November/December 2010 Volume 16, Number 11/12

## The Strongest Link: Libraries and Linked Data

Gillian Byrne and Lisa Goddard Memorial University Libraries, St. John's, Newfoundland and Labrador {gbyrne,lgoddard}@mun.ca

doi:10.1045/november2010-byrne

(This Opinion piece presents the opinions of the authors. It does not necessarily reflect the views of D-Lib Magazine, its publisher, the Corporation for National Research Initiatives, or the D-Lib Alliance.)

### Abstract

Since 1999 the W3C has been working on a set of Semantic Web standards that have the potential to revolutionize web search. Also known as Linked Data, the Machine-Readable Web, the Web of Data, or Web 3.0, the Semantic Web relies on highly structured metadata that allow computers to understand the relationships between objects. Semantic web standards are complex, and difficult to conceptualize, but they offer solutions to many of the issues that plague libraries, including precise web search, authority control, classification, data portability, and disambiguation. This article will outline some of the benefits that linked data could have for libraries, will discuss some of the non-technical obstacles that we face in moving forward, and will finally offer suggestions for practical ways in which libraries can participate in the development of the semantic web.

## Introduction

For many years now we have been hearing that the semantic web is just around the corner. In 2008 Tim Berners-Lee declared the semantic web "open for business" (Paul Miller, 2008). The reality for most libraries, however, is that we are still grappling with 2.0 technologies. Few among us have yet embraced web 3.0, also known as the web of linked data, or the semantic web. The promise of the semantic web is a dazzling one. By marking up information in standardized, highly structured formats like Resource Description Framework (RDF), we can allow computers to better "understand" the meaning of content, rather than simply matching on strings of text. This would allow web search engines to function more like relational databases, providing much more accurate search results - the ability to distinguish between a book that is written about a person, as opposed to a book that is written by a person, for example. For most librarians this concept is fairly easy to understand. We have been creating highly structured machine-readable metadata for many years, after all, and we already understand the benefits.

The second part of the linked data vision is where things really begin to get heady. By linking our data to

shared ontologies that describe the properties and relationships of objects, we begin to allow computers not just to "understand" content, but also to derive new knowledge by "reasoning" about that content. As a simple example: Shakespeare wrote Macbeth. "Wrote" is the inverse of "WrittenBy" therefore Macbeth was written by Shakespeare. The real power of the semantic web lies in this ability for "intelligent" search engines to disambiguate terms (Apple the computer vs. apple the fruit, for example), to understand the relationships between different entities, and to bring that information together in new ways to answer queries. E.g., Show me all of the articles that have been written by people who have ever worked at any of the same institutions as Lisa Goddard.

As Linked Data initiatives proliferate there has, unsurprisingly, been increased debate about exactly what we mean when we refer to Linked Data and the Semantic Web. Are the phrases interchangeable? Do they refer to a specific set of standards? A specific technology stack? For the purposes of this paper we use the term "Semantic Web" to refer to a full suite of W3C standards including RDF, SPARQL query language, and OWL web ontology language. (W3C, 2010) As for "Linked Data" we will accept the two part definition offered by the research team at Freie Universitat Berlin, "The Web of Data is built upon two simple ideas: First, to employ the RDF data model to publish structured data on the Web. Second, to [use http URIs] to set explicit RDF links between data items within different data sources." (Isele et al., 2009) We can see from this definition that Linked Data has two distinct aspects: exposing data as RDF, and linking RDF entities together.

This article will outline some of the benefits that linked data could have for libraries, will discuss some of the non-technical obstacles that we face in moving forward, and will finally offer suggestions for practical ways in which libraries can participate in the development of the semantic web.

# What benefits will libraries derive from linked data?

With their expertise in search, metadata generation, and ontology development, librarians should actually be well positioned to understand and implement linked data. Libraries suffer from most of the problems of interoperability and information management that other organizations have, but we additionally have an explicit mandate to organize information derived from many other sources so as to make it broadly accessible.

