

## **Integrating Discovery and Access to Canadian Data Sources. Contributing to Academic Library Data Services by Sharing Data Source Knowledge Nation Wide**

**Peter Webster**

Patrick Power Library, Saint Mary's University Library, Halifax, Canada.

Peter.webster@smu.ca



Copyright © 2019 by **Peter Webster**. 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:**

*This paper will describe the national project that has documented and indexed over 200 online data repositories in Canada. Available repositories cover the life and biological sciences, Chemistry, Physics, and Medical sciences, but also Mathematics and Computing in addition to other subjects. They provide data from academic, government and private research organizations. They represent a vast pool of openly accessible data available to academic researchers as well as citizen scientists.*

*Access to available data is an increasingly important part of scientific research and student exploration. Academic libraries across the country are working to ensure that they can effectively discover and provide access to data sources.*

*The paper will look at shared efforts to provide integrated search tools for these Canadian data sources, through regional consortia, government initiatives, and cooperative national projects.*

*Many academic libraries large and small lack the expertise to work with a wide range of specialized data. This paper will focus on efforts to share specialized knowledge and expertise across the country, rather than duplicating effort at each individual library. This process is allowing the development of shared library reference services for data resources. The paper will look at the existing library program for sharing national government data knowledge. It will look at how this program is being used as a model and a platform to share reference training and expertise for other data sources.*

**Keywords:** Canada, research data, National data inventory, Data Liberation Initiative, Portage Federated Research Data Repository

---

In recent years, the use of data has become an increasingly important part of academic research. Science agencies worldwide have supported the view that increased use of data analysis will be critical to driving innovation and discovery, and that open and shared access to large pools of data is leading to major advances in many fields of study. As one of many examples, The United States National Science Foundation (NSF) recently described their research data initiative as:

A national-scale activity to enable new modes of data-driven discovery that will allow new fundamental questions to be asked and answered at the frontiers of science and engineering... and will generate new knowledge and understanding and accelerate discovery and innovation” (United States National Science Foundation, 2019).

There is a strong international movement toward greater openness and accessibility for all kinds of data. Agencies, such as The World Bank, The International Monetary Fund (IMF), The United Nations (UN), and the Food and Agriculture Organization (FAO) are making major efforts to collect, aggregate, and provide open access to their data.

In Canada, as in many countries, the national government is working toward greater transparency and accessibility of government-produced data. The Canadian government has recently released “Canada’s 2018-2020 National Action Plan on Open Government”, including the Canadian Government “Roadmap for Open Science” (Government of Canada, 2018). Canadian funding agencies are endorsing open access for data produced by research grant funding (Science Canada , 2018).

Hundreds of research organizations, science and public interest groups, in many fields of study offer online data repositories. Many specialized data collections are available in order to collect share and distribute data. National and international government agencies rapidly moving to make more government data openly available online. Canadian provincial and municipal agencies have followed suit, making increasing amounts of their data available.

Academic libraries and library organizations in Canada are supporters of open data. They are actively involved in developing data management and repository resources for university produced research data. The Canadian Association of Research Libraries (CARL-ABRC) has been central in open access policy developments, and in developing data management resources. As an example, CARL-ABRC recently released the following statement on open government and open science.

From CARL-ABRC’s perspective, easy, timely and open access to government information and publicly funded research is of inestimable value for students, educators and researchers alike. Open science, for example, accelerates scientific discovery, enables international collaboration and coordination, and ultimately supports economic prosperity. (Canadian Association of Research Libraries)

The U.S. Association of College and Research Libraries (ACRL) *Research Planning and Review Committee* made the following statement about the growing need for library data services, in their 2018 Top Trends in Academic Libraries report:

The establishment of data science programs at numerous institutions has led to the need for librarians to adapt and integrate growing management, accessibility, and technical subject expertise to support data scientists. Professional associations and information science programs should continue to expand and enhance training in data management

and data analytics to prepare librarians in using and addressing big data questions with colleagues and patrons (“Top trends in academic libraries”, 2018).

### **Academic Library Reference and Research Data Services.**

The use of data in scholarly research and the amount of available data are increasing rapidly. Canadian academic libraries are seeing a growing demand for reference assistance in finding and using data. In many cases, open data sources have replaced former paper reference works.

