

# Research Data Management Training Landscape in Canada

## A White Paper

Prepared by the Portage Training Expert Group  
on behalf of the Canadian Association of Research Libraries (CARL)

Jane Fry (Carleton University)  
James Doiron (University of Alberta)  
Danny Létourneau (Université de Montréal)  
Laure Perrier (University of Toronto)  
Carol Perry (University of Guelph)  
Wendy Watkins (Carleton University, retired)

JANUARY 2017

Portage Network  
Canadian Association of Research Libraries  
[portage@carl-abrc.ca](mailto:portage@carl-abrc.ca)

[www.carl-abrc.ca](http://www.carl-abrc.ca)

**portage**  
SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE  
SHARED STEWARDSHIP OF RESEARCH DATA

# Table of Contents

- Introduction..... 2
- How is RDM expertise distributed in Canada? ..... 3
- Does Canada have the capacity for RDM? ..... 5
  - The National Data Summit .....5
  - Research Data Canada .....5
  - Other Discussions on RDM Capacity .....6
- How should Portage address RDM training in Canada? ..... 7
- Summary ..... 10
- Acknowledgements ..... 11
- Appendix 1: RDM Training Principles ..... 12

# Introduction

The Tri-Agency [Statement of Principles on Digital Data Management](#) identifies specific research data management responsibilities of key Canadian stakeholder groups involved in funded research.<sup>1</sup> Overall, the Tri-Agency's document heightens the need for a collaborative, national perspective on RDM that has been missing in Canada. This statement can be viewed as a challenge to stakeholders to evaluate their roles and responsibilities. What remains to be seen is if these groups will take up this gauntlet.

The [Portage](#) Training Expert Group is investigating strategic directions for RDM professional development that will address part of the academic library community's response to Tri-Agency expectations. In providing a national outlook, our central focus is to integrate the RDM expertise of our community with other research stakeholders and to collaborate in training initiatives within the broader RDM community.

This White Paper will provide a high-level perspective on RDM training by answering three fundamental questions.

1. What is the current RDM expertise landscape in Canada?
2. Does Canada have the capacity to meet the demand for training?
3. How can Portage help increase Canadian RDM expertise through training and other means?

To provide answers, we have investigated developments in RDM training in Canada and internationally, both inside and outside the library community. This has given us insight into significant issues and gaps in RDM training in Canada, resulting in a set of potential actions and recommendations. We intend to build upon and collaborate with other RDM training specialists. Some are in the process of developing their own training materials and we hope to benefit from their work. In return, we welcome others to use our training contributions.

Portage is committed to outreach in higher education through collaborative relationships with Canada's four regional library consortia: CAUL (Council of Atlantic University Libraries); BCI (Bureau de coopération interuniversitaire); OCUL (Ontario Council of University Libraries); and COPPUL (Council of Prairie and Pacific University Libraries). As such, our Training Expert Group will support both national and regional

---

<sup>1</sup> The Tri-Agency consists of [CIHR](#) (Canadian Institute of Health Research), [NSERC](#) (Natural Sciences and Engineering Research Council) and [SSHRC](#) (Social Sciences and Humanities Research Council).

training priorities. We exist to support them in their training efforts and see the sharing of information to be essential.

## How is RDM expertise distributed in Canada?

Canadian RDM expertise was located at an earlier time in one of our national institutions. In 1973, the Machine-Readable Archives (MRA) Division was formed as part of the Public Archives of Canada. This Division provided a home for data generated by the federal government and was a source of expertise on archiving data for the rest of Canada's RDM stakeholders. This central location provided a recognized source for RDM best practices in preparing data for dissemination and preservation. Unfortunately, the MRA did not survive a 1987 reorganization of the Public Archives. Data were not deemed to be a specialized format and the holdings were moved into subject-matter areas<sup>2</sup>. As a result, the centralized repository of professional expertise built over time within this Division was lost. Without these data specialists, a recognized, national centre of RDM expertise disappeared in Canada and instead became widely distributed across Canada in a variety of institutional contexts and without a common thread for collaboration or sharing.