One of the library buzz phrases of recent years has been "evidence-based decision-making", the idea that business decisions should be based at least in part on quantitative data about resource use and demand. Practically speaking, libraries have not done a great job of this because usage data is held in many different formats across different systems. Linked data would allow us to publish weblogs, COUNTER statistics, and survey data in a common format (RDF) that can be easily aggregated and queried. This data could additionally be cross-referenced with related data from outside agencies, like government demographic data, or data on grades and student success from university systems.

Having a common format for all data would be a huge boon for interoperability and the integration of all kinds of systems. One can easily imagine libraries working with their vendors to collaboratively develop a large shared knowledgebase that could act as a library "linking hub". The linking hub would expose a network of tightly linked information from publishers, aggregators, book and journal vendors, subject authorities, name authorities, and other libraries. These large stores of linked data could provide a true "integrated library system" handling all of the functions of selection, ordering, cataloguing, authority control, taxonomy development, and search. True integrated systems could replace our current mishmash of proxy servers, resolvers, search applications, Integrated Library Systems (ILSs), Electronic Resource Management Systems (ERMSs), and data crosswalks, while reducing the overhead of local records maintenance and vastly increasing the amount and quality of information that libraries have at their disposal.

Linked data could also help us uncover that holy grail of library technologies – smart federated search. If all of our electronic resource providers exposed their data in RDF and used linking hubs to apply a common controlled

vocabulary, then we could eliminate many of the problems currently associated with federated search engines like lack of granularity, inability to support sophisticated queries, poor relevancy ranking, inaccurate deduping, and slowness. For years we have been talking about having a "Google-like" search for library resources, well, semantic search could take us far beyond the current string-matching capabilities of search engines like Google.

In fact, Google and a handful of other major players are already working very hard to improve their own services by leveraging natural language processing, semantics, and structured data. (Schonfeld, 2010) Structured data will allow search engines to generate well-defined facets for disambiguation of results, and increase search precision while reducing noise and "choice overload" (Spivak, 2009). Semantic search engines also hope to take advantage of social data that is currently held in walled systems like Facebook to create expert recommender and referral systems that can help you to choose a restaurant, or a university course, or even a mate based on the preferences and connections of your friends and peers. Semantics will also help with real time search of data like Twitter streams where there are no inbound links or html tags that can to help determine relevancy. If the semantic search vision holds true then librarians can look forward to vast improvements in web search that will allow search engines to provide sophisticated query capabilities similar to those provided by conventional relational databases.

# What are the major obstacles for libraries?

Linked data tools, projects and resources for libraries have proliferated over the last couple of years, from simple MARC-RDF converters to entire semantic platforms. Although the tools will undoubtedly benefit from further refinement, technology is no longer the major obstacle to linked data implementation. A number of libraries have already spawned successful semantic web projects, and software aimed at non-expert users is increasingly available. In the <u>Appendix</u> you will find a small selection of tools and platforms that libraries may be interested in; due to space constraints this should not be considered an exhaustive list, especially in a field growing so rapidly.

A fundamental challenge for the development of linked data in libraries is lack of awareness. While there are pockets of libraries, such as digital libraries, doing complex work in the field, the concepts of linked data and the semantic web have yet to hit the consciousness of many librarians. A look at library conference proceedings and the major article databases in the field show a predominance of web 2.0 topics, with relatively little published on linked data, particularly in North America. This is beginning to change; in 2008 there were two special issues of major library journals dedicated to semantic web topics, *Library Review* (Special Issue: Digital libraries and the semantic web: context, applications and research) and *Cataloging & Classification Quarterly* (Knitting the Semantic Web). Linked data is beginning to appear in more mainstream library literature as well, as evidenced by recent columns promoting linked data by Dan Chudnov and Roy Tennant.

The cause of linked data is not helped by the problem that, although the promise should appeal greatly to librarians, the 'killer example' isn't out there yet. Linked Data becomes more powerful the more of it there is. Until there is enough linking between collections and imaginative uses of data collections there is a danger librarians will see linked data as simply another metadata standard, rather than the powerful discovery tool it will underpin.

