Data Management
Position Paper

For Innovation, Science, and Economic Development Canada

2017-08-31
# Table of Contents

1. Executive Summary ........................................................................................................ 1

2. Methodology .................................................................................................................. 5
   2.1 How did we undertake our work? ................................................................................ 5
   2.2 How did the Working Group identify user need? ....................................................... 6
   2.3 What are data? ........................................................................................................... 8

3. Research Data Management ............................................................................................ 8
   3.1 What is Research Data Management (RDM)? ........................................................... 8
      3.1.1 Why is RDM important and how does it lead to greater research impact in Canada? 9
      3.1.2 Who is involved? ............................................................................................... 10
      3.1.3 How is RDM used? ........................................................................................... 13
      3.1.4 What are the core RDM functions required to meet the needs of Canada’s research community and administrators? ................................................................. 14
   3.2 Current State ............................................................................................................. 15
      3.2.1 Strengths and Challenges .................................................................................... 15
      3.2.2 How is RDM managed in other international jurisdictions? ............................... 21
   3.3 Future State ............................................................................................................... 25
      3.3.1 The Future We Seek .......................................................................................... 25
      3.3.2 Vision ............................................................................................................... 26
      3.3.3 Principles ......................................................................................................... 26
      3.3.4 Goals ............................................................................................................... 27

3.4 National Coordination and Facilitation ........................................................................ 28
   3.4.1 Why is it important to have a national approach to coordinating and facilitating RDM? ......................................................................................................................... 28
   3.4.2 How could the Government of Canada support the national coordination and facilitation of RDM? ................................................................. 28
3.4.3 How would RDMC undertake its work? ................................................................. 29
3.4.4 How would RDMC be accountable to the research community? ......................... 30
3.4.5 How would RDMC deliver on the core RDM functions? .................................... 30
3.4.6 What level of federal investment is being proposed to fund RDMC? .................... 36

3.5 Key Success Factors ................................................................................................. 37

4. Research Management Information ........................................................................ 38

4.1 Why is RMI important and how does it lead to greater research impact in Canada? .... 39
4.2 The Vision for RMI in Canada .................................................................................. 41
4.3 The Future We Seek ................................................................................................. 41
4.4 Principles .................................................................................................................. 41
4.5 Goals ....................................................................................................................... 41
4.6 Building and Sustaining RMI infrastructure ............................................................ 42

5. Conclusion .................................................................................................................. 43

Appendix A: Core Functions Based Upon Data-Related Activities ............................... 45
Appendix B: RDC Principles ......................................................................................... 50
1. Executive Summary

In November 2016, the federal department of Innovation, Science, and Economic Development (ISED) provided funding to the Leadership Council for Digital Research Infrastructure (LCDRI) to undertake a number of key initiatives, including the development of a position paper on data management. As a result, in early January 2017, a Working Group on Data Management (DM) was established to undertake this work. This paper was prepared with feedback from and consultation with the larger DRI community, and is based on a thorough consideration of the needs for DM in Canada’s academic research community.

Digital infrastructure, driven by new information and communication technologies such as the Internet, is transforming our lives. It enables us to tap new areas of human creativity in ways that are not only changing how we relate to one another, find information, make decisions, and manage and use our time, but are also allowing us to alter our approaches to identifying and solving challenges in our schools, our businesses, and society more generally. We are only beginning to understand its impact and to harness its potential.

Similarly, digital research infrastructure is transforming the practice of research and enabling the rapid creation of massive quantities of data at an explosive rate. These changes are fundamentally reshaping the speed at which our researchers work, the questions they can ask, and the results they can achieve. To realize the amazing social and economic potential that new digital research infrastructure and data offer to Canadians, and to ensure that we are among the first to mine the benefits of this important resource, a strong and vibrant digital research infrastructure (DRI) ecosystem must be in place.

The ability of Canadian researchers to store, access, reuse, and build upon digital research data is essential to their ongoing capacity to remain current and collaborative in their fields, enabling them to generate and contribute critical research that underpins Canada’s economic and social well-being. Accessible digital research data also holds great potential benefits for the Canadian non-profit and private sectors in helping them advance critical social and commercialization goals.

The systematic management of this research data is emerging as an essential component of the DRI ecosystem, and Canada has just begun to lay the foundations that are needed to ensure that the necessary DM tools and resources are in place to support researchers in exploiting the benefits that are available to them through the delivery of high quality, equitable, and effective DM services and platforms. This report identifies key challenges to achieving this goal and suggests how federal investment in DM would be cost-effective in advancing the critical research that will help to drive Canada's future economic and social prosperity.

Although this report deals primarily research data management (RDM), it also discusses the importance of another, if somewhat invisible, aspect of research-related digital activity and its associated infrastructure that is increasingly critical to an effective research ecosystem: research management information (RMI).

What is RDM?

RDM encompasses activities related to data policies, data planning, data element standardization, information management control, data synchronization, data sharing, and database development, including practices and projects that acquire, control, protect, deliver...
and enhance the value of data and information. These activities can be organized into a set of five core functions: policies; standards and protocols; processes and procedures; leadership, advice, support, and training; and tools and platforms.

The collective efforts of organizations, the federal government, individual universities, discipline-specific research organizations, and researchers themselves have resulted in establishing significant strengths within Canada’s RDM community, providing important foundations on which to build for the future, including partnerships with other stakeholders that can be leveraged to develop collaborative new RDM tools and services. In addition, some universities are developing training and offering discipline-specific, as well as general, RDM advice and support directly to researchers on their campuses.

The working group also identified a number of challenges presented by the current state of RDM in Canada, related to:

- Coordination and coherent sector-wide planning,
- Clarity on roles and responsibilities,
- Common policies, standards, and protocols,
- Skills, knowledge, and training,
- Attraction and retention of highly qualified personnel,
- Academic publisher’ interests in data repositories,
- Culture,
- Funding,
- Tools and platforms, and
- Organizational and researcher readiness for new RDM policy.

Our vision for the RDM component of the DRI ecosystem is that of an innovative and coordinated research data management community, providing responsive services and resources that support Canada’s research community in advancing research for the betterment of all Canadians.

Three broad goals are suggested with a view to achieving the vision.

- Build innovative services and resources that are distributed across universities, nationally coordinated, internationally recognized, and sustainable, while being responsive across the full spectrum of researcher needs and disciplines. These services and resources should respect researcher, discipline-specific, national, and university data stewardship policies, and be based on best-practice standards and protocols.
- Advance and adopt RDM processes and procedures that are informed by researcher, university, and discipline-specific needs, to improve the overall quality of research data and to advance best practices. This will require flexible and adaptive tools and platforms supporting data planning, creation, curation, deposit, access, discovery, and reuse.
- Establish a community of practice that is supported by a distributed network of specialists who can provide expert advice, support, and training in RDM best practices to researchers.

These vision and goals lead the first key recommendation in this position paper:

In order to ensure that all Canadian researchers have access to high quality, equitable, and efficient RDM services and platforms, we are proposing that the Government of Canada consider investment in a national body that would
coordinate and facilitate key RDM activities that are most effectively undertaken at the national level in Canada. This new body, **Research Data Management Canada** (RDMC), would build on the critical foundations laid by Portage and RDC. Using a federated service model, it would support and leverage the RDM practices and approaches of universities and discipline-specific communities across the country, while also respecting their independence and autonomy in providing services and platforms to their researchers. It would also establish and maintain innovative, coordinated, and interoperable RDM services, tools, and platforms that would be accessible to administrators and researchers in all Canadian universities, and supported through a community of practice and a network of experts.

LCDRI is proposing a budget of $54 million over 5 years to support the work of RDMC. The proposed budget has been broken into three main cost categories (highly qualified RDM personnel, tools and platforms, and a secretariat office), reflecting the goals and approach proposed. Highly qualified personnel includes secretariat staff, staff related to the network of experts, and contracted experts. Tools and platforms includes software tools, data repositories, and archival storage, including the costs of external service contracts for their operation and maintenance. The operational budget of the secretariat includes operational costs such as IT, travel, and office rental and set-up. The majority of the investment that is required to support RDM in Canada would continue to be made by regional consortia and individual universities. The federal contribution to RDMC that is proposed below would be used to leverage these investments and provide a critical point of coordination for the development of nationally shared RDM policies, standards, protocols, processes, expertise, tools, and platforms.

**What is RMI?**

RMI deals with the activities and workflows associated with effective management of all aspects of the research lifecycle, including applications for research support, research evaluation, managing research funds, research reporting, research dissemination, monitoring compliance, and outcomes and impact assessment.

The practice of research management involves a wide spectrum of individuals, disciplines, jurisdictions (national, regional, international), organizations (research-performing, research-funding, research-publishing) and functional units within organizations (e.g. finance, research services, library). It is an increasingly important and distributed activity for ensuring the smooth, efficient and measurable functioning of the research lifecycle. Coordinated action among these diverse and independent stakeholders is required to enhance the capacity for transparent development and maintenance of open standards, to improve interoperability of systems and platforms, to strengthen international linkages, and to promote innovative implementations of RMI.

Our vision for the RMI component of the DRI ecosystem is for responsive services, founded on high-quality and comparable research management information, that enable Canada’s research and research management communities to optimize resources and productivity, enhancing the Canadian research enterprise.

This vision led to the second key recommendation position in this position paper:
LCDRI is proposing an additional funding envelope of $700,000 over 5 years to support efforts related to RMI, in addition to the budget for RDMC.
2. Methodology

2.1 How did we undertake our work?

In February 2016, the Minister of Science, the Honourable Kirsty Duncan, met with members of the Leadership Council for Digital Research Infrastructure (LCDRI) and asked them to affirm their intention to work collaboratively to ensure that Canada has a strong digital research infrastructure (DRI) ecosystem in place to support the critical research undertaken by Canada’s academic research community. In March 2016, the LCDRI responded to Minister Duncan’s request with a letter that was signed by all LCDRI members, confirming their commitment to work together.

In November 2016, the federal department of Innovation, Science, and Economic Development (ISED) provided funding to the LCDRI to establish a short-term secretariat with full-time staffing support and tasked the group with undertaking the following:

- development of a position paper on research data management;
- development of a position paper on advanced research computing (ARC); and,
- consideration and recommendation of potential approaches for coordination.

In early January 2017, a Working Group on Data Management was struck to develop the required position paper on data management. Over the last eight months, the Working Group has met for six full-day meetings. There were also numerous virtual meetings and discussions among and between members to develop specific pieces of work that were then vetted through the full Working Group.