These library reference services are working to take a wider role in locating and using data sources. However, they have encountered several barriers to developing data reference services. The first major barrier has been the lack of discoverable information about available data sources. The second barrier has been a lack of training for library reference staff, because data use, and particularly the use of scientific data has not been a major part of general academic librarians reference skill set.

### **Canadian Data Repositories Inventory**

In 2015, Saint Mary’s University partnered with several other library organizations, to begin what has come to be called *Canadian Data Repositories Inventory Project*. This project has provided essential information about Canada’s research data landscape. It has also helped address both barriers to library data reference service. This project has created a centralized and largely complete inventory of Canadian research data repositories, by consolidating and expanding on existing lists of Canadian data sources. As this project has progressed, a second step has been to promote awareness about the Canadian data repositories included in the inventory, and to provide national training on discovering and using them.

*The Canadian Data Repositories Inventory Project* was started by Saint Mary’s University, in cooperation with The National Research Council of Canada, Ontario Scholars Portal, the international data repository directory Re3data.org, with input from Research Data Canada’s Standards and Interoperability Committee, and other data contributors.

### **A Collaborative National Data Repository Inventory**

Prior to 2015, several directories of Canadian data sources had developed independently. The National Research Council of Canada offered a web-accessible *Gateway to Research Data*. Several international research data directories including Fairsharing, and Re3data.org listed many Canadian repositories. Scholar’s Portal the online technology service of the Ontario Council of University libraries (OCUL) had developed an extensive *Guide to Research Data Repositories*. Scholars Portal was developing a large Dataverse repository for Ontario provincial, and other Canadian University research data collections. The Canadian National Committee for CODATA International, volunteer group, had produced a semi-annual *Data Activities in Canada Report*, for several years up until 2012. At the same time, many Canadian academic libraries had also developed their own reference lists of data sources useful to their researchers.

The existing situation presented several challenges. The different lists of data repositories duplicated many entrees, each listed included unique resources not identified by others, and many well-known repositories were not listed in any source. The different lists often lacked clear or common definitions. Multiple data source efforts were resulting in duplication of

effort. Several data directory efforts were proving difficult to sustain, and becoming out of date.

The first phase of the inventory project took place in 2015 and 2016. Available Canadian data sources listed in different sources were updated, corrected and merged into a single list. Since then, the project has systematically identified and added data repositories from academic institutions. Most recently, it has added provincial and municipal open data catalogues to the list.

### **Re3data.org. International Cooperation to Inventory Canadian Data Repositories.**

A review of available options showed that the international data directory Re3data.org had the most complete listing of Canadian repositories, though it listed only one third of known Canadian repositories. Moreover, Re3data.org had the most developed search interface functionality. (Re3data.org, 2015. About). Re3data.org was already being used by some Canadian libraries, so improvements to Canadian listings in it would have immediate benefits. The Re3data.org organization, hosted at the Karlsruhe Institute of Technology (KIT) Library in Germany, was also, generously, willing to work with the Canadian project in a systematic way. Therefore, it was decided to use Re3data.org as the ongoing platform for Canadian Data Repository entries, rather than attempt to develop a new online platform. This allowed us to build on existing resources and made the project a truly international effort.