Particularly when compared to web 2.0 applications, linked data can seem rather inaccessible; anyone can create a Twitter account or promote user tagging or even contribute to mashups, but the world of linked data, for the moment, remains firmly in the hands of the experts. Nicholas Joint makes the analogy to the first days of HTML; librarians became interested in the power of placing information on the web, and thus learned HTML, not because they found HTML to be intrinsically interesting (Joint, 2008). Efforts need to be made in promoting linked data as relevant to the practice of every librarian while developing expertise where needed.

While general lack of awareness among librarians is an obstacle that needs to be overcome, a more practical concern is that changing the foundation of library metadata is no trivial task. In a case study of the socio-technological impacts of implementing semantic technologies, it was concluded that the major stresses of implementation would rest with the strategic and operational levels of the organization rather than the technological. The participants in the study perceived the organizational challenges to be overwhelming; "The basis for this seems to be the perception of the impact of ontology production and maintenance, and the strong focus on semantic meta-data production, all of which have to be established" (Bygstad 2009, 979). It's therefore understandable that libraries might focus on improvements to current systems such as discovery layers rather than a complete overhaul. The next section of this paper provides suggestions for how libraries can move forward in spite of this.

Finally, there are obstacles facing the linked data generally that have particular importance to libraries. Privacy is a huge concern for many interested in linked data. Not only will there likely be more information available, "Semantic Web agents will be able to explore the Web more aggressively than humans" (<u>Rezgui 2004</u>, 172). The more data is linked, the more it is possible that personal information will be exposed. Librarians, with their long tradition of protecting the privacy of patrons, will have to take an active role in linked data development to ensure rights are protected.

Related to privacy is trust. Ross Singer asserts that, "the largest hurdle to library adoption of Linked Data, though, may not be educational or technological [but instead] may be an issue of trust" (Singer 2009, 121). He argues that librarians have traditionally preferred to stick to models and resources over which they have complete control, rather than publically editable, user-centred tools. A counter argument to this is that librarians have long trusted subject experts to develop taxonomies for their disciplines. As linked data matures, the ability to access these trusted subject specific datasets and ontologies could become an advantage rather than an obstacle.

Rights management poses potential problems for linked data in libraries. Libraries no longer own much of the content they provide to users; rather it is subscribed to from a variety of vendors. Not only does that mean that vendors will have to make their data available in linked data formats for improvements to federated search to happen, but a mix of licensed and free content in a linked data environment would be extremely difficult to manage. In addition, Eric Hellman notes that, "unfortunately, RDF, the data model underlying Linked Data and the Semantic Web, has no built-in mechanism to attach data to its source" (Hellmann, 2009a). Without a mechanism to attribute datasets, it will be very difficult for libraries to open all the data to which they have access in a linked environment.

It is apparent that, although the list of early adopters continues to grow, linked data has a long way to go before it is seen as a standard foundation for library data. Although some of the challenges are immense, it is interesting that as with the wider world, the majority of issues are non-technical in nature. The technology is ready; it is now a matter of getting libraries and librarians ready as well.

# What needs to happen to move libraries to the next level?

As overwhelming as the idea of changing the underlying foundation of library metadata and resource description might be, movements within the library world are already driving this change through the developments of Resource Description & Access (RDA), Functional Requirements for Bibliographic Records (FRBR), and Functional Requirements for Authority Data (FRAD) standards. The implementation of RDA will drive the issue of linked data in libraries to some extent, as in order for libraries to fully leverage the power of RDF, it must play well with RDA and other emerging library standards and principles. One of the stated goals of RDA is to provide "data that is readily adaptable to new and emerging database structures" (Joint Steering <u>Committee for Development of RDA 2010</u>, Sect. 1.3). However, some librarians following its development have been critical of the RDA's reliance on legacy standards designed for the card catalogue world, such as creating

uniform titles for grouping items, highly structured format descriptions, and emphasis on textual "citation-like" notes fields. "These legacy practices fly in the face of the reality that in the digital world, identity is rarely expressed in a textual way, but instead standard linking technologies with Uniform Resource Identifiers (URIs) are preferred." (<u>Coyle, 2007</u>)