The members of the Working Group are: David Baker, Executive Director, Consortia Advancing Standards in Research Administration Information (CASRAI); Susan Haigh, Executive Director, Canadian Association of Research Libraries (CARL); Chuck Humphrey, Director, Portage Network; Mark Leggott, Executive Director, Research Data Canada; Martha Whitehead, current and now past President, CARL. In May 2016, the Working Group also welcomed Donna Bourne-Tyson to its midst when she became president of CARL. Meeting organization and facilitation were overseen by the LCDRI secretariat.

The Working Group shared and sought feedback from the DRI community on its work as it evolved. The two major consultations with the broader DRI community took place as part of a workshop with all members of LCDRI on February 28, 2017 and at the LCDRI Summit on June 27, 2017.

![Figure 1. Face-to-Face Meetings of the Data Management Working Group](image)
2.2 How did the Working Group identify user need?

Having a clear understanding of the RDM needs of researchers and administrators who are responsible for maintaining good public stewardship of the data was clearly central to the Working Group’s ability to address its mandate of articulating how best to support excellence in RDM practices in Canada. In order to properly articulate these needs and to ensure that nothing was missed, the Working Group identified the key data-related activities/responsibilities of researchers and administrators. These activities are shown in Figure 2. The activities in the centre of the circle (secure, discover, document and curate, and store) occur repeatedly throughout the research process. The activities in the outer ring (plan, create, process, analyze, disseminate, preserve, and reuse) are data-related activities associated with a particular part of the research process and are often iterative.

The Working Group developed descriptions to clarify what specific research actions take place during each of the data-related activities. These descriptions are as follows.

**Store:** writing (accessing/placing/putting) and reading (retrieving) digital content on a variety of physical media. Content is used for different purposes across the lifecycle, including copies for active use, archival use, and dissemination. Active storage enables the use of digital content within research currently being conducted. Archival storage enables long-term access and preservation by protecting the digital integrity of content for current and future uses.

**Secure:** protecting, controlling, and complying with legal and ethical conditions on the use of the data.

*Figure 2. Data-Related Activities During the Research Process*
**Document and Curate:** describing, explaining, and communicating the context in which the data exists, the workflow of the data, and the technical details of the data (the metadata). Employing metadata standards provides structure and form in documenting data and enables the interoperable use of the documentation and the data it describes. There are many levels of metadata that can be articulated (e.g., study level description, sample description, variable description, instrument description, etc.), describing, identifying, and explaining the data for discovery purposes.

**Discover:** searching, finding, mobilizing, locating, interpreting, assessing, and visualizing metadata, and retrieving research data.

**Plan:** identifying resources, expertise, and services required to develop, manage, and share high-quality data; planning for resource, time, and cost management; preparing/arranging/identifying/planning for the deposit of data so that it is discoverable in a data repository and reusable; developing and implementing policies at universities, and reviewing policies for compliance at project level; determining data management requirements from the research design; identifying data management practices and workflow; planning consent for sharing.

**Create:** identifying, acquiring and creating (i.e., experimenting, observing, measuring) data for the research; generating data (e.g., simulating) and corresponding metadata; consulting policies that structure or define procedures for collecting data, such as ethics procedures and approval.

**Process:** preparing research data for analysis; checking, validating, cleaning, describing, transforming, aggregating, manipulating, reducing, anonymizing; describing workflow and processing; consulting expertise in data wrangling and statistics, as well as policy and procedures in domain-specific instances.

**Analyze:** describing, comparing, interpreting, and modelling patterns within the data; deriving variables/data; consulting expertise for statistical research methodologies and domains; as well as policies and procedures in domain practices.

**Disseminate:** sharing the data; transferring data from project to repository; consulting policy for data deposit agreements, licenses, access (conditions for reuse), and preservation; promoting and discovering; consulting expertise in data curation, submission information package management (i.e., formats, identification of files, manifest that identifies types of files); ensuring metadata is developed robustly.

**Preserve:** preparing, enhancing, and storing data, metadata, documentation, and code for long term access and reuse; migrating data to the best formats and through suitable media; storing data and back-up; applying best practices for digital preservation processing (i.e., characterization and normalization); consulting policies to apply to a trusted digital repository (i.e., format, selection, etc.), and expertise for digitization, digital preservation, and metadata.

**Reuse:** combining existing datasets to create new data for research and analysis; ensuring data discovery, access, and reuse of existing data (secondary data analysis); using resources to perform reuse; consulting policies concerning attribution, provenance, and licensing, as well as expertise on data wrangling, search skills, and secondary analysis skills.
2.3 What are data?

Data\(^1\) are facts or observations captured with a minimum of contextual interpretation. Data may be in any format or medium (analog or digital). Data elements can include text (and other symbols), numbers, images (and other graphical representations), video or audio. Data includes metadata.

There are two types of data pertaining to the university research environment that are the focus of this data management position paper: research data and research management information.

Research Data are used as primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity, and as evidence in the research process and/or are commonly accepted in the research community as necessary to validate research findings and results. All other digital and non-digital content have the potential of becoming research data. Research data include metadata and may be experimental data, observational data, operational data, third-party data, public-sector data, monitoring data, processed data, or repurposed data.

Research Management Information (RMI) is information used primarily to facilitate research management by research-funding and research-performing organizations. Examples include information about the people, organizations, funding, equipment, projects, outputs, outcomes and impacts of the research lifecycle. RMI include metadata recorded with the information, such as the version number of a classification scheme used to classify the expertise of a Principal Investigator or a specific project. Common synonyms for RMI include: admin data, research administration data, research information, and research documentation.

3. Research Data Management

3.1 What is Research Data Management (RDM)?

Research Data Management is activities related to data policies, data planning, data element standardization, information management control, data synchronization, data sharing, and database development, including practices and projects that acquire, control, protect, deliver and enhance the value of data and information.

To be effective, these activities must be coordinated and shared not only among researchers in specific universities, but also among researchers across Canada and throughout the world. Highly trained experts in the field who are able to help researchers navigate the complexities of research data management are essential to this process.

A second essential element is a network of data repositories that support the curation and processing of data, so that current and future generations can find, access, reuse, and manage

---

\(^1\) These definitions were developed through consultation with the CASRAI IRiDuM Glossary and the Data Management (DM) Working Group (WG). They were then shared for input from LCDRI members. There is a wide and varied understanding of the terms, which can create confusion when working in a diverse and multi-disciplinary community such as Canada’s academic community. The group did its best to be inclusive and understandable in the definitions, in order to help the community to be successful in using the same shared definitions going forward.
it. These data repositories must then be linked to an archival storage network that ensures that data deemed worthy of preservation is stored properly for the long-term. All of the data repositories and archival storage platforms that are included in these networks are housed across a diverse community of domestic and international organizations/institutions, which makes ensuring their interoperability challenging, but critical.

3.1.1 Why is RDM important and how does it lead to greater research impact in Canada?

The ability of Canadian researchers to find, access, reuse, and manage data is essential to their ongoing capacity to remain current, competitive, and collaborative in their fields, both at home and internationally. It enables them to generate and contribute critical research that underpins Canada’s economic and social well-being in key areas such as advanced manufacturing, agri-food, clean technology, digital industries, health/bio-sciences and clean resources, as well as artificial intelligence and machine learning.

The Tri-Agency Statement of Principles on Digital Data Management underscores the importance of effective research data management in support of this goal:

Research data are gathered through a variety of methods, including experimentation, analysis, sampling and repurposing of existing data. They are increasingly produced or translated into digital formats. When properly managed and responsibly shared, these digital resources enable researchers to ask new questions, pursue novel research programs, test alternative hypotheses, deploy innovative methodologies and collaborate across geographic and disciplinary boundaries. The ability to store, access, reuse and build upon digital research data has become critical to the advancement of science and scholarship, supports innovative solutions to economic and social challenges, and holds tremendous potential for Canada’s productivity, competitiveness and quality of life.²

The true impact and outcomes of a world in which researchers collaborate across domestic and international boundaries to build almost effortlessly on existing data are, in fact, not yet fully understood. In the same way that the Internet has shifted how we conceive of problems and relate to one another, information and communication technologies have the potential to shift fundamentally how we undertake research and the critical innovation that this research may yield. Effective and efficient research data management practices that are sustained, coherent, and coordinated are essential to unlocking this potential. In order for us to be able to unleash this new world of possibility and maximize Canada’s return on its investment in our research community, the systems, policies, tools, and platforms must be in place for researchers across disciplines to find and use data easily.

---

3.1.2 Who is involved?

RDM is a distributed activity involving many actors, including researchers, librarians, university administrators, professionals in discipline-specific and non-profit organizations, and public servants in government, all having roles in its delivery. Given the number of players involved, the landscape can be challenging to navigate. One approach is by jurisdiction: local, regional, national, and international. Each level presents some aspect of support or enablement to researchers, but often in different or critical ways.

Local

A number of universities across Canada have begun to invest in campus RDM service delivery. The number of universities providing these services and the scale of investment required to support them are expected to increase significantly as researcher needs change and grow, and as new national and international requirements for research data management are articulated. There are a number of offices and individuals at the university level who already have or will have responsibility for supporting this effort. They include:

a) Libraries (i.e., librarians/data librarians, academic IT personnel, data management committees);
b) Research Offices (i.e., vice-presidents research, research office staff, university research policy committees);
c) Information Technology Departments (IT) or Computing Services (i.e. Chief Information Officers (CIO), university data governance committees); and
d) Research Ethics Boards.

Researchers themselves are also often engaged actively in supporting RDM through domain-specific communities of practice and professional societies.

Provincial/Regional

Academic libraries have a long tradition of working collectively to provide access to research collections, resources, and services. They have built communities of practice supporting inter-university collaboration and have established four regional organizations covering all provinces and territories to coordinate and manage the delivery of shared services, including RDM. These are the Bureau de Coopération Interuniversitaire (BCI), the Ontario Consortium of University Libraries (OCUL), the Council of Prairie and Pacific University Libraries (COPPUL), and the Council of Atlantic University Libraries (CAUL). This regional base strengthens individual libraries through a network that acts on behalf of their collective interests and that produces greater returns than any single library is able to achieve on its own. Together, the consortia also ensure representation of Canada’s geographic, linguistic, cultural, and jurisdictional diversity, which is critical to ensuring that RDM support is provided to researchers across Canada in ways that are appropriate and easily accessible.

There are a number of other existing provincial/regional organizations that have an impact on or role in the delivery of research data management policies and services. These organizations include:

a) Provincial Research Funding Agencies/Ministries (e.g., Ontario Ministry of Research and Innovation, les Fonds de recherche du Québec, le Ministère de l’Économie, de la Science et de l’Innovation du Québec, Nova Scotia Research & Innovation Trust, Innovate NL, New Brunswick Innovation Foundation, BC Knowledge Development
Fund, Alberta Economic Development & Trade, Innovation Saskatchewan, and Research Manitoba); and,

b) Regional Research and Education Networks (i.e., ACORN-NL (Newfoundland), ACORN-NS (Nova Scotia), Aurora College (Northwest Territories); BCNet (British Columbia), Cybera (Alberta), MRnet (Manitoba), ECN (New Brunswick and PEI), ORION (Ontario), RISQ (Quebec), SRnet (Saskatchewan), and Yukon College (Yukon)).