Internationally, several examples of national centres of RDM expertise exist. The United Kingdom has both the [UK Data Archive](#) (UKDA) and the [Digital Curation Centre](#) (DCC). The United States has several domain-based centres providing national RDM expertise, including the [Interuniversity Consortium for Political and Social Research](#) (ICPSR) and [DataONE](#). Other widely known institutions with global recognition in RDM expertise include the [Australian National Data Services](#) (ANDS) and the [World Data System](#) (WDS) of the International Council for Science. In each of these cases, the value of their services is widely acknowledged and their reputation is built solidly on the generous sharing of their training materials and knowledge.

In Canada, other initiatives have advanced the cause of national RDM expertise since the demise of the MRA but with little success. In 1996, the Data and Information Panel of the Canadian Global Change Program<sup>3</sup> published "Data Policy and Barriers to Data Access in Canada: Issues for Global Change Research"<sup>4</sup> containing five categories of

---

<sup>2</sup> Not all the data are lost. A small fraction may be found at the Library and Archives Canada's (LAC) [Open Data site](#). Missing, for example, is the large store of tree-ring data that were part of the collection. These would certainly be useful to climate scholars among others.

<sup>3</sup> The Royal Society of Canada organized the working groups within the CGCP through funding provided by the federal government and administered by Environment Canada.

<sup>4</sup> Don C. McKay and David Henderson, "Data Management in Canada: Issues for Global Change Research," *Government Information in Canada*, 1, No. 2.6 (1994).

recommendations: infrastructure, archiving, documentation, access, and standards. This report called for national leadership in data management but no action was taken<sup>5</sup>. Similar recommendations for national RDM centres of expertise were made but they also failed to be implemented<sup>6</sup>.

While attempts to establish a national centre of expertise have been unsuccessful, pockets have developed across the country, although they have tended to exist largely in isolation of one another. For example, the National Research Council sponsors a [Canadian CODATA chapter](#) that connects our country to this interdisciplinary scientific committee of the International Council for Science. [CODATA](#) is an important RDM body internationally but the Canadian chapter is too poorly resourced to play a central role at the national level.

Another example is the [Data Liberation Initiative](#) (DLI), a subscription service between Canada's academic community and Statistics Canada. This program helped establish data services within academic libraries across Canada. Through its peer-to-peer training program, DLI has taught data service skills to librarians over the past 20 years. DLI has provided libraries with a solid foundation for expanded roles in RDM and many of those trained by DLI are now active in Portage Expert Groups.<sup>7</sup>

RDM expertise in Canada today is distributed across the country in multiple institutions at different jurisdictional levels. No systematic coordination of this expertise exists and the diverse nature of research data tends to isolate such expertise within disciplinary and agency silos of RDM specialists. In our environmental scan of RDM training across Canada, we discovered that the vast majority of it is provided by the library community. This positions libraries to be leaders in this area.

Portage is a distributed organization that is engaged in building a community of practice across the country. Its Network of Expertise is an essential component that exists to support experts locally, as well as to share their knowledge and skills with the wider community of practice.

---

<sup>5</sup> Much of this history has been adapted from Chuck Humphrey's excellent blog on research data in Canada, which describes our beginnings. Personal recollections are also included where warranted. <https://preservingresearchdataincanada.net/2012/12/05/hello-world/>

<sup>6</sup> The 2001 and 2002 National Data Archive Consultation jointly sponsored by the National Archives of Canada (NA) and SSHRC and the 2004 National Consultation on Access to Scientific Research Data (NASCARD) made these recommendations.

<sup>7</sup> It should be mentioned that, in the interests of space, we have chosen to highlight only two examples of initiatives involving pockets of RDM expertise across the country.

## Does Canada have the capacity for RDM?

Recent national discussions on RDM repeatedly raise the question: does Canada have the capacity to deal with the scale of research data being produced? This section reviews this question in light of the 2011 National Data Summit, a 2014 report from Research Data Canada, and a few other important discussions on this topic.