This discontent with the development of RDA has spurred several initiatives of interest to linked data followers. In 2007, the Dublin Core (DC) Metadata Initiative RDA Task group was formed, with a charter to "define components of the draft standard "RDA – Resource Description and Access" as an RDF vocabulary for use in developing a Dublin Core application profile" (<u>Dublin Core Metadata Initiative, 2010</u>). The major goals of the Task Group are:

- 1. Define RDA modeling entities as an RDF vocabulary (properties and classes).
- 2. Identify in-line value vocabularies as candidates for publication in Resource Description Framework Schema (RDFS) or Simple Knowledge Organization System (SKOS).
- 3. Develop a DC Application Profile for RDA based on FRBR and FRAD.

It is hoped by many this work will provide tools to push forward with linked data within the framework of RDA – "We'll have a standard set of terms that everyone can use, whether they're reading Resource Description and Access and cataloguing a book at the National Library of Canada or whanging some metadata into a blog's RSS feed in New Delhi." (Denton, 2010)

Several other projects are pushing linked data tools into RDA and FRBR. Richard Newman and Ian Davis created the Expression of Core FRBR Concepts in RDF. A work in progress, it includes FRBR group 1,2 and 3 entities. (Davis, 2005) Ian Davis used this to represent a FRBRized version of Harry Potter and the Goblet of Fire in RDF. (Davis, 2006). Martha Yee, a Cataloging Supervisor at the UCLA Film & Television Archive, has also experimented with building an RDF model of FRBR-ized cataloging rules, which presents some of the problems of using RDF to describe bibliographic data. Yee concludes, "If some of the RDF problems described above are insoluble, we may need to work with Semantic Web developers to create a more sophisticated version of RDF that can handle the transitivity and complex linking required by our data" (Yee 2009, 68).

The issue of RDA and its relationship to linked data is a complex one. Some librarians question whether RDA should be sidestepped entirely in favour of developing rules more in sync with linked data principles, while others are focussing on how to make RDA work with semantic web. The outcome of these discussions will have a major impact on the ability of libraries to move to linked datasets for bibliographic data.

As discussed in previous sections, the ability of the library community to build and sustain the ontologies needed for linked data is a substantial challenge. It is highly desirable to automate at least some of the process of building and linking large-scale vocabularies, taxonomies, and ontologies. Enterprise search companies like Sinequa, Endeca, and Automony are refining automatic classification tools that are able to ingest and organize large chunks of semi-structured data from different sources. A Sinequa implementer from one of France's largest wireless providers told Content Management System (CMS) Analyst Theresa Regli that "in over a hundred categorizations, we only have found two small errors in the past year" (<u>Regli, 2009</u>).

Another large implementer agreed, adding "We wouldn't have undertaken an enterprise taxonomy project because we never could have spent the time and money to write scripts or manually tag everything afterwards." (Regli, 2009). As librarians will no doubt appreciate, automated classification solutions will not completely obviate the need for human curation. People will still be necessary to choose authoritative sources for crawling, and to perform quality control at the end. Ideally, software interfaces will eventually allow for social curation & crowd sourcing among identified experts, so that ontologies can be edited and maintained collaboratively. According to Eric Hellman, "In many respects, the most important question for the library world in examining semantic web technologies is whether librarians can successfully transform their expertise in working with metadata into expertise in working with ontologies or models of knowledge" (Hellman, 2010b). Included in that transformation must be an embracing of new methods for creating ontologies.

# How can libraries get involved?

Eric Miller noted in a 2004 talk that libraries have four major roles in the semantic web:

- 1. exposing collections use Semantic Web technologies to make content available;
- 2. web'ifying thesaurus/mappings/services;
- 3. sharing lessons learned;
- 4. persistence. (Miller, 2004)

Looking at these four points, there are opportunities for individual institutions and librarians to push linked data work forward.