National

For researchers to be able to find, access, reuse, and manage their data effectively, it is essential that policies, processes, protocols, standards, metadata, and interoperable preservation storage are shared among a distributed landscape of individuals and organizations. Coordination of this work at the national level is critical. In order to ensure that a researcher in BC is able find, access, and potentially reuse the data of a researcher in Toronto, Halifax, or the Yukon, nationally agreed-upon frameworks and practices must be in place. The same is true of a researcher in Montreal who would like to find and reuse data that has been produced in Berlin or Tel Aviv. The effort required to ensure that these critical RDM foundations are in place goes well beyond the capabilities of a single university or region in Canada.

The Canadian research community has recognized this need, and two community-led groups have taken a leadership role in beginning to facilitate and coordinate Canada’s data management-related activities at the national level. These are:

a) Canadian Association of Research Libraries (CARL). In 2015, CARL launched the Portage Network to contribute to the shared stewardship of research data and to address specific gaps in the national research data-management infrastructure in Canada. The aim of Portage is to coordinate and expand existing expertise, services, tools, and platforms so that all academic researchers in Canada have access to the support they need for research data management.

b) Research Data Canada (RDC). Established in 2011 following a recommendation in the 2011 Canadian Research Data Summit Report, “Mapping the Data Landscape,” RDC is a stakeholder-driven and supported organization dedicated to improving the management of research data in Canada. Since 2016, RDC has been funded by CANARIE.

The federal government has also recognized the importance of coordinated, national approaches to RDM, and has taken a leadership role in the area as well. In November 2016, the department of Industry, Science, and Economic Development asked the LCDRI to include as a priority in its work the development of a position statement on data management, which was welcomed and greatly appreciated by the DRI community. The June 2016 Tri-Agency Statement of Principles on Digital Data Management is an illustration of federal engagement in setting RDM policy principles for federal research grants. Recent Open Science and Open Data initiatives also have an impact on the general environment in which Canadian researchers work.

---

3 An example of a regional research and education network supporting RDM is Cybera’s work with Portage and Compute Canada to find a sustainable delivery model for Jupyter Notebooks, which is a popular open source web application that allows researchers to create and share documents that contain live code, equations, visualizations, and explanatory text.
There are a number of other organizations and associations that are engaged in supporting the delivery of research data management services in Canada.

a) University-Based Organizations
   • Research Institutions (e.g., Marine Environmental Observation Prediction and Response Network (MEOPAR), CBrain (McGill Neurological Institute), Ocean Networks Canada (ONC), Polar Data Centre),
   • University Consortia/Organizations (e.g., Universities Canada; U15, Canadian University Council of Chief Information Officers (CUCCIO); Canadian Association of Research Libraries (CARL); Canadian Association of Research Administrators (CARA); Canadian Association of Research Ethics Boards (CAREB); Canadian Association for Graduate Studies (CAGS); Canadian Research Knowledge Network (CRKN);

b) Discipline-Specific Communities of Practice
   • Multi-disciplinary (e.g., Federation for the Humanities and Social Sciences, Royal Society of Canada), and
   • Single-domain (e.g., Canadian Society of Microbiologists, Canadian Association of Physicists).

c) Other Organizations (e.g., First Nations Information Governance Centre (FNIGC))

International

For Canadian researchers to leverage the opportunities presented by data generated and collected by their peers elsewhere in the world, and to ensure that they remain sought after internationally as collaborators, Canadian research data management practices must coordinate with global practices across all disciplines. There are a number of international organizations that have an impact on or role in the delivery of research data management policies and services. They include:

a) Government Organizations (e.g., the Organization for Economic Co-operation and Development (OECD), the Global Science Forum, the European Commission, the Belmont Forum)

b) Communities of Practice
   • RDM-focused (e.g. CODATA, Consortia Advancing Standards in Research Administration Information (CASRAI), Research Data Alliance (RDA), EUDAT, Netherlands Institute for Permanent Access to Digital Research Resources (DANS), Australian National Data Service (ANDS)); and,
   • Domain communities of practice (e.g., multi-disciplinary - Food and Agriculture Organization of the UN (FAO), Agricultural Information Management Standards) and single-domain (e.g., the International Brain Research Organization).

Other Contributors

The private sector is also having an impact on research data management. For instance, research organizations are increasingly deploying commercial cloud storage and computing resources.

Engaging the private sector in the conversation about the delivery of RDM will be important, as it could be a key partner in the delivery of sustainable and cost-effective services. At the same time, universities and researchers will need to ensure that proper safeguards are in place to protect their ownership and primary stewardship roles for the data that they generate.
Journal publishers such as Elsevier, IEEE, Nature, etc., are another group that have an impact on RDM. Journal publishers play a substantial role in directing and influencing researchers’ approaches to RDM, both through their official publishing policies, and their communities of practice. Nature, which publishes one of the world’s most influential suite of scholarly journals, has a journal dedicated to the publication of data and data articles, and maintains an extensive list of domain repositories where it recommends data be deposited.¹ Journal publishers also increasingly require scholars to deposit data within a 6–12-month period after publication of an article. The Public Library of Science (PLOS), which, like Nature, publishes a series of high-impact journals, maintains a list of recommended data repositories and “require(s) authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception.”⁵ PLOS goes even further in their policy statement: “Refusal to share data and related metadata and methods in accordance with this policy will be grounds for rejection.” This approach is no longer the exception, but is rapidly becoming the norm in scholarly publishing. The overarching rationale for this requirement is simply to make research better.

### 3.1.3 How is RDM used?

As described elsewhere in this paper, RDM is critical to supporting researchers’ ability to find, access, reuse data that has been generated by others. But, RDM also clearly plays an important role in helping principal investigators (PIs) manage their own data, which can be especially challenging in a large research lab where graduate students and collaborators come and go regularly. RDM helps ensure the protection of data during a research project as well as over longer stretches of time, helping PIs meet the increasingly stringent requirements of good research ethics and reproducibility. The principles of good RDM apply equally to the local research context and can only benefit the individual researcher, their team, and the larger community. RDM is even more important when the research involves large volumes of data. Below are two researchers’ profiles, providing examples of how research data management is used in their work.

**Susan Brown: Harnessing the power of computers for critical literacy research**

*The Orlando Project* is a digital history of women’s writing in the British Isles. It is also an experiment in the integration of text and technology. The project contains details of more than 1,300 writers’ careers, including how writers were received, relationships with publishers, intellectual influences, friends, political activities and even illnesses.

University of Guelph digital literary historian Susan Brown, along with two colleagues and more than a hundred collaborators, used advanced research computing to realize the Orlando Project so that data can be found, sifted and reordered according to researchers’ priorities. They can, for example, see shifts in society’s understanding of ethnic or racial identity, or the relationship between childbearing and the ability to publish.

"A lot of humanities scholars are still using Microsoft Word or spreadsheets to track the stages involved in this kind of work,” says Brown. “It might be the reason we don’t see a lot of large scale collaborations like the Orlando Project.”

---


Still, the project, updated twice a year with new discoveries, writing and authors, is now being used as a model for other works of digital scholarship, The Canadian Writing Research Collaboratory among them. Brown and her team created it as a virtual research environment for scholars of our country’s literary and cultural history.

The key to this kind of collaborative research, says Brown, is a robust network for moving large amounts of data around, and 24/7 server access that lets researchers around the world shape the material on their schedule. RDM that tracks where information came from and what has been done to it, and data preservation are essential so later generations of scholars can access the material.

"The future of our discipline depends on figuring out ways of enabling new kinds of research digitally," says Brown.

**Julie Friddell: Don’t put this data on ice**

Julie Friddell describes herself as a steward for accelerating scientific discovery about the cryosphere. It’s her job to make sure data from Canadian and international scientists studying the frozen parts of our world are widely accessible—today and for generations to come.

Friddell is director of the Canadian Cryospheric Information Network and Polar Data Catalogue, based at the University of Waterloo. She and her team of IT professionals run an online database of Arctic and Antarctic data and metadata.

A metadata entry about a study of caribou in the Arctic, for example, might include information about what kinds of indicators of health were collected and by whom, over what time period, from how many animals, and at what locations. It would also provide a link to the university servers where the research findings are stored.

“It gives users an idea of whether the data will be useful to them and whether they want to go to the trouble of downloading it,” says Friddell.

With almost 2,600 metadata records representing over 2.8 million individual data files, the catalogue takes up a lot of storage space. Making sure all that information is secure and reflects the latest storage protocols is a big job. Friddell and her staff make regular updates so their database is always accessible and secure from hackers.

“We have a societal obligation to hold and take care of data that come from publicly-funded research," says Friddell. “It’s not good enough for it to just sit on someone’s laptop. We need to make sure raw data are openly available so others can use them and make new discoveries.”

As the benefits of research data management become clearer there will be a significant uptake by researchers to undertake this important cultural change.

**3.1.4 What are the core RDM functions required to meet the needs of Canada’s research community and administrators?**

Once the data needs/responsibilities of researchers and administrators were explored and articulated, the Working Group turned its attention to identifying the core research data management functions to support these needs/responsibilities. Using the diagram of key data-
related activities as their starting point, the Working Group mapped key RDM functions to each of the activities.

As the group worked, it became clear that the core RDM functions could be structured into five key areas:

- Policies,
- Standards and Protocols,
- Processes and Procedures,
- Leadership, Advice, Support, and Training, and
- Tools and Platforms.

The Working Group then reviewed how each core function would be required to support the data-related activities/responsibilities of researchers and administrators throughout the research process. The results of this exercise can be found in Appendix A, which provides an overview of the RDM resources and services (specifically or generically titled) that are, or should be, available to the research community as part of delivering the core functions.

It is important to emphasize that the responsibility for delivering each of the core functions is shared across a highly distributed RDM community. Once the core RDM-related functions were articulated, the Working Group used this information to discuss and identify jurisdictional roles and responsibilities within the complex landscape of players.

This process of defining core RDM functions, as well as understanding the roles and responsibilities of the various players at home and abroad was incredibly valuable in helping the Working Group reach the conclusions presented at the end of this paper.

### 3.2 Current State

#### 3.2.1 Strengths and Challenges

The current Canadian RDM landscape is dynamic and evolving through the leadership of the many stakeholders. Although no formal coordination structures currently exist among these groups, there is a significant amount of goodwill and commitment to working collaboratively to advance excellence in RDM, in the best interest of Canadians and researchers across all disciplines.

