LISTSERV.LOC.GOV
- Tillett, B. (2003). What is FRBR? : a conceptual model for the bibliographic universe [PDF]. Retrieved 1 July, 2017 from <https://www.loc.gov/cds/downloads/FRBR.PDF>