## **Exposing Collections**

Many libraries simply do not have the resources to transform all their data stores into linked data, but most have small standalone collections of structured data that can be used to develop expertise and technologies within libraries. Taking advantage of resources in platforms that support linked data such as DSPACE and Fedora as well as the many freely available tools to convert data to RDF can operationalize what seems like an insurmountable task.

### Web'ifying thesaurus/mappings/services

Exposing data, while a great first step, is not enough to realize the full potential of linked data. Barbara Tillet makes a convincing argument that this is a particularly significant role for libraries to fulfill. Librarians have been using controlled vocabulary as a corner stone of their services, and "converting these tools and vocabularies to Semantic Web standards, such as the Web Ontology Language (OWL), will provide limitless potential for putting them to use in myriad new ways." Supporting and contributing to the efforts underway such as those of the DCMI, as well as web'ifying locally maintained controlled vocabulary is a natural fit for the profession.

#### Sharing Lessons Learned

Sharing is something that comes naturally to the library community, but in addition to the usual methods of dissemination such as papers, conferences and social media within the profession, librarians must engage with - and contribute to - wider linked data community efforts. The semantic web is about breaking down silos, not building better ones.

#### Persistence

There is a large advocacy role for librarians in linked data. Librarians must demand that vendors develop their own data semantically. Institutions can work on data in catalogues and digital repositories, but if datasets in article databases and electronic resources remain mired in web 1.0, the true benefits of linked data cannot be realised. Additional advocacy efforts around open data, a corner stone of linked data principles, will also further development of linked data; it might be simplistic to proclaim "data wants to be free", but being able to access and reuse publicly funded data will enrich all the resources libraries offer users.

Linked data represents a major opportunity for libraries to integrate our information resources with the wider web. In addition, it provides a foundation for better data management, allowing libraries to store, share and re-use data as needed. The technology is finally ready; it's critical for libraries to begin preparations to become full participants in the world of linked data.

# Appendix - Semantic Web Tools for Libraries

11/3/2015

## **RDF Converters**

- MARC/MODS RDFizer
- Marc2rdf-Modeler
- OAI-PMH RDFizer
- BibTEX -> RDF
- Dublin Core -> RDF Crosswalk
- DC.Metadata Gen
- Simile Project RDFizers
- D2R Server

### **RDF** Publishing Tools

#### <u>Drupal 6 & 7</u>

The current stable version is Drupal 6, which does not include core support for RDF, but there are several contributed modules to produce RDF export (including evoc which allows you to import external vocabularies to Drupal, and expose those classes and properties to other Drupal modules for reuse). In Drupal 7 the ability to map the data structure to RDF and expose this in RDFa will be ported to the Drupal 7 core. This means that even if the site manager has no knowledge about RDF, Drupal 7 sites will expose the common elements like title, author, date, etc. as RDFa.

## Eprints 3.2.1+

Eprints is open-source repository software that is widely deployed in libraries around the world. As of version 3.2.1 Eprints has included several semantic elements, including Export formats: RDF+XML, N3, N-Triples, URIs for derived entities like authors, events, or locations, and an extendable RDF system that uses the BIBO Ontology by default.

## **Fedora**

Popular open source digital library platform Fedora is natively semantic, with an integrated RDF triple store called Mulgara. Fedora allows libraries to offer RDF output of record level citations.

#### **Zemanta**

Bloggers can incorporate RDF content by using semantic tagging services. Zemanta is a real-time semantic analysis tool that plugs into Movable Type, TypePad, and Drupal . As you blog, Zemanta performs on the fly term extraction, disambiguates by examining the surrounding context, and suggests appropriate enrichment material.

#### Semantic Media Wiki

Ontoprise has developed a set of Semantic MediaWiki (SMW)+ extensions to MediaWiki that provide a community-based environment for authoring ontologies and creating semantically enhanced wikis. Semantic mark-up of end-user data is enabled through

structured webforms, easy-to-use tagging and annotation tools. Ability to output data in a variety of visual formats, as well as to XML standards like vCard & BibteX.