The collective efforts of organizations such as CARL/Portage and RDC, the federal government, individual universities, discipline-specific research organizations, and researchers themselves have resulted in establishing significant strengths within Canada’s RDM community, providing important foundations on which to build for the future. For example, a number of research libraries have developed significant RDM expertise that can be shared with others. They have also created partnerships with other stakeholders that can be leveraged to develop collaborative new RDM tools and services. In addition, they are developing training and offering discipline-specific, as well as general, RDM advice and support directly to researchers on their campuses.

These RDM leaders have also helped to increase awareness of the importance of RDM significantly among university administrators, researchers, and funders. As mentioned previously, the federal government’s *Tri-Agency Statement of Principles on Digital Data Management*, as well as its recent commitments to Open Science and Open Government have
assisted in raising the profile of RDM in Canada significantly. International RDM policies and requirements have also helped to drive new understanding of the importance of RDM, as well as the adoption of new RDM practices among researchers and universities.

The leadership efforts of Portage and RDC, two nationally-based organizations with RDM-specific mandates, have been particularly important in putting in place the more formalized building blocks needed to advance excellence in RDM and to set Canadians and Canadian researchers up for success in the future.

Since its inception in 2015, Portage has built, through leadership investments and in-kind contributions on the part of CARL members, a solid foundation for national RDM services and platforms, and has helped raise the national profile of RDM more generally. In particular, it has contributed:

- A growing and distributed network of experts who provide RDM support to academic librarians and researchers across Canada, currently 87 contributing members from 35 universities.
- A prototype for a critical federated research data repository platform, that will help researchers and universities curate and process their data to ensure others can find, access and reuse it in the future. The repository was developed in partnership with Compute Canada and is currently in beta testing. It is on schedule for limited production in September 2017 and production launch in April 2018.
- A growing national research community to which it provides services and which has been facilitated by memoranda of understanding signed between Portage and three of Canada’s four regional academic library consortia, with the final region planning to sign in the fall of 2017. These memoranda will enable the coordinated delivery of RDM services to researchers at small, medium, and large universities across Canada.
- The development of the Portage Data Stewardship Template for the DMP Assistant platform, which will support Canadian researchers in meeting anticipated Tri-Agency DMP requirements and can be customized by Canadian research universities.
- The development of online training modules, including a series for health researchers through CIHR, and two general courses for the broader research community.
- A community of practice website that provides access to RDM platforms, such as DMP Assistant and the Federated Research Data Repository, lists of contacts for Portage’s network of specialists, training resources, Canadian RDM news briefs, and technical reports and publications on RDM.

Another major change that opens up new opportunity for the RDM community in Canada is the evolving IT environment in which it is now operating. We live in a time when information technology makes it possible for a national organization to coordinate and support a comprehensive RDM service in Canada without having to operate its own data centre. Through today’s cloud computing paradigm, technology services can be leased from multiple platform providers in customizable configurations, providing an organization with the resources to offer national RDM services. These infrastructure providers can be institutional (e.g., University of Alberta Libraries providing DMP Assistant), regional (e.g., Scholar’s Portal Dataverse instance), national (e.g., ORCID.CA), or international (e.g., ARL-COS SHARE). The services can be commercial (e.g., Globus transfer which is being used in the Federated Research Data Repository in Canada), open source (e.g., the UBC Library user interface to their digital collection discovery application, which is being used by the Federated Research Data Repository), or universities working together to provide a common solution (e.g., the COPPUL Digital Preservation Network). This new IT environment provides RDM administrators with an
opportunity to provide more cost-effective and efficient services by leveraging the assets and talents of others, rather than having to establish their own tools and platforms from scratch.

What are the current challenges for RDM in Canada?

Despite the many strengths in Canada’s RDM community, a number of significant challenges persist.

a) **Coordination and coherent sector-wide planning.** To ensure that Canada’s researchers are able to find, access, reuse, and manage data, both domestically and internationally, coordination among those responsible for delivering and supporting RDM in Canada is essential. Given the large number of actors, the complexity of their roles and responsibilities, the number of jurisdictions that are involved, and the diversity of requirements with which they must align their efforts, this task is very challenging.

As described above, national, coordinated leadership for RDM is growing through organizations such as RDC and Portage. However, current efforts to bring together RDM stakeholders have not been funded adequately or mandated formally. These two factors, a highly distributed and complex stakeholder environment and a lack of funding for coordinating processes, have made it challenging to undertake strategic and coordinated planning, thereby increasing the risk of duplication of effort. These factors have also made it difficult to develop the shared policies, processes, protocols, and standards that are so essential to ensuring researchers across Canada are able to leverage fully the promise of being able to find, access, reuse, and manage data that has been generated both at home and abroad. Lastly, these factors have created some significant gaps in Canada’s RDM platform. One striking example is that archival storage and the active, ongoing preservation activities to ensure the long-term stewardship, access and usability of data remain under-developed and under-resourced.

b) **Clarity on roles and responsibilities.** Given the diverse distributed nature of the RDM community, establishing a common understanding among all stakeholders of their shared and individual roles and responsibilities is critical to successful outcomes in this area. A paper commissioned by the Tri-Agency Data Management Working Group in 2015, a *Comprehensive Brief on Research Data Management Policies*, provided a detailed overview of the policy environment for RDM and sharing in Canada and internationally. The report’s author, Kathleen Shearer, strongly supports this assertion and outlines potential roles and responsibilities for the various actors in the RDM system in Canada. However, she also underscores that these roles and responsibilities are largely aspirational as “there is no common understanding across stakeholders about where the responsibilities lie for the various aspects of research data management.” This challenge will need to be addressed if Canada is to develop a strong and effective RDM environment.

c) **Common policies, standards, and protocols.** Common policies, standards, and protocols are critical building blocks for excellence in RDM. They allow researchers to share their data with others and to find, access, and reuse data that has been generated by their colleagues, both domestically and internationally. While the Portage network of expertise has begun to look at how to tackle this issue, and some disciplines such as genomics, astronomy, and ocean

---


7 Ibid., 40.
science have developed protocols to guide their own RDM practices, this work is still nascent in many other disciplines.  

Currently, Canada does not have the common or consistent policies, standards, and protocols that it needs to support researchers across disciplines and sectors in managing their data. Similarly, it does not have the common or consistent policies, standards, and protocols for research management information it needs to support structured and efficient RDM planning. Not only does this affect our current and future ability to leverage the enormous potential of our data, it puts at risk our ability to work as a part of international collaborations, which are becoming increasingly important. Ensuring that Canada’s RDM policies, standards and protocols are compatible with global practices is critical.

Other jurisdictions such as the European Union and the United States have invested in this area. Unfortunately, however, Canada has not been as engaged internationally. For instance, Canada was invited to be a founding national member of the Research Data Alliance (RDA), an international, community-driven organization that is dedicated to building the social and technical infrastructure required to enable the open sharing of data. Unfortunately, no Canadian organization stepped forward with the financial support sought by the founding organizations from the US, UK, EU, and Australia. On the other hand, the Australian experience illustrates how its engagement in RDA has established itself as an international leader in RDM and how the Australian National Data Service has leveraged RDM developments in other countries to Australia’s advantage. In addition to providing access to peer networks and opportunities to help shape international practice in RDM, international engagement also offers significant opportunities in terms of reinvention prevention: Canada could benefit from tools and resources that have already been developed elsewhere.

d) Skills, knowledge, and training. The recent emergence of data-intensive research and the growing understanding of the potential of big data across many fields have revealed the largely neglected state of RDM in Canada. As outlined in the Canadian DI Environmental Scan that was prepared for Summit 2014, “there is a significant unmet need for skills upgrading, training, and mentoring in the use of advanced computing, especially in disciplines that have not had extensive engagement in data-intensive research until recently. While improving, there is still a paucity of awareness of RDM principles and good practices among researchers and research universities; relatively few researchers have training in RDM; there are few positions for data managers/professionals; training opportunities are sparse.”

Once again, Portage is helping to address these challenges, and a number of Schools of Information Management have begun to add RDM to their curriculum, which is an important development. In addition, more and more universities across the country are investing in training and have dedicated library staff providing RDM support services to researchers. However, this investment tends to be concentrated in larger universities. More needs to be done to support the skill and knowledge development of frontline staff at universities that have historically lacked the capital and training opportunities needed to serve the RDM needs of their researchers. Addressing this need is particularly acute in the context of the proposed Tri-

---


Agency data policy now in consultation, as it is anticipated that many universities will require support in responding to its new requirements. Scaling up and providing better access to training and mentoring services is essential if ensuring sustainable and equitable access to RDM support is to be achieved at all universities across Canada.

e) **Attraction and retention of highly qualified personnel.** RDM, while dependent on digital platforms, is fundamentally a people-intensive activity. Data professionals work with those creating and using research data to help them with tasks such as planning the management of data research in projects, preparing metadata, and curating long-term preservation of data. Increasing the number of professionals to meet the full range of Canadian researchers in the new DRI-dependent environment needs to be a priority. Ensuring sustained and predictable funding for RDM is also important. Because funding for RDM is unpredictable and project-based, many of the highly qualified and highly sought-after personnel who undertake the work associated with RDM are in precarious contract or term positions, making it harder to attract and retain these critical individuals.

f) **Academic journals.** While journals are becoming important actors in helping to shift attitudes and drive a more RDM-aware culture, their growing understanding of the value of data has encouraged a trend toward the acquisition of previously open-access disciplinary data repositories, increasing the risk that vital research will be placed behind a paywall. To avoid a situation in which journals have a monopoly on data and Canadian researchers and universities are forced to pay for access to this data, publicly owned and managed data repositories and archival storage are essential.

g) **Culture.** As research in all domains becomes increasingly data-driven, awareness among researchers, funders, and universities about the importance of proper data management is also growing. Evidence of this change is found in the proposed Tri-Agency data policy, now in consultation, that would impose RDM requirements on future grant recipients and universities, as well as the data deposit requirements of an increasing number of academic journals. However, despite these important shifts, a number of barriers persist to the development of a strong RDM culture that enables data-sharing in Canada. In her report, Shearer offers a number of reasons for this gap: 1) attitudinal barriers – researchers have indicated that they are concerned about issues such as a loss of control, being scooped, and privacy; 2) technical challenges – a Dutch survey across 15 international jurisdictions concluded that technical challenges such as obsolete software were an issue; 3) a lack of professional expertise and formal training in data management; 4) a lack of formalized and standardized procedures; 5) insufficient peer support for and awareness of the importance of RDM; and 6) insufficient incentives and rewards.10 These barriers will need to be addressed if Canada is to build a culture of and commitment to RDM among universities and researchers.

h) **Funding.** Effective RDM is not something that can be undertaken in fits and starts or in a piecemeal fashion. To ensure that the potential benefits of Canada’s research data are optimized, they must be maintained in a sustainable environment and managed over time.11 Significant funding challenges have made this a difficult state to achieve in Canada. In a

---


competitive environment, the importance of RDM can be lost to competing needs of other components in the DRI ecosystem. In addition, RDM lacks defined revenue streams, and its importance is often overlooked by funding organizations and agencies. When revenue streams do exist, they tend to be project-based, which makes developing and sustaining communities of practice, tools, and platforms challenging. It has also compromised the RDM community’s ability to undertake strategic visioning, coordinated short- and long-term planning, and collaborative implementation.

i) Tools and platforms. Across Canada, significant local innovation in RDM tool and platform development helps support the unique needs of researchers from a variety of disciplines in the management of their data. Research universities are also developing private or adopting shared data repositories to meet their researchers’ active data storage needs. National organizations, such as Portage and RDC, support this innovation, fostering a Canadian community of practice to increase this local capacity in collaboration with stakeholders across the country, through the development of national services and software that cut across university and disciplinary boundaries. For instance, an internationally used research data management planning tool, DMPOnline, has been adapted, modified and translated into French for use by all Canadian research universities. As well, a model of federated data repositories is currently in development by CARL, Portage, and Compute Canada.

