Forming the Canadian Linked Data Initiative

Caitlin Tillman  
*Victoria University Toronto, University of Toronto*

Joseph Hafner  
*McGill University (Montreal, Quebec)*

Sharon Farnel  
*University of Alberta (Edmonton, Alberta)*

http://docs.lib.purdue.edu/iatul/2016/data/1

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
BUILDING THE CANADIAN LINKED DATA INITIATIVE
A PRESENTATION FOR THE 2016 IATUL CONFERENCE IN
HALIFAX, CANADA.

By:
Caitlin Tillman, University of Toronto, Toronto, Canada
Joseph Hafner, McGill University, Montreal, Canada, and
Sharon Farnel, University of Alberta, Edmonton, Canada

Abstract:

In the fall of 2015, librarians from the University of Toronto, McGill, the University of Montreal, University of Alberta, and the University of British Columbia met to discuss the potential of collaborating on linked data efforts. As a result of that meeting, and with the inclusion of Library and Archives Canada, Bibliothèque et Archives du Québec and Canadiana.org, over the next six months the Canadian Linked Data Initiative (CLDI) was born. This paper describes the problem the collaboration is trying to solve, discusses the processes and tools the group is using to work collaboratively to articulate, plan, and find funding for a suite of projects that will coordinate the metadata production activities of our Libraries’ technical services units, special collections, and digital libraries in both English and French, and then concludes with an articulation of some of the challenges still ahead.

Presentation:

June 2015 - Sept 2015 (CAITLIN)

Since the story of the conception and creation of the Canadian Linked Data Initiative - a lively and fascinating story about transformation and about collaboration - would be almost meaningless without a basic understanding of what linked data is and what it can do for the library community and, since in our experience building this collaboration, we have noted that while the words “linked data” or “open linked data” have often been heard, it is less often understood what linked data is and why it is important to academic libraries, I'll first address those two things, and then we'll launch into the larger story of the Canadian Linked Data Initiative.

Briefly and simply, and without expertise, here are the basics and to provide some context. First, linked data for libraries is about cataloguing – or metadata creation. For all of my life, and for all of the careers of everyone in this room, libraries have created metadata, by and large, according to the MARC standard. And now, after several years of announcing MARC’s demise, there is actual, real evidence that change is coming.

The MARC standard, developed in the 1960s to represent the bibliographic data recorded on printed catalogue cards, does not easily translate to a web environment. This is because the data stored in a MARC record, while machine-readable by MARC-aware applications, is not directly machine-linkable. This means that human mediation is required to understand the data found in a MARC record and connect it to data found anywhere else on the web. In the web environment, the need for human intervention to link data limits its usefulness and expressiveness and has confined the MARC record and its high-quality, library-created, metadata to library catalogues. Translating textual data into machine-linkable “identifiers” such as ISNI or VIAF identifiers, and explicitly defining the relationships among the data using a more commonly understood RDF schema, will allow libraries to integrate our rich, high-quality data with the worldwide information community.
That sounds great – and it will be great – but a shift away from MARC to a linked data framework can be equated in terms of complexity, expense, and difficulty to the shift from card catalogues to OPACs. This time, however, the main challenge will not be technological; it will be cultural. Not only does linked data represent a new way of thinking about resource description, it also provides the opportunity for libraries to change the current practice of creating specialized metadata in different units of one library (e.g., cataloguing, special collections, archives, maps and data, the institutional repository, digital collections). To be successful, libraries will have to work across departments, institutions, and countries to create new workflows and tools and adapt to a new conceptual understanding of descriptive metadata. And thus begins the story of the CLDI.

A bit of background: In my capacity as Associate Chief Librarian Collections and Materials at the University of Toronto, I attend a bi-annual meeting of the Technical Services Directors of Large Research Libraries – a meeting of the top 21 ARL institutions plus Stanford, the Library of Congress, New York Public Library, and the National Library of Medicine, also referred to as the ‘Big Heads’. Each year I have returned to Toronto reassured that we are on the right path, we are doing the right things, at the very least we are in the same boat as everyone else, and this made me feel calm and happy. I enjoyed this meeting that made me feel calm and happy - until last June. At the 2015 ALA Annual meeting a large portion of the agenda and much of the discussion centred on the linked data initiatives of our peers. Although there are three Canadian institutions represented in the top 25 ARLs (UBC, Alberta and Toronto) none of the largest, most research-intensive university libraries in Canada were in a position to contribute substantially to the conversation. I did not return from that meeting feeling calm and happy – and was further disheartened to note that although there are small, dispersed initiatives in Canada, our peer institutions in the United States and in Europe are quickly outpacing us in their efforts to learn, develop, and apply new metadata standards. So wanting that calming and happy feeling back, I went to talk to my boss, Larry Alford, Chief Librarian at the University of Toronto Libraries and although he did not really understand what linked data was or why it was necessary that he care, it was clear to him that it mattered to me. [Small interjection here of particular relevance to this audience -- it was a universal experience of each of the U5 libraries that the cataloguers and Technical Services Directors were aware of linked data and its potential but Library Directors were not. One of our first discussions was about how to talk to the ULs about why linked data is important to libraries]. So like any good leader, Larry listened, asked some questions and then he told me to do something about it. So, I did what comes naturally in a crisis – I turned to my friends.