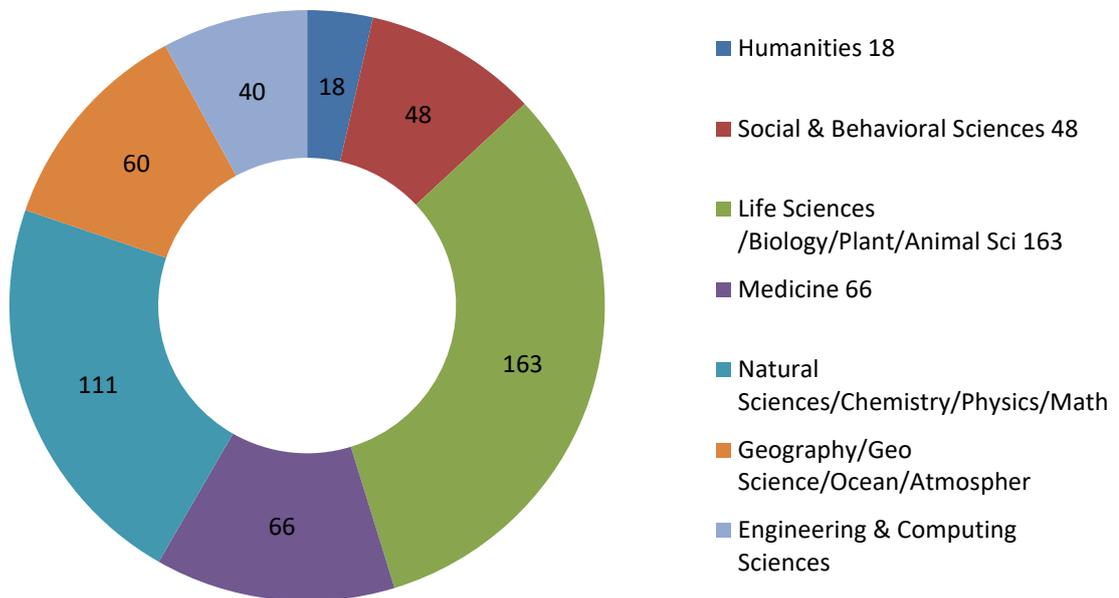
The Re3data.org directory also had existing standards for deciding what should be considered an ongoing “data provider” separate from a one-time data project, and what should be considered a “service provider” which offers access to multiple data providers. They were also applying standardized subject identifiers to each repository and applying standards for assigning national origin to repositories (Re3data.org, 2019). These standards were critical to a consistent data repository inventory. Standards were not well developed or consistent in existing listings of Canadian data resources. Although Re3data.org’s standards and definitions might be expanded or improved upon in the future, they have been an essential tool for the inventory project.

Since 2015, hundreds of additions and corrections, and over one hundred and fifty new Canadian data repository entries have been added to Re3data.org. There are currently over 225 Canadian data repositories listed in Re3data.org.

### **Subject and Organization Breakdown of Canadian Data Repositories in Re3data.org.**

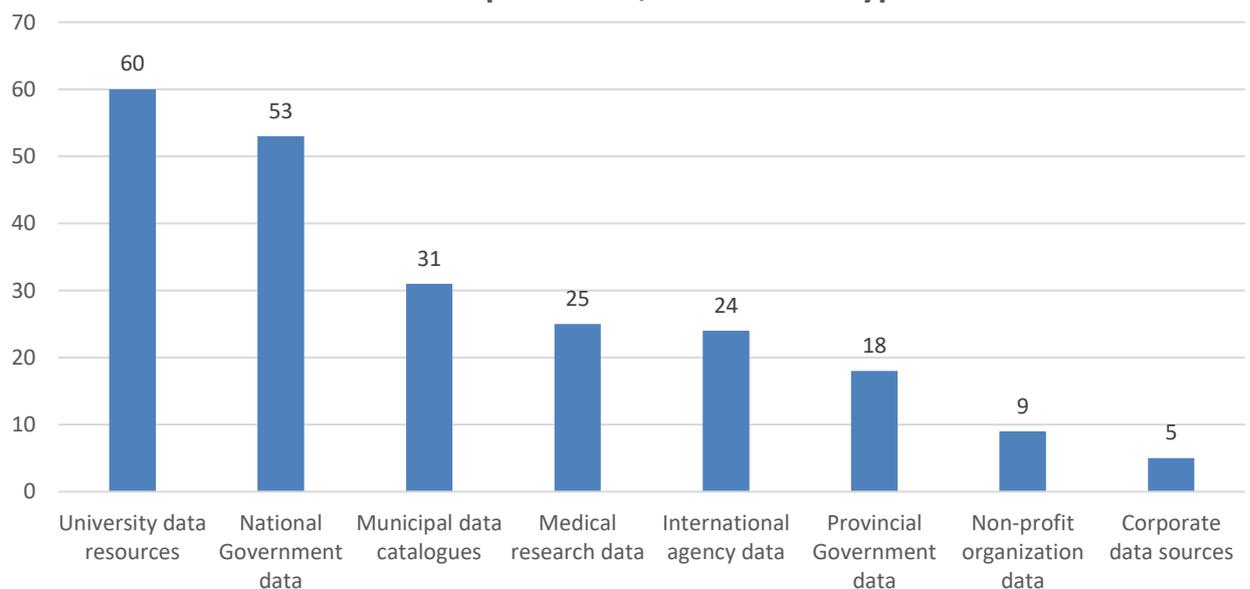
For the first time this inventory makes it possible to get a clear picture of Canada’s data repository landscape. It is possible to see the kinds of organizations that are making data available, and the subject distribution of Canadian data repositories, as listed below.