Scaling local innovation to meet the range of requirements of diverse communities of practice will be impossible without national coordination and support. Significant gaps in RDM infrastructure capacity compound this challenge. For instance, increasing the supply and ensuring the interoperability of active and archival storage infrastructure poses a significant challenge. Tools are needed to better capture and incorporate research management information (RMI, aka admin data) to support collaboration among universities, funding agencies, and disciplines. As well, shared processes and tools for migrating research data stored in obsolete software must also be developed. Meanwhile, a lack of coordination among RDM stakeholders has resulted in insufficient infrastructure that cuts across sectors and domains for widespread adoption. The poor coordination of RDM initiatives has also resulted in unequal adoption of RDM tools and best practices by researchers and supporting players.

The number of platforms providing aspects of RDM support is increasing in Canada and nascent archival storage services are being planned. But there are not enough of them at the level and scale of service required by the research community, particularly given the impending Tri-Agency policies on RDM. Furthermore, they are not coordinated, putting Canada’s research investment at risk for the future. The true potential of the new data universe will only be fully realized if researchers can find, access, reuse, and manage the data of others.

One of the most significant gaps in the current ecosystem is the provision of archival storage. A 2015 report by RDC\(^\text{12}\) suggested that storage requirements for NSERC and CIHR research data in the long tail (i.e., not big data) would be 45 Petabytes for a 5-year period. If one were to include CIHR-funded projects as well as all other publicly funded research data, the number would be three or four times this. In speaking with Canadian researchers who want to share their data openly, RDC has found that the biggest gap is the availability of archival storage for specific datasets. A primary reason for this gap is that no single organization has the mandate to fund and support the provision of archival storage. For example, CFI sees as its role to fund

only active storage for approved research projects. Academic libraries are striving to fill the archival storage gap either individually or through regional library consortia. Their focus to date has been to protect their investments in acquired digital collections. A wider injection of funds is desperately needed to achieve archival storage for research data on a national scale.

**j) Organizational and researcher readiness for new RDM policy.** As discussed elsewhere in this document, new domestic and international requirements for RDM are increasing. At home, the highly-anticipated *Tri-Agency* data policy would place new RDM requirements on both researchers and universities. As stated above, while significant progress has been made at some universities and a number of discipline-specific communities have been investing in this area, university and researcher commitment to and engagement in RDM has not been consistent and has often been hampered by the lack of system coordination, funding, and human resources. These issues will become particularly acute if researchers and universities are required to comply with the anticipated Tri-Agency data policy requirements. Not only will researchers need tools, expertise and training to complete their data management plans, they will also require access to data repositories and archival storage, neither of which are well-developed RDM areas in Canada at this point in time. They will also require a strong RMI infrastructure to enable efficient exchange of information among diverse administrative applications, e.g. data management aspects of ethics and research proposals.

The RDC submission to the Industry Canada DRI consultation argued that, when looking at the question of university/organizational readiness, you must consider both capability and capacity to deliver. The submission further notes that “in some institutions the capability exists, but within current resources it cannot scale to a capacity that could serve the entire cohort of researchers. In more institutions, there is neither the capacity nor the capability. It is also unrealistic to imagine that many smaller institutions could effectively maintain both capacity and capability.” The Portage network of expertise is making progress in tackling the issue of university and organizational readiness for pending Tri-Agency data policy requirements. However, while it will add foundational RDM capabilities to RDM practice in Canada, significant resources will be required to bring it to scale.

### 3.2.2 How is RDM managed in other international jurisdictions?

The challenges in establishing the foundational RDM tools, platforms, and practices described above are not unique to Canada. Countries around the world are dealing with similar challenges in managing their data assets: some have well established RDM programs, others are in the early stages. What is common to all is the recognition that research data is increasingly important to the national and international innovation fabric.

An important international trend is the movement toward coordinating shared and distributed RDM responsibilities at the national level. This trend is especially strong in Europe, where it has been partially driven by the European Commission’s strong commitment to Open Science and Data as a key driver of the EU’s shared economy. It also recognizes that research universities (many of which have been in existence for decades or centuries) have well-developed

---

14 Ibid., 3.
15 Ibid., 3.
infrastructure and service offerings, and duplicating this effort make little economic sense. The increasingly collaborative and interdisciplinary nature of most research makes this level of coordination and facilitation necessary. Some examples are highlighted below.

- In Australia, the Australian National Data Service (ANDS) has been established as part of the Australian National Research Infrastructure Strategy, and provides a suite of services to support researchers, such as national registry and discovery services for research data, unique identifier services, and skills training and outreach that operate in conjunction with national and domain-specific infrastructure services. Recent efforts in 2016–17 brought the business plans of the three primary national research organizations (ANDS, Nectar, and RDA) into alignment to ensure that investment “efficiencies gained from synchronising our projects will allow us to provide benefits across the research community.”

- In the United Kingdom, Jisc, a not-for-profit higher education and research organization, receives funding from the Higher Education Funding Council for England to provide shared digital infrastructure and services to partner institutions. One of Jisc’s recent priorities was establishing the Research Data Shared Services program, which will provide a new “a la carte” RDM platform, as well as an RDM services and consultancy initiative to support researchers and their organizations.

- In the United States, a vision for a National Data Service (NDS) is emerging that aims to provide a suite of RDM services and infrastructure to researchers. This initiative being undertaken by a consortium of volunteer participants from approximately 50 different research organizations.

- In Japan, the National Institute of Informatics (NII) has begun work to establish a national-level framework for RDM, sharing, publication, and reuse that is expected to be implemented by 2020. The NII is developing the necessary research data platforms and associated services in collaboration with higher education and research institutions.

- Korea is developing a National Data Service, tentatively called KORENDS, that will create a national-level framework and services for RDM. The NDS will provide RDM services to researchers at science and technology research institutes funded by the Korean Government, as well as institutes of higher education.

- In Finland, the Ministry of Education and Culture has been funding a National Research Data Initiative for over a decade, and the recent Open Science and Research Initiative resulted in the establishment of common research data infrastructure and services that are provided to researchers by national and institutional stakeholders such as the National Library, Finnish Data Archive, Helsinki University Library, the FIN-CLARIN consortium of Finnish universities, and CSC – IT Center for Science.

- In the Netherlands, the national networking (SURFnet), and HPC entities (SURFsara) have formed one organization (SARA), that includes RDM as a core part of its service mandate. The Data Archiving and Networked Services initiative in the Netherlands (DANS) has been an international leader in the development of trusted digital repository services.

---

16 Drawn from the draft document being developed by the National Data Services Interest Group of the Research Data Alliance, which has additional details on the types of services, budgets, etc. https://docs.google.com/document/d/17iUyJ2icY9gFzMZGPWjyy5E0tUoukAtt4BFeRonefv4.


frameworks, and has been providing long-term archival storage for the nation’s researchers since its inception.

- In the Czech Republic, CESNET was founded as a national research and education network that also provides researchers with data storage infrastructure and RDM services. Like many national services, CESNET is an association of all universities in the Czech Republic and the National Academy of Sciences.
- In Qatar, a research data curation service of the Qatar National Library (QNL) aims to offers guidance, training and data storage services to researchers. The service is funded from the budget of the QNL, with the involvement of the Qatar National Research Fund, the Qatar Foundation, and other stakeholders.

Another important trend in international RDM is the development of national RDM policies and frameworks. In her Comprehensive Brief on Research Data Management from 2015, Shearer found that many funding agencies and institutions in other international jurisdictions are ahead of Canada in introducing RDM policies for their research communities. The objectives of these policies largely focused on improving the efficiency of research, supporting the reuse of data for new insights and discoveries, fostering collaboration, and facilitating greater transparency.

Shearer further found that the jurisdictions with the most comprehensive policy environments are the United Kingdom, United States, Australia and the European Union. Details of their policies varied across regions, agencies and domains, but they also shared a number of elements, including requirements around standards and metadata, data sharing, and data retention and/or long-term preservation. A requirement for data management plans (DMPs) was also common, as were policies containing provisions for the protection of confidentiality, intellectual property, and sensitive data.

Examples of RDM policy frameworks in other countries are numerous (some examples are listed below), and as international agreements and legislative frameworks evolve to encourage more cross-border research collaboration, these policy frameworks will likely become more synchronized. This is highlighted by the recent release of “Legal Interoperability of Research Data: Principles and Implementation Guidelines.”19 by CODATA/RDA, which defines an RDM policy framework that was agreed to by a broad international stakeholder group.

This document is also an excellent example of the role that RDA plays in the international RDM ecosystem — a role that Canada needs to embrace more wholeheartedly to ensure that national RDM practices match those being developed elsewhere. RDA has over 6,000 members, and its technical standards and policy work has been adopted by the European Commission20 and other agencies that recognize the value of RDM and interoperability on an international scale.

- In the United Kingdom, Research Councils UK has issued a set of principles that call on constituent councils to implement RDM policies, although requirements may vary according to the funder. An even broader community of stakeholders in the UK (all public funders, universities, and the Wellcome Trust) developed the “Concordat on Open

Research Data," which was released by the Minister of State for Universities and Science in July 2016. The Wellcome Trust refers to the Concordat when it states: “There is international consensus on the need to share and preserve research datasets in a way that maximises their long-term value.”