#### **OpenCalais Semantic Proxy**

SemanticProxy translates the content of any URL on the web to its semantic representation in RDF, HTML or Microformats. Give SemanticProxy.com the address of a web page. Get rich semantic metadata about the people, companies, events and relationships on that page.

### **Calais Viewer**

To see the entity extraction process in action paste some unstructured text into the Calais Viewer. It will return the major entities, topics, and relationships. The Calais web service automatically attaches rich semantic metadata to the content you submit. Using natural language processing, machine learning and other methods, Calais categorizes and links your document with entities (people, places, organizations, etc.), facts (person "x" works for company "y"), and events (person "z" was appointed chairman of company "y" on date "x")

### <u>Virtuoso</u>

Virtuoso Universal Server is a middleware and database engine hybrid that combines the functionality of a traditional RDBMS, ORDBMS, virtual database, RDF, XML, free-text, web application server and file server functionality in a single system. The open source edition of Virtuoso Universal Server is also known as OpenLink Virtuoso.

### Talis Platform

Supporting data publishers and developers, the Talis platform provides dedicated cloud storage for RDF data stores. The content and metadata becomes immediately accessible over the Web and discoverable using both SPARQL and a free text search system with built in ranking of results according to relevance to the search terms.

### Linking Hubs

The major production linking hub of interest to libraries is the <u>Library of Congress Authorities and Vocabularies</u> <u>Service</u>, which offers all LCSH, including Genre/form headings, Children's subject headings, Subdivision records, and Validation records in SKOS and JSON formats. In the future, the Library of Congress plans to release other vocabularies this way including the Thesaurus of Graphic Materials, MARC Geographic Area Codes, MARC Language Codes, and MARC Relator Codes.

The <u>Virtual International Authority File (VIAF)</u>, a joint project of several national libraries with support from OCLC, was released in linked data format in September 2009. It contains over 9.5 personal names drawn from over 30 national name authority files.

There are many other linking hubs that may be useful to libraries, including <u>Dewey Summaries</u>, <u>MESH</u> headings, and <u>RAMEAU</u> subject headings from the French national Library.

# **Ontologies & Schemas**

<u>MarcOnt</u> is an initiative in development that attempts to capture concepts from MARC21 and other legacy bibliographic systems and transform them into machine-readable data. It involves a set of tools including an Ontology, Mediations Services, RDF Translator, MarcOnt Portal and is closely connected to the JeromeDL semantic digital library platform.

<u>Bibliographic Ontology Specification</u>, known as BIBO, has as its goal to "be used as a citation ontology, as a document classification ontology, or simply as a way to describe any kind of document in RDF." It is used as a format for exporting data from several semantic platforms, including Talis Aspire.

<u>Open Archives Initiative Object Reuse and Exchange</u> (OAI-ORE) was developed to solve the problem of indentifying an aggregation of resources on the web, such as a Flickr photo stream. OAI-ORE uses Resource Maps to provide aggregations with individual URIs and machine-readable data describing it, including information about the aggregation itself (like who created it) as well as information about the relationships between the resources in the aggregation. OAI-ORE can be expressed in several semantic formats such as RDFa, RDF/XML, turtle and n3. A powerful tool for digital libraries, OAI-ORE is being used by Chronicling America to provide linked data about aggregations of items such as the individual pages in an issue of a newspaper.

Other ontologies/schemas that may be of use to libraries include the <u>EDI Ontology</u> and the FRBR Ontology from <u>SchemaWeb</u>.

### Projects

<u>LIBRIS</u> is a Swedish union catalogue that contains approximately six million bibliographic records from 175 libraries. LIBRIS contains a URI for every resource and exposes data using FOAF, SKOS, BIBO and Dublin Core. It also links to external linked data sources such as dbpedia and LCSH.