As it happens, and luckily, the collections and cataloguing directors of the Canadian U5 libraries (McGill, Montreal, UBC, Alberta and Toronto) meet regularly – by teleconference – to discuss common issues and challenges in technical services. At our July meeting I suggested that we should collaborate to “do something about linked data in Canada” and they all said: “yeah let’s do that”. In addition, a participant from each institution agreed to travel to Toronto in early September 2015 to attend the Access conference so that we could meet in person to talk about how to proceed.

In the next several months a collective structure was developed and the formulation of a collaborative plan began …

**Sept 2015 to April 2016 (JOSEPH)**

The Access Conference is a Canadian library technology conference bringing librarians, technicians, developers, programmers, and managers together to discuss cutting-edge library technologies. Following-up that meeting with a first meeting about linked data seemed fitting; this way we would hear various speakers on the topic at the conference, but could also bring in other speakers from the United States to share their experiences with us.
Our focus was on the group of U.S. Libraries who had received a Mellon grant and were collaborating on another Mellon grant related to linked data and their partners, which include: Stanford, Cornell, Harvard, Columbia, Princeton and the Library of Congress.

This worked out well for us, because we had people from each of our institutions attend either or both the Access Conference and our own mini-conference, along with opportunities during the conference to talk with various people.

We met with Phil Schreur, Stanford; Beacher Wiggins and Sally McCallum from the Library of Congress; Roy Tennant and John Chapman From OCLC; and Pat Riva from IFLA and the BAnQ. These speakers along with the presenters at the conference gave us a great lay of the land for what was happening with linked data in the U.S. with the partnership, and from a global perspective.

We decided after these meetings that we wanted to find a way forward as a group to collaborate on issues related to linked data in Canada. We knew that while Canada has done one large collaborative project using linked data related to WWI historical photographs called “Out of the trenches”, we could build on that first project and find ways to work together as Canadian institutions on issues related to linked data. We felt we were better off trying to do this together, knowing there are limitations in time and resources at each of our institutions, but pooling together would be ideal.

We shifted the focus of our collaboration at this point to be about Linked Data and decided we wanted to keep the momentum going. Our next step was to plan some meetings together via conference calls and then meetings for when several of us were attending ALA Midwinter in Boston.

Before Boston we decided to reach out to Diane Beattie at Library and Archives Canada who then agreed to have LAC become part of our group. We also decided to reach out to Bibliothèque et Archives du Québec and Canadiiana.org to see if they would be interested in being part of our endeavour. With the inclusion of these institutions, we formulated an initial manifesto for the Canadian Linked Data Initiative:

Working cooperatively, this group will articulate, plan, and find funding for a suite of projects that will coordinate the metadata production activities of our library’s technical services units, special collections, and digital libraries in both English and French. Working together and with partners in Canada, US, and Europe our goal is to provide a path to linked data readiness for our institutions and leadership for the adoption of linked data by libraries across Canada.

We decided that we wanted to continue to bring people together working with linked data already, so we had a meeting that included some of our group and guests from other organizations including:

- University of Toronto - Caitlin Tillman and Marlene Van Ballegooie
- McGill University - Joseph Hafner and Dawn McKinnon
- Université de Montréal - Julie Cardinal
- University of Alberta - Sharon Farzel
- University of British Columbia - JoAnne Newyear-Ramirez
- Library and Archives Canada - Diane Beattie
- Library of Congress - Beacher Wiggins and Sally McCallum
- Stanford University - Phil Schreur
- The George Washington University - Jackie Shieh
- Zepheira - Eric Miller (for intros), John Richardson, Gloria Gonzalez
- University of Manitoba – Lisa O’Hara
SirsiDynix kindly offered us the use of their office space at ALA and we held a 2 hour meeting to review various linked data initiatives and learn from the various people there. They encouraged us to find a way to work together as Canadian institutions to find a way forward.