- In the United States, the National Institutes of Health (NIH) and the National Science Foundation (NSF) have adopted their own policies regarding research data sharing, which have recently begun to require researchers to submit a DMP with their funding application, and to deposit their data where appropriate. The White House’s Office of Science and Technology Policy (OSTP) has also released a requirement for federal agencies with annual research and development expenditures greater than $100 million to develop plans to manage digital data resulting from their own funded research, whether generated internally or externally. This requirement is articulated in the context of international collaboration in a December 2016 report “Principles for promoting access to federal government-supported scientific data and research findings through international scientific cooperation.”

- In Europe, researchers funded by Horizon 2020 have been required to deposit their final research publication for a number of years, and as of July 2017, new grant recipients will be required to deposit data, although they can apply to opt out as per the funding guidelines. Researchers also must comply with requirements to produce research data management plans.

- In Australia, a policy directing requirements for responsible research was issued jointly by university and national research councils. The “Australian Code for Responsible Conduct of Research” includes guidance on data management plans, stewarding research data, and more.

- Many funding agencies (public and private) are issuing RDM policies that are tied to research grants. The Wellcome Trust’s RDM Policy framework has been in place for a number of years and is probably the most forward-thinking of the policies. It was modified most recently in July 2017, and now includes the requirement to deposit “materials that will hold clear value as a resource for others in academia or industry,” which includes physical assets such as cell lines. Other organizations, like the Bill and Melinda Gates Foundation, are leaders in the development of data deposit policies. In one recent example, the Gates Foundation changed their publication policy to restrict grantees’ research outputs to journals that support open access/data.

---


3.3 Future State

3.3.1 The Future We Seek

In ten years, with appropriate strategic planning, coordination, and investment, we see a future in which RDM in Canada will be transformed.

At the community level, the transformation will result in:

- a nationally coordinated RDM community that provides researchers with robust RDM services and resources in both official languages;
- coordinated RDM services that are offered to researchers on all Canadian campuses and managed collaboratively by all regional and local service providers, such as libraries, offices of research, and computing services;
- a research data management community and delivery system that is recognized internationally as a leader;
- innovation for local and domain-specific RDM services and resources that is supported and encouraged nationally;
- libraries adopting the long-term preservation and discovery of research data as a core responsibility;
- a national network and communities of practice of highly qualified personnel for RDM services that support an integrated set of RDM platforms, as well as expertise that can help those serving researchers at all universities;
- a strategic planning process that is informed by, and responsive to, continual feedback from users across all disciplines, changes in policy, technology, local innovation, and continuous operational learning and improvement;
- Canadian data repositories being developed and interoperating collaboratively in a federated environment, nationally and internationally;
- international standards being used to advance the certification of research data repositories in Canada;
- an environment in which a variety of research support services, such as registries for persistent identifiers, and current research information systems (CRIS), are integrated and supportive of RDM;
- RDM service providers being a recognized and integrated part of a broader digital research infrastructure ecosystem that is delivering seamless and integrated access to Canadian researchers;
- the unique data interests of Indigenous Peoples being an integral part of the RDM community;
- a widely accepted societal risk management framework defining research access to sensitive data, especially regarding the privacy and security of personal information in health research;
- research data and metadata being integrated into scholarly communications; and
- a sustainable and transparent system for coordinating RDM within Canada’s digital research infrastructure being firmly established.

For researchers, the transformation will result in:

- easy and intuitive access to RDM expertise, training, and resources across a full spectrum of user needs and all disciplines;
• simplifying and reducing the amount of work required for researchers to be in compliance with the data policies of federal granting agencies, other funding agencies, and universities;
• efficiencies in RDM, saving time and resources;
• recognition and reward for the curation and sharing of research data;
• increased domestic and international collaboration by allowing greater interoperability and accessibility to research data;
• new connections among researchers that are enabled by improved RDM practices;
• an increase in interdisciplinary research that results in leading edge outcomes through new and exciting knowledge creation and innovation;
• early career researchers having greater access to research data and a richer research environment, enabling them to achieve earlier and more significant results; and
• a culture of data sharing and recognition of the importance of excellence in data stewardship as the norm across all disciplines.

For the management of research data, the transformation will result in:

• RDM practices across disciplines that are based on the FAIR Guiding Principles, making research data Findable, Accessible, Interoperable and Reusable;
• advances in data documentation and curation practices that make research datasets independently understandable and improve the reproducibility of research findings; and
• improved RDM practices that produce higher quality research data and that increase the analytic value of the data and subsequent outputs.

3.3.2 Vision

An innovative and coordinated research data management community, providing responsive services and resources that support Canadian researchers in advancing the research that is critical to building and sustaining Canada's economic and social prosperity.

3.3.3 Principles

The following principles are vital to achieving the vision and desired future state, in which Canadian researchers across a broad spectrum of needs and disciplines are able to access RDM services and platforms seamlessly, with those services and platforms developed and delivered locally, regionally, and nationally, and aligned with international efforts.

• Responsive to the needs of the research community.
• Adaptive to changes in research and data management.
• Innovative in developing local and domain-specific solutions that are scalable.
• Coordinated among key local, regional, and national research service providers.
• Integrative by design to facilitate easy incorporation of new and interoperable services and platforms.
• Collaborative and cooperative working relationships among RDM stakeholders, as well as with those in the broader digital research infrastructure ecosystem, to enable seamless service to researchers.
• Distributed to maximize stakeholder cooperation and to leverage their engagement in a community of practice.
• Accountable to Canada’s research community, and to those funding it.
• **Sustainable** to allow its ongoing operation.
• **Diverse** by valuing and respecting the importance of diversity.

### 3.3.4 Goals

To achieve the vision and desired future state, the following goals must be pursued rigorously by the Canadian RDM community.

• **Build** innovative services and resources that are distributed across universities, nationally coordinated, internationally recognized, and sustainable, while being responsive across the full spectrum of researcher needs and disciplines. These services and resources should respect researcher, discipline-specific, national, and university data stewardship policies, and be based on best-practice standards and protocols.
  
  o Develop a national federated storage network to support Canadian research data repositories that provides both active and archival storage for the reuse and preservation of data by researchers.
  
  o Identify and secure sustained and predictable funding for both operational and capital investment, including a coordinating body to support the community in providing well-managed processes for coordinated and collaborative service delivery across Canada.
  
  o Coordinate local and domain-specific resources and services to align with international RDM standards and protocols; and
  
  o Develop university plans and procedures for RDM in all Canadian higher education establishments in accordance with federal granting agencies and other funder requirements.

• **Advance and adopt** RDM processes and procedures that are informed by researcher, university, and discipline-specific needs, to improve the overall quality of research data and to advance best practices. This will require flexible and adaptive tools and platforms supporting data planning, creation, curation, deposit, access, discovery, and reuse.
  
  o Develop a national software framework that supports the development of innovative local and discipline-specific tools to support researchers with RDM workflows and that is aligned with similar frameworks internationally.
  
  o Develop and ensure widespread adoption of RDM definitions, taxonomies, and unique identifiers by the broader research community.
  
  o Integrate metadata production into RDM tools at the project level and throughout the research lifecycle to automate its capture and reuse.

• **Establish** a community of practice that is supported by a distributed network of specialists who can provide expert advice, support, and training in RDM best practices to researchers.
  
  o Increase the capacity of Canada’s higher education sector to support RDM services and resources for researchers and organizations supporting research. This includes encouraging collaborative relationships among institutional units, such as libraries, offices of research, research ethics boards, IT services, and external RDM organizations.
  
  o Continue to strengthen the capacity of libraries to provide front-line support for RDM as key service points and loci of expertise on campus.
  
  o Support, at national and local levels, the development of a variety of RDM training resources, such as online courses, webinars, guides, presentations, and in-person workshops.
Engage Canadian non-governmental organizations and government RDM agencies as contributing partners in this community of practice.
Facilitate and encourage Canadian participation in international RDM organizations and the adoption of best practices recognized by the international community.

3.4 National Coordination and Facilitation

3.4.1 Why is it important to have a national approach to coordinating and facilitating RDM?

When researchers are able to find, access, and reuse current and historic data easily and efficiently, the opportunity for scientific breakthrough and discovery increases. To be most effective, these activities must be supported in ways that permit data-sharing among researchers across Canada and internationally. Accomplishing this task would be difficult, if not impossible, for any single university or region to do on their own. It would also be highly inefficient, causing duplication of effort and increasing the possibility of confusion, rather than leveraging shared opportunity.

National coordination is essential to enabling coherent, efficient, and effective RDM in Canada, and to prevent barriers to data sharing. Specifically it:

- ensures that all researchers and administrators across Canada have access to RDM services and platforms, regardless of their discipline, geographical location, or the size of their university;
- enables collaborative and efficient development of the policies, standards, protocols, processes, and procedures essential to ensuring that researchers can find, access, and reuse data generated in Canada and elsewhere in the world;
- ensures that tools and platforms, such as data repositories and archival storage, are interoperable across Canada and internationally, facilitating access and sharing, rather than creating barriers;
- facilitates consistency of practice and approach to RDM across Canada by supporting strong communities of practice, access to networks of experts, and shared training;
- leverages expertise and shared investment across universities, government, and other funders, increasing quality, impact, and financial efficiency; and
- builds a collaborative RDM culture that engages researchers and university administrators across Canada, creating awareness of and support for RDM and ensuring that rather than stepping on each other’s feet, they are standing on each other’s shoulders to advance RDM in Canada.

3.4.2 How could the Government of Canada support the national coordination and facilitation of RDM?

In order to ensure that all Canadian researchers have access to high quality, equitable, and efficient RDM services and platforms, we are proposing that the Government of Canada consider investment in a national body that would coordinate and facilitate key RDM activities that are most effectively undertaken at the national level in Canada. This new body, Research Data Management Canada (RDMC), would build on the critical foundations laid by Portage and
RDC. Using a federated service model, it would support and leverage the RDM practices and approaches of universities and discipline-specific communities across the country, while also respecting their independence and autonomy in providing services and platforms to their researchers. It would also establish and maintain innovative, coordinated, and interoperable RDM services, tools, and platforms that would be accessible to administrators and researchers in all Canadian universities, and supported through a community of practice and a network of experts.

3.4.3 How would RDMC undertake its work?

The verbs that best describe how RDMC would undertake its work are: support, coordinate, facilitate, and leverage. This would be accomplished in three ways:

1) **Creation of a network of experts** to support those providing on-campus, front-line delivery to researchers. This would ensure that RDM support staff, no matter their location in Canada and regardless of their size or needs, would have access to both the general and specialized expertise they need to serve their research community most effectively. These experts would also work collaboratively across universities and disciplines to coordinate the development of the common standards, policies, protocols, tools, and platforms needed to facilitate the ability of a researcher in Toronto to find, access, reuse, and manage data generated by a researcher in Calgary, Fredericton, Iqaluit, or Beijing, as well as to help them manage data within their own labs.