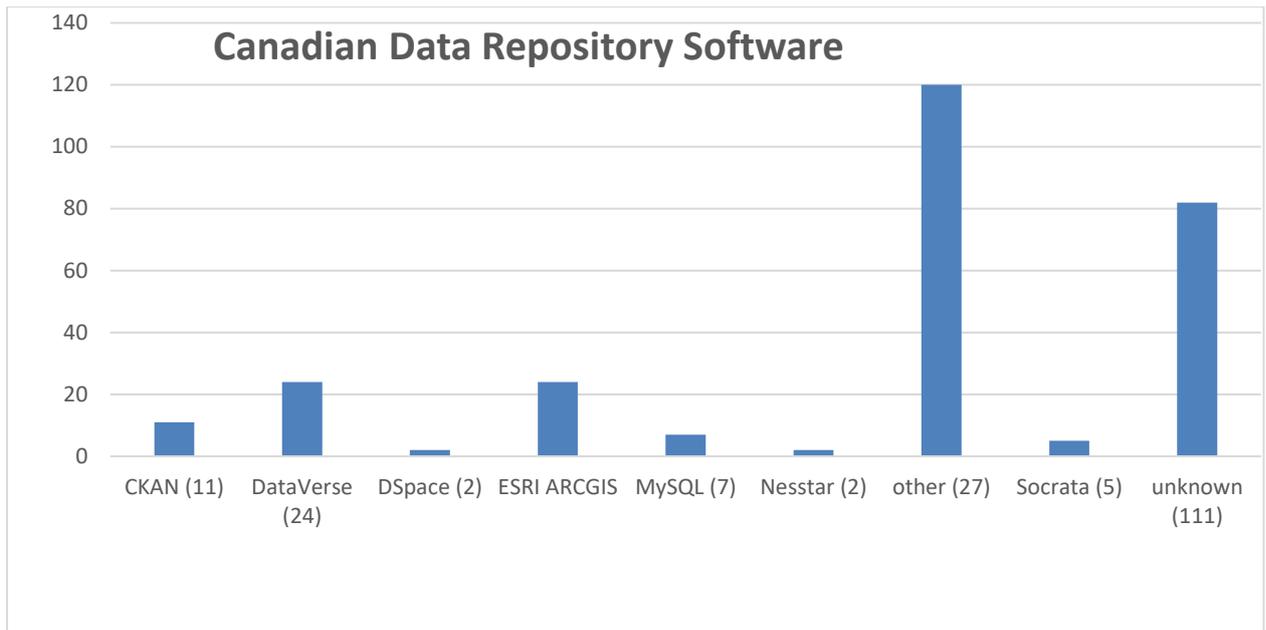
## Canadian Data Repositories by Subject