The Library of Congress has developed its <u>Chronicling America Digitized Newspaper Database and Directory</u> as linked data, using Dublin Core, Bibliographic Ontology, FOAF, and Object Reuse and Exchange (OAI-ORE). The database contains digital views of historical newspapers as well as a directory of newspapers in the United States, 1690 to the present.

The <u>Open Library</u> has a goal of providing "One web page for every book". It provides a URI for each item in its system. The project is currently exploring other linked data approaches in order to enhance "the ability for people to connect our records with many more systems online."

Journal publishers are also doing work with semantic publishing:

The DBLP database provides bibliographic information on major computer science journals and conference proceedings. The database contains more than 800,000 articles and 400,000 authors. The <u>D2R Server</u> is based on the XML dump of the DBLP database, which has been read into a MySQL database. The complete RDF view on the database consists of approximately 15 million RDF triples, which are served in small easily consumable chunks.

The Royal Society of Chemistry launched its <u>Prospect Structure Search</u> in 2007. It uses enhanced HTML to incorporate standard metadata in articles that links related articles together and allows for search on structures and sub-structures in RSC articles.

Elsevier has experimented with Structured Digital Abstracts (SDAs) in the journal <u>FEBS</u> <u>Letters</u>. The abstracts of ninety papers were converted to machine-readable data, with information including all named entities such as genes and the results of the papers using controlled vocabularies.

Other efforts include the <u>ChemSpider Journal of Chemistry</u> and <u>Public Library of Science</u> (PLoS) <u>Neglected Tropical Diseases (NTD</u>.

Hosted on the Talis Platform, the <u>Linked Periodicals Project</u> provides a conversion of the journal metadata provided by CrossRef, Highwire and the NLM outputs citations as RDF/XML, JSON, Turtle.

# References

[1] Bygstad, Bendik, Gheorghita Ghinea and Geir-Tore Kl�boe. 2009. Organisational Challenges of the Semantic Web in Digital Libraries: a Norwegian Case Study. Online Information Review 33 no. 5: 973-985. http://dx.doi.org/10.1108/14684520911001945 (accessed August 18, 2010).

[2] Coyle, Karen and Diane Hillman. 2007. Resource Description and Access (RDA): Cataloging Rules for the 20th Century. DLib Magazine 13 no.1/2. <u>http://www.dlib.org/dlib/january07/coyle/01coyle.html</u> (accessed August 18, 2010).

[3] Davis, Ian and Richard Newman. 2005. Expression of Core FRBR Concepts in RDF. Contributor Bruce D'Arcus. <u>http://vocab.org/frbr/core.html</u> (accessed August 18, 2010).

[4] Davis, Ian. Harry Potter in FRBR. Internet Alchemy (June 13, 2006). http://blog.iandavis.com/2006/06/harry-potter-in-frbr (accessed August 18, 2010).

[5] Denton, William. RDA + DC + FRBR + FRAD + RDF = OMG!!! The FRBR Blog (February 12, 2010). http://www.frbr.org/2007/05/04/rda-dc-frbr-frad-rdf (accessed August 18, 2010).

[6] Dublin Core Metadata Initiative. Front Page. DCMI/RDA Task Group Wiki (October 13, 2009). <u>http://dublincore.org/dcmirdataskgroup/</u> (accessed August 18, 2010).

[7] Hellman, Eric. The Blank Node Bother and the RDF Copymess. go to hellman (November 5, 2009). <u>http://go-to-hellman.blogspot.com/2009/11/blank-node-bother-and-rdf-copymess.html</u> (accessed August 18, 2010).

[8] Hellman, Eric. Can Librarians Be Put Directly Onto the Semantic Web? go to hellman (August 4, 2009). <u>http://go-to-hellman.blogspot.com/2009/08/can-librarians-be-put-directly-onto.html</u> (accessed August 18, 2010).

[9] Isele, Robert, Anja Jentzsch, Chris Bizer and Julius Volz, 2010. Silk - A Link Discovery Framework for the Web of Data. (October 6, 2010) <u>http://www4.wiwiss.fu-berlin.de/bizer/silk/</u> (accessed August 19, 2010).