### The National Data Summit

The Canada Institute for Scientific and Technical Information (CISTI) established the Research Data Strategy Working Group (RDSWG) in 2008 as a grassroots response to the 2004 National Consultation on Access to Scientific Research Data. One of the RDSWG contributions was a commissioned analysis of the gaps in data stewardship in Canada. This report was updated in 2011 and served as a background document for the 160 senior ‘data evangelists’ who attended the National Data Summit in September 2011. The ensuing report, [Mapping the Data Landscape: Report of the 2011 Canadian Research Data Summit](#), outlined a five-year plan of action for various stakeholders.

This report concluded that Canada did indeed have a base-level capacity for RDM and suggested how RDM training might be delivered in Canada to increase this capacity. Of particular significance, training highly qualified personnel and establishing curricula throughout the educational sector were explicitly identified as components of research infrastructure. The report included a high-level goal of RDM, “Capabilities and Education”, putting forth the need for:

1. Data management training modules implemented within research methods courses;
2. Training in data management available to researchers across Canada; and
3. Expertise developed to support researchers and to help with their data management needs.

This report’s outlook for Canadian research data stewardship envisioned educating all stakeholders in the research process in RDM best practices and providing researchers with the necessary skills to manage their research data.

### Research Data Canada

The RDSWG was transformed into Research Data Canada (RDC) in 2012. Based on a recommendation from the National Data Summit, this organization was to help achieve the Summit’s vision. The RDC Committee on Education and Training was responsible for implementing the Summit’s training goals. Membership in this committee consisted of national stakeholders from a number of Canadian universities, as well as an international representative from the DCC in the UK.

In January 2014, this Committee released a report with [four key recommendations](#):

- Create an entity similar to the Digital Curation Centre (DCC) in the UK that would be a national resource to all sectors;
- Adopt a cascading training model to maximize the diffusion of knowledge about RDM;
- Implement training requirements for RDM with stakeholder communities; and
- Engage internationally, through participation in initiatives such as the Research Data Alliance.

As a final recommendation, the Committee advised RDC to relegate the planning and implementation of its report to CARL's RDM project, known at the time as ARC.

## Other Discussions on RDM Capacity

In Québec, the “Rapport final du Groupe de travail sur la gestion des données de recherche”<sup>8</sup> made recommendations regarding librarian training, covering a three-year period starting in the fall of 2016. The first planned activity was an “Introduction to Research Data Management” day in November 2016. This is to be followed by a series of five webinars held over the 2017-18 academic year and complemented by a series of five webinars on the following topics: Data Management Planning; Choosing a Repository; Citing Data and Measuring Data Impact; Metadata for Research Data; Ethical Issues, Licenses, Conditions for Sharing and Reuse.

Research Data Canada has produced a [series of webinars](#) that provides training in several areas. The annual regional training by the DLI has undertaken RDM planning and data management training on an ongoing basis for the past few years. Some universities have already developed research data programs and policies for their institutions. Some excellent examples include: the [University of Alberta](#), [Queen's University](#), and the [University of British Columbia](#).

The volume of data produced in Canada indicates some level of RDM capacity. The question of training in this country, however, has been left largely to individual stakeholders. Little evidence exists of formal discussions about training across RDM communities and when it happens locally in different universities, the larger concern over the shared stewardship of research data is not addressed. This latter factor has long plagued the Canadian RDM landscape either because of the absence of shared RDM infrastructure to manage data after projects conclude or because of very limited access to such infrastructure within specific domains. Portage is working to fill such infrastructure gaps through a variety of RDM platforms that collectively cover the

---

<sup>8</sup> The Bureau de coopération interuniversitaire (BCI) were the authors of this report, which was not made public.

data lifecycle. This framework of platforms will result in new training needs as researchers look to develop RDM workflows for their projects.

## How should Portage address RDM training in Canada?

The Portage Training Expert Group has developed the following recommendations to direct Portage's engagement in RDM training in Canada. The priorities of this group are intended to support RDM training across the academic library community without disrupting regional or local training programs. Some jurisdictions are well prepared to offer training while others are not. Portage is available to those needing help, thereby leveling the 'playing field'.

These recommendations have been drawn from this group's discussions and from the evidence provided by the many meetings and reports described above. Foremost, this group seeks to identify immediate training priorities, targeting delivery within three to four months of the release of this paper. We plan to implement our recommendations concurrently with the delivery of training.