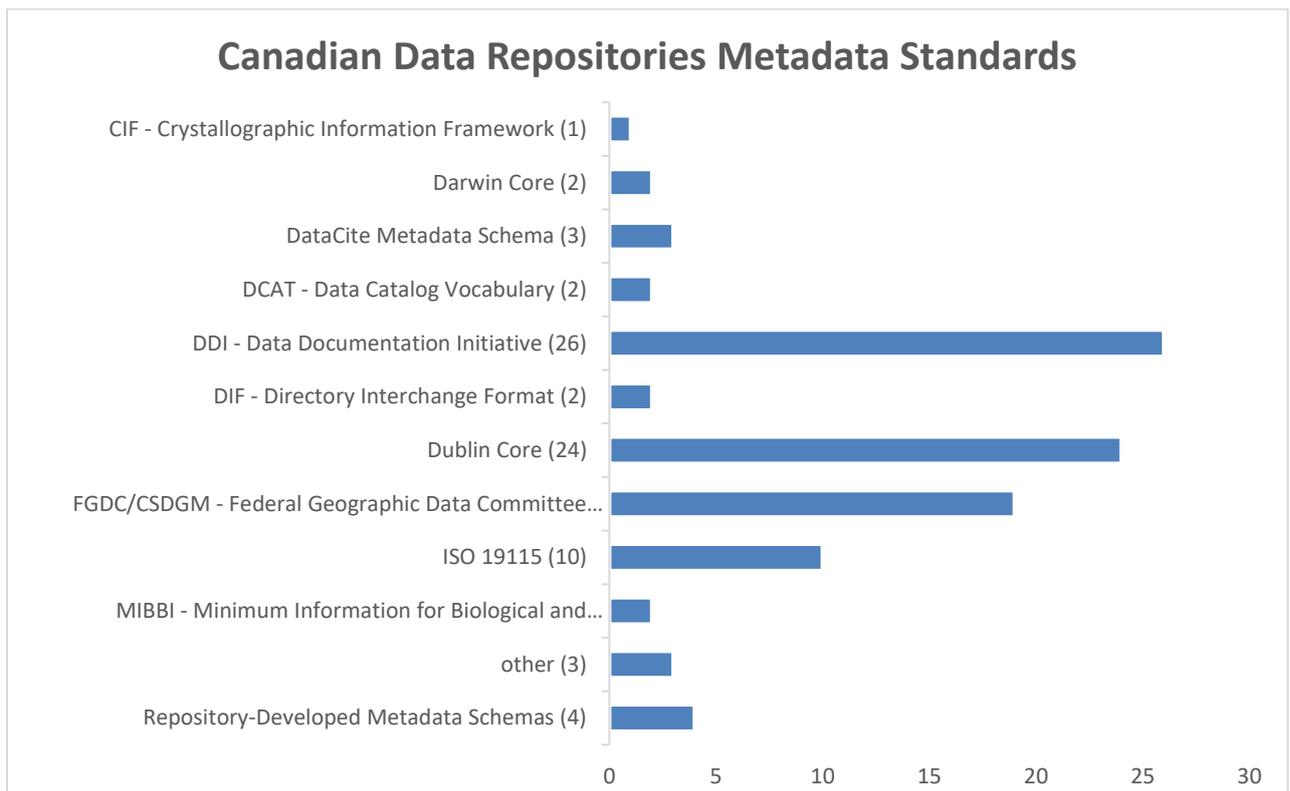
(Note that most repositories are assigned multiple subject headings. Therefore, this breakdown by subject is not a count of the number of repositories.)

## Canadian Data Repositories, Institutions Type





(Inventory work so far has provided valuable information about data repository software being used, but software for many repositories still needs to be identified.)



Work on the inventory has made Re3data.org a reliable and comprehensive reference tool for Canadian, as well as other data sources. It provides a wealth of new information about the kinds of data being collected in Canada. It provides additional information about the terms of use, access and sharing arrangements for each data source, as well as information about metadata standards, and software being used.

Re3data.org has also worked with the Canadian project to customize their software so that spreadsheets of Canadian, or any other national entries, can be easily downloaded for use by any library wishing to offer their own selected data reference listing and guides. Saint Mary's University has developed a spreadsheet tool for downloading the latest complete metadata for Canadian Research Repositories listed Re3data.org.

**The Canadian Data Repositories spreadsheet download tool is available here <http://library.smu.ca/re3data/Re3Data.xlsx>** (Saint Mary's University, 2019)

### **Building on Existing Networks for Canadian Data Repositories Training.**

As information about Canadian data repositories in Re3data.org has grown, it has become an ongoing task to educate Canadian libraries about this resource, and about the many important data sources listed.

Canada has several existing library networks that focus on skills training, including a well-developed government data-sharing network. These existing networks are a key resource for developing Canadian data source knowledge in academic libraries across the country.

