

Cat Among the Cataloguers

Susan Warren

City Libraries, City of Gold Coast, Gold Coast, Queensland, Australia.
swarren@goldcoast.qld.gov.au

Marian Morgan Bindon

City Libraries, City of Gold Coast, Gold Coast, Queensland, Australia.
mmorganbindon@goldcoast.qld.gov.au



Copyright © 2017 by Marian Morgan Bindon and Susan Warren. This work is made available under the terms of the Creative Commons Attribution 4.0 International License:

<http://creativecommons.org/licenses/by/4.0>

Abstract:

The clearest message received when surveying both non-users and users alike was: “Remind us that you are there, and then we will use you more!” (Houen, 2011).

Gone are the days when public library collections comprised of only physical print materials, and when our customers used the library in person, yet our catalogues and their underlying bibliographic data have remained fundamentally the same.

It’s no longer viable for libraries to simply maintain information resources for customers to access. We have to make the information visible to seekers where they begin their search, on the web.

City Libraries is the second largest public library service in Australia. We have over 1 million MARC and Dublin Core records waiting to be further discovered. While several public libraries in the United States have successfully implemented Linked Data initiatives in Australia, public libraries, and libraries in general, are cautiously watching international developments.

This paper explores the defining of City Libraries’ road map for the journey towards transitioning from the old document-centric bibliographic model to a more data-centric model - Linked Data. Key short-term impact trends driving this journey are increasing the value of the customer experience and ensuring City Libraries are more relevant and accessible in the eyes of our community.

Our journey begins in a moment of uncomfortable turmoil, knowing we need to forge a new path, unsure where the path might lead, or even what it looks like. This paper will discuss the barriers implementing Linked Data in our environment, issues relating to data standards, management and delivery, and how to achieve a structured implementation to get the most value from the possibilities and potential of Linked Data.

Keywords: Linked Data, Public Libraries

The Gold Coast is a metropolitan region south of Brisbane on Australia's east coast. Situated in South East Queensland, the Gold Coast stretches along 57 kilometres of coastline and is currently home to over half a million people. Beyond the city's beautiful beaches are vast, sub-tropical rainforests with breathtaking scenery. The Gold Coast is Australia's premier tourist destination, welcoming close to 12 million visitors each year. Gold Coast is the sixth-largest city in Australia, making it the largest non-capital city.

The city's public library service is Australia's second largest, comprising of a network of 12 branch libraries, one mobile library, a local studies library, and a corporate library. Gold Coasters love their libraries. Our libraries are a highly popular, visible and valued by the community. We have 253,000 registered borrowers and approximately 3 million physical and 2,472,561 virtual library visitors every year. City of Gold Coast Libraries deliver on Council's aspirations to provide a leading edge library service which meets the changing informational, informal learning, recreational, and cultural needs of our resilient and socially inclusive community.

The city's usage and uptake of online technologies is above the Australian national average with 83% of Gold Coast homes having an Internet connection (Australian Bureau of Statistics, 2017). In addition, City Library customers have the highest download rates of digital magazines in Australia through a multi-platform distribution subscription service. We recognise the future of libraries lies with our customers online and meeting these expectations has led us to linked data.

Linked data motivation

The purpose and perceived benefits of linked data have been widely discussed in academic literature. There is an abundance of information and discussion on linked data frameworks, schemas, and mapping data. However, most libraries need to articulate a specific local business case, the 'why' for implementation before their organisation is prepared to invest in new technologies. It is important for every library to understand its own unique context and motivations for moving to linked data.

The purpose of linked data is facilitating the re-usability, cross-linking, integration and sharing of data (Berners-Lee, 2009, 2010; W3C, 2011). Published literature and general discussion on the motivation for public libraries to move to linked data points to the perceived benefits to community, and increased exposure to resources and content as the major drivers. Our city's vision is to be an active digital city, and City Libraries' business objectives are to reach, engage and connect with new communities and to have discoverable and accessible collections. This vision and strategy underpins our decision to investigate linked open data with the primary goal to expose our records through Web searches, the first place our customers turn to when searching for information.

While 44% of Gold Coast residents are library members, like all public libraries, City Libraries aims to attract more customers. The issue lies in how to appeal to elusive non-library users. In a three-year study, Houen identifies that some of the assumptions around non-library users have been confirmed: “the internet is a strong rival, that non-users would rather buy books than borrow them (several of the non-users were super keen readers) and many have their needs met through other channels” (Houen, 2011). The clearest message received when surveying both non-users and users alike was: “Remind us that you are there, and then we will use you more!” (Houen, 2011). In order to remind users and non-users that libraries exist, our data must be accessible where they are, in rival territory, the Web. The Web represents the biggest alternative—and some say threat to libraries—while also offering significant opportunities.