2) **Support for communities of practice** composed of RDM stakeholders at the discipline-specific, local, regional, national, and international levels. These communities of practice would provide RDMC with an important conduit to facilitate the building of RDM skills and knowledge in universities across Canada. They would also allow it to tap and leverage RDM expertise both at home and abroad, rather than having to develop in-house capacity in all areas. In addition, these communities of practice would also be leveraged to facilitate collaborative development of the shared RDM standards, policies, protocols, tools, and platforms required to ensure RDM excellence in Canada. RDMC will support a community of practice for RDM, working with Canadian universities to deliver local RDM services, increasing the capacity needed to strengthen local RDM services, developing and delivering RDM training in both official languages, establishing service level agreements with providers of infrastructure platforms for RDM plans, data curation, data preservation, and data discovery, and overseeing the coordination, planning, and funding of the deployed RDM operations.

3) **Collaboration with other partners** within the broader DRI ecosystem and other sectors, who would, through contracts and other arrangements, be part of the federated research data service model. For instance, RDMC would leverage the new IT environment that enables more distributed access to and delivery of RDM services, tools, and platforms. RDMC will not need to operate expensive data facilities independently nor to create new software from scratch. Instead, RDMC would work with and leverage the assets and skills of many different infrastructure providers, both domestically and internationally, such as universities and private industry, all of whom would contribute components such as archival storage, repositories or software tools, to the nationally coordinated RDM service. RDMC will seek partnerships that capitalize on domain-developed RDM innovation and will work to introduce these advances to the wider research community.
3.4.4 How would RDMC be accountable to the research community?

The LCDRI is proposing that an Advisory Committee, consisting of representatives from regional organizations, domains, and universities, be established to provide advice and direction to RDMC.

3.4.5 How would RDMC deliver on the core RDM functions?

To coordinate and facilitate RDM services and platforms at the national level, it is essential to employ a service model that is well suited to leveraging Canada’s diverse and existing RDM assets and expertise, that is closely aligned with the Canadian research context, and that is responsive and agile. The LCDRI is proposing a federated service model that includes distributed delivery of each of the core RDM functions: policies; standards and protocols; processes and procedures; leadership, advice, support, and training; and tools and platforms. See Figure 3.
i) Policies

RDM policies are foundational to ensuring that researchers can find, access, reuse, and manage their data. They articulate the frameworks through which RDM occurs, facilitating consistent approaches and practices among administrators and researchers both at home and abroad. The *Tri-Agency Statement of Principles on Digital Data Management* is a good example of policy guidelines that are helping shape RDM practice in Canada. The *Tri-Agency Statement* articulates the RDM-related expectations and responsibilities of the various stakeholders who have a role to play in RDM, including researchers, research communities, research universities, and funders.

The *Tri-Agency Statement* is just one example of guidelines to shape data policy development and implementation. Examples of other important RDM policy developments include the proposals for university strategies in managing research data and for the deposit of datasets upon which published research findings are based. Both need to reflect and link with internationally recognized best practices.

Sensitive data and access to it for legitimate research use in Canada is another example of an important issue requiring RDM policy. Research areas where this is particularly important include access to health data and data concerning Canada’s Indigenous Peoples.

Successful development and implementation of these policies requires a community of well-trained and knowledgeable RDM administrators and researchers whose efforts are effectively coordinated. For instance, in the case of access to sensitive data, a review of relevant legislation and policy regarding privacy and ethics as it relates to the protection of research data needs to be conducted and a framework for community consent and approval of the preservation and reuse of this data needs to be developed. Many different communities and individuals will need to be included in this work to ensure consistency of approach and practice. In addition, the effort required will go beyond the financial and human resource capabilities of many universities.

RDMC would play a key role in helping to support the development and implementation of essential RDM policies such as those described above. Through its communities of practice and network of experts, it would leverage RDM expertise, both domestically and globally, to help establish and maintain the RDM policy frameworks that Canada needs. In addition, it would support their implementation by providing tools, resources, and training to front-line service staff in universities across Canada, to ensure that researchers are able to access the high quality RDM support that they need on their campuses. The significance of this type of support was illustrated during an interview with a privacy officer in an agency, who was struggling to understand emerging data policy. She relies on a community of practice to interpret privacy legislation and policy when implementing RDM practices and procedures in her organization. When she encounters a tough question from a researcher regarding an aspect of privacy implementation, she turns to an RDM community of practice to seek advice on how others in her position have addressed the problem.

ii) Standards and Protocols

Standards and protocols are key to the efficiency and effectiveness of many aspects of RDM, such as data instrumentation, formats, security, transfer, curation, and storage. They support the implementation of RDM policies by articulating the critical information needed to guide RDM practice in support of these policies. As well, they support researchers by ensuring that RDM is
operationalized to maximize their ability to understand data generated by others, support their workflows, systematize the curation and storage of their data, and ensure that their RDM tools and platforms can function in an interoperable way. Increasingly, standards also enable those machine-to-machine transactions needed to manage massive volumes of research data.

A joint initiative between CASRAI and Portage to develop shared definitions for key elements of the data management plan (DMP) template is a good example of how RMI standards and protocols can reduce confusion and support implementation of RDM and how the worlds of RDM and RMI intersect. Having consistent and shared definitions of DMP elements will reduce the amount of time and effort required by researchers to fill out DMPs, increase accuracy, and, through an API, enable elements to be exchanged between software platforms, making relevant information shareable.

Coordination can increase the speed by which standards are adopted by facilitating access and knowledge about use. For instance, a group of national organizations (CASRAI-CA, CRKN, CARL, RDC, CANARIE, Compute Canada, and individual universities) have recently come together to coordinate implementation of an international standard for unique researcher identification (ORCID.CA). The opportunity to leverage the collective reach of these organizations, as well as their unique skills and expertise, has greatly increased the speed of adoption of this standard.

RDMC would support the development and implementation of RDM standards and protocols by leveraging existing domestic and international expertise in key areas such as RDM workflow and interoperability of tools and platforms. It would also identify gaps and facilitate the collaborative development of new standards and protocols, through communities of practice and contractual arrangements with other specialized service providers.

**iii) Processes and Procedures**

Processes and procedures are critical to supporting efficient and effective data workflows. They are also critical to facilitating data curation and security, both of which are most effective when managed throughout the research process.

Coordinating the implementation of processes and procedures can be challenging because of the diversity and number of RDM stakeholders engaged in them. For instance, researchers and their data managers implement and administer RDM processes and procedures in their projects. There also are data workflow specialists within disciplines who assemble automated procedures for commonly managed data types and processes. In addition, local universities often have RDM consultants who assist researchers with data curation or who are responsible for the curation of collections of data from their campuses.

Through its network of experts and communities of practice, RDMC would facilitate the implementation of RDM processes and procedures in Canada. It would ensure that administrators and researchers have access to the general and discipline-specific expertise that they need in areas such as research workflow, data curation, and security. It would also identify gaps and leverage domestic and international expertise to develop new processes and procedures, as appropriate.
iv) Leadership, Advice, Support, and Training

As discussed elsewhere in this paper, Canada has not yet developed a strong culture of RDM, and the skills/resources that researchers need to change this situation are unevenly distributed and often underdeveloped. Leadership, as well as advice, support, and training on RDM practice is critical to changing this situation. For instance, training programs for front-line service staff are needed to ensure that researchers have access to knowledgeable and up-to-date RDM service providers on campus. Also, having a network of experts who provide both general and specialized advice would help ensure a consistent practice and approach to RDM across Canada. It would also enable equitable access to RDM services, regardless of a researcher’s location or the size of their university.

RDMC, through its network of experts and communities of practice, would facilitate the development and delivery of collaborative and coordinated RDM advice, support, and training across Canada, ensuring that Canadian RDM practice remains current and that no researcher or university is left behind. It would also nurture relationships with key partners in RDM training and support, coordinate service providers, and maintain service agreements. Lastly, RDMC would provide a central point for RDM leadership in Canada, with responsibility for developing and maintaining a strategic, national RDM framework and for coordinating Canada’s RDM presence abroad.

v) Tools and Platforms

There are several technology tools and platforms that provide the essential infrastructure needed to enable RDM. These include software, data repositories, and archival storage. Each of these components is discussed separately below.

Portage has demonstrated how a federated approach to RDM infrastructure can be implemented successfully in the new IT environment, with diverse and multi-sectoral infrastructure providers all contributing components to a nationally-coordinated service. RDMC would bring this model to scale and provide the secretariat support required to coordinate and facilitate its implementation and ongoing operation. This would enable researcher access to RDM services, regardless of their location or local capacity. These tools and platforms would not replace those developed and maintained by universities for local research uses. Rather, the intention would be to augment these local tools and platforms and to leverage them, as appropriate, for wider community use.

Software tools to enable RDM

RDM software tools are required to enable all data-related activities of the research process: planning, creating, processing, analyzing, disseminating, preserving, and reusing research data. Access to these tools facilitates best practices across these activities, is critical to producing high-quality data, and ensures that research data are findable, accessible, interoperable, and reusable (the FAIR principles). RDMC, as a federated national service, would be able to assemble diverse partners, leverage expertise, and coordinate the collaborative development of critical RDM software tools.

The DMP Assistant is an example of coordinated implementation of a tool within a federated environment. The Assistant helps researchers prepare data management plans before starting a project and maintain an active record of RDM activities throughout a project. Portage established a two-year agreement with the University of Alberta Libraries to provide the DMP
Assistant to the research community as a free, bilingual tool. Through its DMP Expert Group, Portage also developed a Data Stewardship template, guidance text, and a help desk to support researchers with their DMPs. The University of Alberta maintains the tool while Portage supports the research community’s use of the tool. This provides a helpful model for future software tool development and implementation.

The proposed budget below requests financial resources for developing and leasing or sharing software tools that support researchers throughout the research process.

**Data repositories**

Data repositories are essential to supporting the data-related activities of disseminating, preserving, and reusing data. They facilitate data discovery, access, and preservation by ingesting data, supporting its curation, processing it for preservation, indexing it for discovery, storing it for safe keeping, and supporting its retrieval for reuse. Providing the infrastructure and services required to enable these activities on an individual project-by-project basis would require large investments by universities and governments across Canada. Furthermore, researchers would have to expend enormous amounts of their time to support such services. A more efficient, collaborative response is required. In Canada, some universities have begun to step forward to provide data repositories, including university libraries (e.g., UBC Abacus), domains (e.g., Canadian Astronomy Data Centre and CANFAR), regional consortia (e.g., Scholars Portal’s Dataverse instance), national agencies (e.g., DFO). However, not all researchers or universities currently have the access they need to these services.