The most important existing network for data training is the Data Liberation Initiative (DLI). This program began in 1996, as a joint project between Statistics Canada and Canadian universities (Statistics Canada, 2018). Over 75 universities now share central access to dedicated Statistics Canada resource staff to answer reference questions, and to direct researchers to specific data sources. The program is very successful at working across the complex Statistics Canada organization to locate data and get advice on data methods and use. The DLI program is focused on Statistics Canada resources. These o include a great deal of scientific data, particularly in areas like environment and health research.

The DLI has always featured a major component of data training. Each participating university library designates several data contact people. These contact people are brought together each year for data training at the regional level. Combined national level training takes place every several years. The DLI program maintains an ongoing "Boot Camp" program to train library staff who are new to working with data.

From the outset, this program has highlighted developing basic statistical literacy. DLI contacts are trained to understand how numerical datasets are tied to descriptive metadata. They gain experience in understanding the use of codebooks and essential field, variable and data set definitions.

The DLI program ensures that nearly all Canadian academic libraries have librarians and library staff with a knowledge of data sources and the ability to address data related questions. The ongoing training and peer support offered by the DLI program is of particular importance to many smaller universities where data and science work, is just one of multiple tasks for library staff.

In recent years, DLI training sessions go beyond Statistics Canada related data, to cover other data sources. DLI training sessions have been done on key Canadian data repositories, including the Ontario Scholars Portal Dataverse for university research data, provincial government data catalogues, and the Canadian Open Government Portal. DLI sessions have covered the use of application programming Interfaces (API) to extract data sets from many data repositories. They have included sessions on accessing research data held by major publishers like Nature/Science and Elsevier. They have included sessions on using major international open data sites including The World Bank, International Monetary Fund (IMF),

National Aeronautics and Space Administration (NASA), or World Food and Agriculture Organization (FAO). All of these can be discovered using Re3dat.org.

The DLI training programs have been an ideal venue for passing on knowledge about the data resources identified by the Canadian Data Repositories Inventory program. Therefore, several regional and national training sessions have been carried out looking at the work of the repository inventory, and at use of key Canadian data sources found in Re3data.org. (See Appendix 1)

In addition to the DLI program, other programs have helped develop library data reference services. Statistics Canada offers the Research Data Centres (RDC) program. This program supports detailed data research done at different universities across the country. It provides controlled access to confidential Statistics Canada datasets. This program also has a training component. The training is primarily aimed at developing the data skills of university researchers working with Statistics Canada data. Nevertheless, many librarians have participated in this training, learning to use common statistical analysis software such as SPSS, SAS, and Stata.

The Partnership of Provincial and Territorial Library Associations of Canada also offers a national online professional development & continuing education program. The first training session through this program on finding Canadian data resources using Re3data.org is scheduled for August 2019

A review of academic library information guides across Canada, finds that there are over 45 individual institutions that have “Data and Statistics” subject guides that list Re3data.org as a resource. A number of these guides now link directly to Re3data.org Canadian data entries (Re3data.org, countries=Canadian, 2019).

### **Getting the Word Out Via the Data Liberation Initiative Listserv.**

In addition to training, one of the key components of the Data Liberation Initiative (DLI) is an online Listserv for DLI Library representatives and Statistics Canada staff to share data knowledge and answer questions coming from university researchers. The DLI program receives on average 1000 questions each year.

The listserv is primarily intended to provide a fast and reliable channel of communication from Statistics Canada staff to University Libraries. Nevertheless, library data representatives commonly use the listserv to share information among themselves. A review of DLI listserv messages over the last 5 years shows that the listserv is used for many data related discussions beyond Statistics Canada. Use of the listserv for a broader range of data related questions is growing. Information about activities of the Inter-university Consortium for Political and Social Research (ICPSR) in the United States, and questions about ICPSR data are periodically shared the DLI listserv.

Since 2009, Science Boot Camps for Librarians have developed across the United States (Harvey, 2019). In 2015, the first of these in Canada “The True North Boot Camp” was announced via the DLI listserv. Information about “True North Boot Camps has continued to appear on the DLI listserv since then.

The DLI Listserv has been the ideal venue for notices about the progress of the Canadian Research Data Inventory as the project has advanced.