The Web provides an opportunity for libraries in the form of linked data, enabling serendipitous results in information searching, further exposing libraries to both users and non-users. Serendipity in this context can be defined as providing “...services that will connect people with online ephemera they would not normally find on their own” (Krotoski, 2011). Serendipity has a long tradition in the history of science as having played a key role in many significant discoveries; this includes taking home 20 books from your library visit when you only intended to borrow one. Serendipity is also an integral part of the web browsing experience. Linked data allows for the possibility of serendipity in information searching, finding information one didn't even know they were looking for, just like browsing the library shelves. In this way, searches performed via search engines such as Google, could allow users to “bump” into the information they need from their local public library (Krotoski, 2011).

Another incentive for City Libraries to enter the linked data environment is to enrich professional practice, and participate in the evolving conversation surrounding information retrieval. All libraries need to have a strong voice in the evolution of linked data, in order to enhance the ways in which users access the information they require. It is common practice for library software vendors to direct the features and functions of the tools libraries use to interact with their customers. This places software development in the hands of system vendors, with the vendor controlling the data structure within the system. Generally, libraries are reliant on vendor's enhancement roadmaps and implementation timeframes to enable advancement in any technology based offerings. Academic, special and public libraries all have very different and unique motivations, needs and strategies related to the technologies employed to support respective operations—yet a relatively small niche market of library vendors serve us all. Libraries have an opportunity to take the reins in the inevitable transition to linked data and to tell library vendors what we require. In addition, freeing library data from insulated library management systems may provide an opportunity for libraries to look beyond traditional library vendors when considering technology partners. The opportunity to work with

innovative web specialists from non-library environments could potentially allow libraries to offer dynamic and unique web based services.

A public library is a library for the public. As catalogue data is an important discovery resource for the library's end-user services, the public library should keep in mind that it is the public, the customers; the data should be made to facilitate resource discovery. Cataloguers have produced catalogue data as a tool for librarians so that they, in turn, could use it to assist their customers. This is not the case anymore. Customers search the online catalogue independently and are more involved in self-discovery. Nevertheless, cataloguing rules, classification schemes and metadata formats still tend to be 'librarian readable' and remain a mystery to most customers. Linked data simply focuses on making the data rich and expressive, so that it can be applied and combined in as many ways as possible through 'speaking' in a way our customers and the Web understands which is very appealing to City Libraries.

Making linked data happen

A common sentiment, in literature about linked data in libraries, is that libraries must be not only on the Web, but of the Web. They must change from the old document-centric model to a more data-centric model, and linked data is one way to accomplish that. Librarians are in an excellent position to take the data already contained in millions of catalogue records and set it free to live on the Web. Linked data is a logical evolutionary step for the established principle of freedom of access to information.

Moving away from cataloguing in MARC and transitioning to a linked data ecosystem in any given library service, is not without obstacles and some of the hurdles which need to be considered include:

- modifying ILS and discovery services to optimise indexed bibliographic data based on linked data orientated structures
- ensuring library suppliers are able to provide bibliographic records based on linked data
- upskilling library staff, providing cataloguing and interpreting linked data skills
- convincing your organisation to pay for linked data and
- re-educating library users and non-users that they can now search like Google on the library catalogue

Each one of these requirements is a huge undertaking and considering the low risk profile associated with many library environments, it is understandable that linked data is not yet widely implemented in public libraries.

Current library system limitations and the complexity of the technology are a major concern, as well as the intimidating thought of transitioning away from MARC, a standard that is as old as many librarians working in the industry.

There is also the challenge and significant resources, time and effort required to prepare data to be fit for purpose, as well as converting masses of legacy data from MARC format to a new standard. For libraries like ours who utilise shelf-ready services, the outsourcing of cataloguing and end processing, there is significant financial investment in acquiring non-standard catalogue records for new resources.

Even without considering the challenges already mentioned– the fact remains that transitioning away from MARC toward any new bibliographic infrastructure system signifies an immense adjustment. As with any major shift in our profession, resistance to change plays a significant role. Publishing our records on the open web demands different technologies, work practices, skills sets and expertise, and represents a very major shift for libraries.