**RECOMMENDATION 1:** The Portage Training Expert Group will prepare a set of RDM training principles (see [Appendix 1](#)) to help guide discussions and decisions about training directions and priorities.

Approaches to RDM training and the philosophies behind them vary in Canada. A tension has emerged between the development of resources focused on very specific details of individual RDM requirements by funders and the need to educate stakeholders on the broader aspects of data stewardship. Data stewardship encourages RDM best practices for research projects regardless of the presence of funding or policy mandates. Proper stewardship of research data requires a deeper commitment to data management practices on behalf of all stakeholders.

Such tensions over the objectives of RDM training will undoubtedly be part of the ongoing discussions about training priorities. One approach to guide this dialogue is a set of training principles. The DLI training principles serves as an example. At the outset of DLI, libraries lacked personnel with the skills to handle Statistics Canada's data files, aside from a handful of well-established data services. In 1997, a set of principles was prepared to guide the DLI training program, with the goal of supporting those who manage data services locally. These principles have steered the

successful development of DLI training for more than two decades and have contributed to a national training culture of cooperation, sharing, and mentoring. Similarly, guiding principles can play a critical role in providing clear direction to stakeholders and in helping establish a solid foundation for RDM training in Canada.

**RECOMMENDATION 2:** The Portage Training Expert Group will document current training issues and gaps, will identify priorities, and will work with other Portage Expert Groups to devise and deliver RDM training materials and courses.

Our investigation into RDM training resulted in a thorough review of the major issues and gaps in the current state of this training in Canada. These topics include, but are not limited to:

- Metadata for documenting the research lifecycle;
- Citation and reference practices for research data;
- Ethics as it applies to research data; and
- Publishing and scholarly communication and its relationship to research data.

Building on the established levels of RDM-related knowledge and skills of librarians, issues and gaps in RDM will be clarified and addressed through an overall plan. Close and regular collaboration between Portage Expert Groups will assist in developing targeted, relevant training that will address identified needs and priorities within our community of practice. We see these experiences eventually leading to the establishment of a curriculum for RDM training throughout the educational sector.

**RECOMMENDATION 3:** The Portage Training Expert Group will design a delivery framework to guide Portage in its RDM training activities.

Our environmental scan of the RDM training landscape revealed many forms of delivery, often dependent on the needs of individual stakeholders. Our review showed that training courses differ in style, format, and costing structure. They vary from week-long, highly specialized, in-person, cost-recovery courses to freely accessible online tools. We also noted that discipline-specific training is beginning to happen through related associations. In 2013, Research Data Canada compiled a list of [notable training opportunities in RDM](#).

More recently, collaborative ventures have advanced RDM training to a broader audience internationally. An example of this is a five-day course “[Curating and Managing Research Data for Reuse](#)” developed by the University of Michigan’s ICPSR and UK Data Archive. Mainly targeted at data service providers, this course is offered in-person at a cost. Another example is the online [MANTRA](#) course developed by the University of Edinburgh. This free, non-credit course is targeted at postgraduate students, early career researchers, and information professionals. The design allows for self-paced progress through a series of tutorials. A spin-off version of MANTRA has been produced through an international collaboration with the University of North Carolina as an [online MOOC course](#) available in two streams: certificate or non-certificate training.

The group’s discussions on the variety of delivery forms has resulted in a draft framework to guide the planning of training modules. We intend to apply this framework with our early training activities and to modify it subsequently based on our experiences.

**RECOMMENDATION 4:** Knowledge of the audience for which training is being conducted is essential to a successful experience for the learner and the instructor. The Portage Training Expert Group will work with the Portage Research Intelligence Expert Group to identify specific audiences for which training materials will be prepared and training delivered.

Understanding the trainee is essential to establish credibility and relevance with those being trained. Generic training materials will provide a base for content coverage, but such content needs to be tailored to meet specific stakeholder training priorities. A ‘one-size-fits-all approach’ will be avoided. Audiences will have content delivered that will be relevant to their specific requirements.

Training librarians and data stewards is the top priority for Portage. However, we will also work with other stakeholders to prepare training materials and to deliver courses to those identified on the Portage website as critical partners:

- Researchers;
- Service and infrastructure providers; and
- Policy makers and administrators.