## **CARL-ABRC Portage Project FRDR.**

The Canadian Association of Research Libraries (CARL-ABRC) Portage project is a major initiative, developing infrastructure for managing university created research data, and for ensuring that this data is discoverable and openly accessible. The Portage project has developed a network of services and expertise, shared by libraries nation-wide. It is very similar to the Data Liberation Initiative (DLI) network.

This project has developed The Federated Research Data Repository (FRDR) discovery tool and data repository (Federated Research Data Repository, 2017). The FRDR online interface searches selected Canadian data repositories, and links to each individual dataset in these repositories. FRDR also serves as a direct data repository.

In 2018, the FRDR project became involved with the Canadian data inventory. The FRDR project is now using the Canadian inventor entries in Re3data.org to select repositories, which are suitable for data set metadata harvesting. FRDR has harvested data set links to some 40 Canadian data repositories so far. Participants in the FRDR project will carry on updating of the Canadian data repository inventory entries in Re3data.org.

### **In Conclusion.**

There have been several important benefits to developing a complete and consistent listing of Canadian data repositories in an open online platform. The use of Re3data.org provides an easy to use, reliable and richly featured discovery tool for Canadian data source information. It is a collecting point for information about subjects being covered, data standards being used, repository software, data sharing policies, licensing and much more. Working with Re3data.org, the project has also created download tools, so its collection of metadata can be used to better understand Canadian data resources and to help develop future data services. It offers an effective reference tool that is being used by many academic libraries across Canada.

The Data Liberation Initiative (DLI) partnership between Canadian university libraries and Statistics Canada is an existing network for training and exchange of information about data resources. This network has been an invaluable channel for disseminating information and providing training about the widening collection of Canadian data resources.

The project to collect a complete listing of Canadian Data resources and to disseminate knowledge about them has been an important step in the ongoing development of Canadian academic library data services.

### **Acknowledgments**

Evelyn Reuter, Karlsruhe Institute of Technology (KIT) & Re3data  
Karen Morgenroth, Tina Bradford, National Research Council Canada  
Amber Leahey, Scholars Portal Ontario  
Noramay McGruer, Saint Mary's University  
Allison Watt, Saint Mary's University  
Suzanne van den Hoogen, Saint Mary's University  
Dany Savard, York University

## References

- Canadian Association of Research Libraries (CARL-ABRC). (2019). **Open Government**. <http://www.carl-abrc.ca/influencing-policy/open-government/>
- Federated Research Data Repository. (2017). Find and Share Canadian Research Data. Retrieved from <https://www.frdr.ca/repo/?locale=en>
- Government of Canada. (2018). Canadas 2018-2020 National Action Plan for Open Government. Retrieved from <https://open.canada.ca/en/content/canadas-2018-2020-national-action-plan-open-government>
- Harvey, Barbara (2019). "How Librarian Science Boot Camps Fill STEM Gap". *JoVE May 23, 2019*. Retrieved from <https://www.jove.com/blog/library/librarian-guest-authors/how-librarian-science-boot-camps-fill-stem-gaps/>
- Re3data.org (2015). About. Retrieved from <https://www.re3data.org/about>.
- Re3data.org. (2019). Countries=Canadian. Retrieved from [https://www.re3data.org/search?query=&countries\[\]=CAN](https://www.re3data.org/search?query=&countries[]=CAN)
- Re3data.org. (2019). Suggest a repository: re3data.org Registration Policy. Retrieved from <https://www.re3data.org/suggest>
- Saint Mary's University.(2019). "The Canadian Data Repositories from Re3data.org Spreadsheet download tool". Retrieved from <http://library.smu.ca/re3data/Re3Data.xlsx>
- Statistics Canada (2018) "About the DLI" <https://www.statcan.gc.ca/eng/dli/guide/section1>
- Science Canada (2018). Research Data Management. Retrieved from [http://www.science.gc.ca/eic/site/063.nsf/eng/h\\_547652FB.html](http://www.science.gc.ca/eic/site/063.nsf/eng/h_547652FB.html)
- "Top trends in academic libraries". (2018). *College & Research Libraries News*, 79(6), 286–300. Retrieved from <https://crln.acrl.org/index.php/crlnews/article/view/17001/18750>
- United States National Science Foundation. (2019). *Harnessing the Data Revolution*. Retrieved from <https://www.nsf.gov/cise/harnessingdata/>