Building our knowledge and understanding of linked data concepts and tracking developments is difficult. There exists very few detailed implementation case studies published or statistics supporting the benefits, or otherwise, from libraries that have implemented linked data. The scarcity of implementation documentation and tools is cited in a recent survey by Karen Smith-Yoshimura, *Analysis of International Linked Data Survey for Implementers* (2016). The survey provides an overview of linked data implementations projects, and among other things difficulties faced publishing linked data cited by respondents. In the order of the most cited are:

1. Steep learning curve for staff
2. Inconsistency of legacy data
3. Selecting appropriate ontologies to represent our data
4. Establishing the links
5. Little documentation or advice on how to build the systems
6. Lack of tools
7. Immature software

While there are teething problems, the use of linked data is rapidly moving from the experimental, prototype phase into operational technology and, in time, these issues will be addressed as standards are formalised and tools established. It will take many years for the industry and Library Management Systems to migrate fully to the new standards and for now many libraries will have to continue with their existing workflow in obtaining and loading MARC records in their current library systems.

Where to from here

Where does this leave public libraries interested in the tools and technologies that facilitate improved user experience based on linked data?

Like many libraries we do not want to wait years to be discovered on the Web. And while we might have the financial capacity and motivation to make linked data happen, we are not in a position to develop, update and administer our own technical infrastructure to support the linked data ecosystem.

There are options for libraries to dip their toes in the linked data waters—after all best practices aren't always about technological capacity but about the experience libraries are able to provide to their customers.

A number of information discovery service vendors are working with Bibliographic Framework (or BIBFRAME) to harvest and convert MARC records to linked data using the BIBFRAME vocabulary. Records are enhanced with library location data making them geographically relevant to customers and the resulting data is linked across the libraries' local link graph and external Web sources. Once published to the Web, via the library.link network, library resources are discoverable through Web search engines or third party applications. Customers using a search engine would happen upon library data and be then taken to the local OPAC or discovery layer, at the record level.

Linked data discovery services offer levels of participation from one-time full database MARC record extraction through to subscription services with regular update harvests of MARC records from the Library Management Systems. The BIBFRAME data works in parallel with existing discovery layers and OPACs similar to other popular value added back-of-house applications.

In addition to the primary linked data service, some vendors offer data enrichment whereby data is supplemented with read-alike recommendations, appeal factors, and other reader information to provide additional access points and additional linked data formats. There isn't a single format that takes care of all linked data needs: Bibframe was developed specifically for the library community, schema.org is used by most search engines, and Facebook Open Graph is used for social data. All of these webs of data can connect to one another, so the more compatible data is with different formats, the more connected it will be. This effectively creates more access points, more relationships, more links, and more data for algorithms to understand and respond to.

Additional enrichment offerings also include reader's advisory service where additional metadata such as subject headings, genres, tone, pace, writing style, and appeal terms are added to BIBFRAME records, generating 'readalikes' and other information to help readers explore a library's collection. One vendor cites an enrichment layer will add 2-10x more data access points to your linked data.

Although initially introduced as a replacement for MARC, in our view BIBFRAME is far from being an either-or proposition given the MARC legacy. These solutions allow libraries BIBFRAME to coexist within the current cataloguing landscape as a means for sharing bibliographic data over the Web.

City Libraries have commenced implementing a subscription based Linked Data solution and an add-on enrichment layer. While this solution does not tick all the boxes on our wish list, it is a useful and easily implemented medium-term solution.

City Libraries will continue with existing workflow in obtaining and loading MARC records in our current environment. Vendor assisted linked data services provide us with a low risk, affordable option that works with existing data, requires minimal staff time or technical expertise.

The opportunities around linked data are exciting, and will be an integral part of the information landscape in the years to come.

References

Australian Bureau of Statistics. (2017) *GC households have internet connection 2016*. Retrieved from <http://stat.abs.gov.au/itt/r.jsp?databyregion>

Berners-Lee, T. (2009), "Linked Data", Retrieved from www.w3.org/DesignIssues/LinkedData.html

Google Search Statistics. *Internet live statistics*. Retrieved 3 April 2017 from <http://www.internetlivestats.com/google-search-statistics/>

Houen, R. (2011) In pursuit of non-users. *Scandinavian Library Quarterly*, 44(2). Retrieved from <http://slq.nu/?article=norway-in-pursuit-of-non-users>

Krotoski, A. (2011). Digital serendipity: be careful what you don't wish for. Retrieved from <https://www.theguardian.com/technology/2011/aug/21/google-serendipity-profiling-aleks-krotoski>

Mitchell, J. (2012) How Google Search Really Works. Retrieved from https://readwrite.com/2012/02/29/interview_changing_engines_mid-flight_qa_with_goog/

Smith-Yoshimura, K. (2016). Analysis of International Linked Data Survey for Implementers. *D-Lib Magazine* 22(7/8). doi:10.1045/july2016-smith-yoshimura

Sullivan, D. (2016) Google now handles at least 2 trillion searches per year. Retrieved from <http://searchengineland.com/google-now-handles-2-999-trillion-searches-per-year-250247>

W3C (2011). Library Linked Data Incubator Group Final Report, W3C, Cambridge, MA.