We also got organized with an Intranet hosted by U of T to have an online working space that would facilitate our discussions and the working groups that we decided to create. Some of these groups have been meeting for a few months now, and others are just forming. These working groups include:

- Steering and Planning Committee
- Grant Committee
- Digital Projects
- Education and Training
- BIBFRAME Editor
- Identifiers
- French - Groupe de travail francophone
- IT Services
- User services

We decided one of our next steps is to host a Canadian Linked Data Summit in the fall, because we think will be good to reach out to other Canadians working in this area who aren’t yet part of our group and hear about their projects. We would also like to bring in international speakers to share their linked data projects with us, and we will have a day of workshops and breakout sessions organized by our various working groups. We will end the three day Summit with a ‘planning for the future’ day for the core members. This will take place from Monday, October 24 to Wednesday, October 26, 2016 in Montreal and McGill University will be one of the local hosts. We are looking into a SSHRC Connections grant for support of this meeting.

**Challenges & Concluding remarks - May 2015 & beyond (CAITLIN)**

We have accomplished a lot in a relatively short period of time with the support of our library directors and the enthusiastic participation from librarians in each of our institutions, who volunteer bits of their time, working on this project while continuing to do the jobs they were doing before we decided we should collaborate. But this is a massive transformation on a worldwide scale, and so it also comes with equally large challenges. While, once explained, it is easy to recognize a need and importance for libraries to a transition from MARC to linked data, it cannot be done individually and no one institution can do it alone; it is perhaps true that no one country can do it alone. Like the linked data metadata we propose creating, the move towards a linked data future touches all parts of the library, all cultural institutions everywhere, and individual practitioners. The challenges of making the transition in this large way are many. A few we have identified are the following:

**How do we get all units in the library (and possibly beyond) to commit to one standard?** At the moment, in most academic libraries metadata is created in several units in several ways. Each “subset” of library materials - Archives, Rare books, ejournals, digital collections, and repositories - all have different, all be it somewhat similar, metadata standards. Many university libraries are now in the process of setting up new Research Data Management units that appear also be working in isolation, or within themselves, to create metadata - metadata that may or may not use a standard similar to one of the other standards within the library. If each unit in each library institution continues to create metadata in its own way, we will create for ourselves a Big Data problem.

**How do we get government bodies, universities, public cultural institutions, and individual experts to work in concert?** One of the great things about linked data is that it has caught the interest of individuals in libraries and cultural institutions all over the world. Those
interested people have formed their own consortium, LODLAM (Linked Open Data in Libraries, Archives and Museums), hold summits, organize workshops, and generally learn from each other all over the world. It is very much an un-conference world where creativity and individual collaboration thrive. However, many library directors, those who are in the position to make a large, nation-wide change, are not aware of their work or the possibilities linked data holds. Linked data needs to be mainstream while still keeping the rogue practitioners working.

How do we scale this project across 9000km, multiple cultural institutions, and two languages? We have already had some small success in discussing the need, in Canada, to have documentation in two languages. In January we spoke with Eric Miller and others at Zepheria and by the end of February they had translated their BIBFRAME Lite vocabulary into French. This was a first boost of support but more is needed for this to be a truly Canadian effort. In addition to the need to work in two languages are the geographic challenges. Canada is a large country; while technology has made meeting across county easier and we do have regular teleconferences, nothing beats meeting face to face. Linked data is a complex issue in which language matters; words signify different things to different people. What I understand to be metadata or the cloud may be different from what my colleagues at McGill or Alberta understand, and in working together face to face these understandings can be more easily confirmed.

How does the user fit in? Presumably we’re interested in the user experience and what differences linked data library metadata will make, but we need people to work on this full time, learning from projects around the world such as the Oslo public library, and bringing those findings back to us to present a possibility for Canada. In the past few weeks we have been developing our connections to individual experts, trying to establish a method for us all to work together.

And finally, how do we fund this transformation in Canada? In some instances in Europe the move towards a linked data future for libraries and cultural institutions is being funded by governments. In the US, the Library of Congress and large, mostly private institutions, have their own resources plus grant funding driving some major projects. Currently Mellon is funding the Linked Data for Libraries and the Linked Data for Production initiatives - collaborations between Library of Congress, Stanford, Cornell, Columbia, Harvard, and Princeton. In Canada, where we traditionally tend to ride the middle ground and blend US and European practises, we will need to find our own way.

How do we meet all of these challenges? At the moment, I don't know; we none of us do. But by forming a group, comprised of academic libraries, government libraries, and individual practitioners from a variety of cultural institutions, we have taken the first steps towards creating a functioning, flexible, and funded framework which will enable Canada to play a part in the larger world-wide shift to linked data, and ensure that Canada's cultural heritage isn't lost on the expanding web.