[10] Joint Steering Committee for Development of RDA: Resource Description and Access. Frequently Asked Questions (January 18, 2010). <u>http://www.rda-jsc.org/rdafaq.html</u> (accessed August 18, 2010).

11 Joint, Nicholas. 2008. The Practitioner Librarian and the Semantic Web. Special issue, Library Review 57 no.3: 173-186. <u>http://dx.doi.org/10.1108/00242530810865466</u> (accessed February 20, 2010).

[12] Miller, Eric. The Semantic Web and Digital Libraries. DC 2004 / SILF 2004. Shanghai, China (November 13, 2004). <u>http://dc2004.library.sh.cn/english/prog/ppt/talk.ppt</u> (accessed August 18, 2010).

[13] Miller, Paul. 2008. Sir Tim Berners-Lee: Semantic web is open for business. ZDNet.Com (February 26, 2008). <u>http://blogs.zdnet.com/semantic-web/?p=105&tag=trunk;content</u> (accessed August 18, 2010).

[14] Regli, Theresa. 2009. The Death of Taxonomies, revisited. CMS Watch (November 13, 2009).

http://www.dlib.org/dlib/november10/byrne/11byrne.print.html

http://www.cmswatch.com/Blog/1737-Death-of-Taxonomies-Revisited (accessed August 18, 2010).

[15] Rezgui, Abdelmounaam, Bouguettaya, Athman and Mohamed Eltoweissy. 2004. SemWebDL: A Privacy-preserving Semantic Web Infrastructure for Digital Libraries. International Journal on Digital Libraries 4 no.3:
171-184. <u>http://dx.doi.org/10.1007/s00799-004-0081-0</u> (accessed August 18, 2010).

[16] Schonfeld, Erick. 2010. Bing, Google, and the enigmatic T2: The race for a complete semantic search engine. TechCrunch (January 22, 2010). <u>http://www.techcrunch.com/2010/01/22/t2-bing-google-radar-semantic-search-race/</u> (accessed August 18, 2010).

[17] Singer, Ross. 2009. Linked Library Data Now! Journal of Electronic Resources Librarianship 21 no.2: 114-126. <u>http://dx.doi.org/10.1080/19411260903035809</u> (accessed August 18, 2010).

[18] Spivak, Nova. Eliminating the need for search – help engines. Minding the Planet. (January16, 2010). http://www.novaspivack.com/uncategorized/eliminating-the-need-to-search (accessed August 18, 2010).

[19] World Wide Web Consortium (W3C). Semantic Web. (October, 2010). http://www.w3.org/standards/semanticweb/ (accessed October 8, 2010).

[20] Yee, Martha M. 2009. Can Bibliographic Data Be Put Directly Onto the Semantic Web? Information Technology and Libraries 28 no.2: 55-80. <u>http://escholarship.org/uc/item/91b1830k</u> (accessed August 18, 2010).

# About the Authors



**Gillian Byrne** is the Head of Electronic Resources and Serials & Acquisitions at Memorial University of Newfoundland Libraries. Prior to that, she was the Head of Circulation and Document Delivery and an Information Services Librarian at the Queen Elizabeth II Library, with responsibilities for the Library web presence, federated search and the Explore a Topic portal. Gillian has published and presented on web usability and information seeking behavior, and is excited to explore what linked data will bring to users.



Lisa Goddard is the Acting Associate University Librarian for Information Technology at Memorial University of Newfoundland Libraries. She is currently head of Memorial's Digital Archive Initiative and manages the university's Research Repository. Lisa's current research interests include the semantic web & linked data, digital libraries, text mining, Open Access publishing, and scholarly communications.

(On November 17, 2010, the spelling of Barbara Tillett's name, and factual errors in the description of the Virtual International Authority File (VIAF), were corrected.)

Copyright  $\ensuremath{\mathbb{C}}$  2010 Gillian Byrne and Lisa Goddard

# PRINTER-FRIENDLY FORMAT