It must be remembered that reaching out to all stakeholders is essential to build an RDM culture in Canada.

**RECOMMENDATION 5:** The Portage Training Expert Group will prepare a model Memorandum of Understanding (MOU) to help establish successful partnerships in developing RDM training materials. These agreements will describe basic terms for collaboration and will promote common interests that intersect the training objectives of other stakeholders, nationally and internationally.

Several models exist to guide the development of an RDM training MOU. Portage member institutions have a long history of collaborative ventures and the recent agreement between CARL and COPPUL regarding mutual interests in Portage will prove useful. Other successful collaborations within RDM will also be helpful, including international examples provided by the ICPSR and UK Data Archive.

Involving multiple stakeholders across the country in RDM training will help reinforce the importance of RDM in all stages of the research lifecycle, as well as generate buy-in from these stakeholder groups. Pulling in expertise from different stakeholder groups significantly enhances the ability of Portage to develop and deliver comprehensive programs that address a variety of RDM training needs across the Canadian research data landscape.

## Summary

This White Paper provides a high-level perspective on RDM training for Portage. RDM developments in Canada have lagged behind some of the countries typically considered to be our peers, such as the United Kingdom and the United States. This was evident in our environmental scan of the different training activities being developed and offered. Some excellent international training modules are available to Canadian stakeholders but without Canadian-specific content. The Portage website provides an opportunity to prepare and disseminate materials rich in Canadian RDM content.

RDM expertise already exists in Canada. However, this expertise remains largely siloed in specific disciplines and jurisdictions. Training resources need to be organized collaboratively across these divisions to capitalize on the knowledge and resources of these stakeholder communities. While our overview of RDM training is not exhaustive, it does provide a robust representation of the current landscape. It is also imperative that in building a foundation for RDM expertise, a national research data culture is also cultivated that represents the underlying principles and values of such expertise in Canada.

Our recommendations are directed at preparing the Portage Training Expert Group for an impending demand for RDM training from within the library community as it increases engagement with researchers and other RDM stakeholders. We see these recommendations guiding our Training Expert Group both in high-level, ‘blue-sky’ discussions and in dealing with more practical issues related to RDM training.

We envision implementing our recommendations over a two-year period. Some activities will have a higher priority and will happen within the next three to four months. An example is the delivery of Portage and RDM information sessions across the country. Other activities, such as Canadian-specific training modules, will take longer to develop.

The Portage RDM Training Expert Group is prepared to close the gap between Canada and those countries leading in RDM training. Through collaboration with the other Portage Expert Groups, the library community, other Canadian RDM stakeholders (including the research community), and our international counterparts, we are poised to make this leap forward.

## **Acknowledgements**

We would like to thank Eugene Barsky (University of British Columbia), Jay Brodeur (McMaster University), Natalie Clairoux (Université de Montréal), Myrian Grondin (Université de Montréal), Susan Haigh (CARL), Chuck Humphrey (Portage), Steve Marks (University of Toronto), Jeff Moon (Queen’s University) and Martha Whitehead (Queen’s University) for their peer review of our paper and for providing invaluable feedback on its content and style.

## Appendix 1: RDM Training Principles for the Training Expert Group

1. We will respect the training missions of the four regional library consortia<sup>9</sup>.
2. We will work closely with the four regional library consortia
  - To help identify training requirements
  - To support training programs
3. We will work with other RDM stakeholders in collaborative training initiatives:
  - Researchers;
  - Librarians and data stewards;
  - Service and infrastructure providers; and
  - Policy makers and administrators
4. Training will be offered at different levels of expertise to and by the library community.
5. Training materials will be supportive of local campus training initiatives.
6. Audience-specific training is fundamental.
7. A variety of training modes and technologies are open for consideration and dependent on the needs of the groups being trained.
8. All training materials will respect both official languages.

---

<sup>9</sup> CAUL (Council of Atlantic University Libraries); BCI (Bureau de coopération interuniversitaire); OCUL (Ontario Council of University Libraries); and COPPUL (Council of Prairie and Pacific University Libraries)