Using its federated data service framework, RDMC would coordinate and facilitate the establishment of a network of existing data repositories. It would develop a set of mutually accepted standards and protocols to ensure the flow of data and metadata among platforms and services. This collaborative network of repositories would increase the sustainability of each repository, by sharing services that each needs to access and preserve their data collections. Some repositories could also provide specific services to other repositories that have not implemented a feature, e.g., preservation processing or ingesting big data. Furthermore, this network of repositories would enable the implementation of succession policies for backing up one another in the event a repository is forced to cease operating.

In addition, university data repositories could be leveraged to provide services to researchers at universities without a data repository. Portage has demonstrated this potential through its partnership with Compute Canada and UBC in the development of the Federated Research Data Repository, and through its collaboration with the Dataverse North consortium in Canada. In both instances, Portage is leveraging platforms maintained by infrastructure providers to support the services of a community of practice. The Portage Network of Expertise is working with local campuses to support researchers in their use of these platforms and to establish an environment in which these platforms can interoperate and share functionality, especially for preservation processing. RDMC would bring these initiatives to scale and make them available to universities across Canada.

This proposal requests financial resources to purchase archival storage that would be contributed as part of RDMC’s federated service model, augmenting and leveraging archival storage that certain universities or regional consortia develop and maintain as storage providers. This allocation of archival storage would only be a portion of the overall storage for digital preservation in Canada but it should incentivize a variety of organizations to contribute archival storage to a federation of archival storage providers. We believe these organizations
could contribute an additional 25% to the total archival storage costs for Canadian research data, at scale.

To integrate RDM services, Canada’s network of data repositories would be aligned with archival storage providers. Portage is currently engaged in a project to demonstrate the scalability of this approach. Through its secretariat, RDMC would provide the support needed to coordinate and facilitate the implementation and ongoing operations of such arrangements.

**Archival Storage**

A significant gap in Canada's RDM infrastructure is the provision of the archival storage required for the long-term safekeeping of research data. Following the model for federated research data service that has been articulated above, we propose that archival storage also be federated. This would entail collaborating with organizations developing archival storage for the protection of ever-increasing and valuable digital collections. The library community is a prominent player in providing such services. Examples include the Ontario Library Research Cloud for the Ontario Council of University Libraries and the Digital Preservation Network of the Council of Prairie and Pacific University Libraries. In addition, the Digital Preservation Working Group of CARL is studying today's capacity for digital preservation in Canada. Aligning the archival storage requirements for RDM with the library community and other national digital preservation initiatives provides a way of sharing resources that strengthens the overall storage capacity for Canadian digital collections, of which research data is an important part.

This proposal requests financial resources to purchase archival storage that would be contributed as part of RDMC’s federated service model, augmenting and leveraging archival storage that certain universities or regional consortia develop and maintain as storage providers. This allocation of archival storage would only be a portion of the overall storage for digital preservation in Canada but it should incentivize a variety of organizations to contribute archival storage to a federation of archival storage providers. We believe these organizations could contribute an additional 25% to the total archival storage costs for Canadian research data, at scale.

To integrate RDM services, Canada’s network of data repositories would be aligned with archival storage providers. Portage is currently engaged in a project to demonstrate the scalability of this approach. Through its secretariat, RDMC would provide the support needed to coordinate and facilitate the implementation and ongoing operations of such arrangements.
3.4.6 What level of federal investment is being proposed to fund RDMC?

As described in Table 1, LCDRI is proposing a budget of $54 million over 5 years to support the work of RDMC. The proposed budget has been broken into three main cost categories (highly qualified RDM personnel, tools and platforms, and a secretariat office), reflecting the goals and approach proposed. Highly qualified personnel includes secretariat staff, staff related to the Network of Experts, and contracted experts; details are in Table 2. Tools and platforms includes software tools, data repositories, and archival storage, including the costs of external service contracts for their operation and maintenance. The operational budget of the secretariat includes operational costs such as IT, travel, and office rental and set-up.

It should be stressed that the majority of the investment that is required to support RDM in Canada would continue to be made by regional consortia and individual universities. The federal contribution to RDMC that is proposed below would be used to leverage these investments and provide a critical point of coordination for the development of nationally shared RDM policies, standards, protocols, processes, expertise, tools, and platforms that are needed to ensure that all researchers across Canada have access to DM services and platforms, regardless of their discipline, geographical location, or size of their institution.

### Table 1. A Five-Year Budget of Research Data Management Canada

<table>
<thead>
<tr>
<th></th>
<th>FY1</th>
<th>FY2</th>
<th>FY3</th>
<th>FY4</th>
<th>FY5</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Highly Qualified RDM Personnel</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network of Expertise</td>
<td>$1,099,150</td>
<td>$2,085,493</td>
<td>$3,375,519</td>
<td>$3,666,427</td>
<td>$4,297,305</td>
<td>$14,523,893</td>
</tr>
<tr>
<td>Contracted Experts</td>
<td>$295,000</td>
<td>$295,000</td>
<td>$295,000</td>
<td>$295,000</td>
<td>$295,000</td>
<td>$1,475,000</td>
</tr>
<tr>
<td><strong>SUBTOTAL</strong></td>
<td>$1,394,150</td>
<td>$2,380,493</td>
<td>$3,670,519</td>
<td>$3,961,427</td>
<td>$4,592,305</td>
<td>$15,998,893</td>
</tr>
<tr>
<td><strong>Tools and Platforms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDM Tools</td>
<td>$730,000</td>
<td>$892,050</td>
<td>$1,098,129</td>
<td>$1,360,271</td>
<td>$1,693,822</td>
<td>$5,774,271</td>
</tr>
<tr>
<td>RDM Platforms</td>
<td>$2,196,500</td>
<td>$2,248,999</td>
<td>$2,055,655</td>
<td>$2,002,312</td>
<td>$2,310,192</td>
<td>$10,813,658</td>
</tr>
<tr>
<td>Archival Storage</td>
<td>$4,320,000</td>
<td>$3,720,000</td>
<td>$2,677,500</td>
<td>$1,792,500</td>
<td>$2,595,000</td>
<td>$15,105,000</td>
</tr>
<tr>
<td><strong>SUBTOTAL</strong></td>
<td>$7,246,500</td>
<td>$6,861,049</td>
<td>$5,831,284</td>
<td>$5,155,083</td>
<td>$6,599,014</td>
<td>$31,692,929</td>
</tr>
<tr>
<td><strong>Secretariat Office</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HQP</td>
<td>$849,250</td>
<td>$946,413</td>
<td>$987,795</td>
<td>$1,031,377</td>
<td>$1,077,295</td>
<td>$4,892,130</td>
</tr>
<tr>
<td>Operational Expenses</td>
<td>$344,000</td>
<td>$237,240</td>
<td>$251,348</td>
<td>$266,386</td>
<td>$282,418</td>
<td>$1,381,392</td>
</tr>
<tr>
<td><strong>SUBTOTAL</strong></td>
<td>$1,193,250</td>
<td>$1,183,653</td>
<td>$1,239,144</td>
<td>$1,297,762</td>
<td>$1,359,713</td>
<td>$6,273,522</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>$9,833,900</td>
<td>$10,425,195</td>
<td>$10,740,946</td>
<td>$10,414,272</td>
<td>$12,551,032</td>
<td>$53,965,344</td>
</tr>
</tbody>
</table>
To this end, given the diversity of the Canadian RDM landscape, and the ongoing efforts by individual universities to develop and deploy RDM services, RDMC’s budget would clearly need to consider, intersect with, and support these efforts. For example, over time, each university will adopt and deliver its own unique suite of RDM services. RDMC would ensure that these local practices are respected, and that its own services and resources augment them, where appropriate, exploiting the richness of this federated approach at the national level.

3.5 Key Success Factors

Possible quantitative metrics as signs of success:
1. Number of RDMPs published/deposited in VP Research Offices.
2. Number of institutional RDM strategies published.
3. Number of datasets made accessible from Canadian research projects.
4. Number of Canadian research software projects funded to integrate best practices and standards into research platforms.
5. Number of researchers and RDM practitioners who take RDMC-approved training for RDM.
6. Number of domain Champions linked to RDMC initiatives.
7. Number of software platforms and services developed to facilitate interoperability of research software, as well as integration with international best practices.
8. Number and value of funded RDM projects.
9. Volume of Canadian research outputs in national, federated archival storage.
10. Number of international RDM agreements entered into.

Possible Qualitative indications of success:
1. Canadian researchers, innovators, and entrepreneurs are making effective use of public research data.
2. The RDMC team intersects effectively with all public research funders, research organizations using public funds, and organizations supporting these two communities.
4. Research Management Information

This report deals primarily with research data and research data management. However, there is an important, if somewhat invisible, aspect of research-related digital activity and its associated infrastructure that is increasingly important for an effective research ecosystem: research management information (RMI). RMI deals with the activities and workflows associated with effective management of all aspects of the research lifecycle, including applications for research support, research evaluation, managing research funds, research reporting, research dissemination, monitoring compliance, and outcomes and impact assessment.

There is an overlap of RMI and research data through shared terminology and unique identifiers.

Research Data metadata is a critical enabler of the FAIR data principles. The metadata that is directly managed and collected along with the core research data (e.g. experiment temperatures, dataset versions, etc.) is well covered in various parts of this document. But there is “another data set” that is equally critical, but often mis-labelled as “metadata”: RMI which contributes to a fuller understanding of any particular research data. Examples of RMI include: RDM Plans, CV information (including ORCIDs) about participating researchers, profile information about sponsoring organizations, etc.

---

This report deals only briefly with the issues for Canada in enhancing and sustaining an efficient and effective RMI infrastructure, recognizing that further review and discussion will likely be required.

4.1 Why is RMI important and how does it lead to greater research impact in Canada?

RMI is critical to the productivity and performance of any research-producing nation. Policy makers, research funders, publishers and research-performing universities are faced with pressures for increased return on investment, transparency, and accountability, all with increased efficiency and minimization of administrative burden on researchers and universities. Strengthening the infrastructure for RMI underpins improvements in the policy and administrative functions in the research ecosystem.

Funders and research-performing universities share a need to exchange digital RMI efficiently and effectively, but lack an agreed approach to the meaning of information elements that ensures everyone interprets the exchanged content in the same way. Internationally, there is a drive for more research on research, essentially the study of research systems—sometimes called the “science of science policy.” But we lack meaningful and comparable evidence to address the questions that are being asked (e.g. “What should we fund?”). Effective operation of the often contentious research assessment landscape requires comprehensive RMI systems in which statistical and qualitative evidence of performance in research and research impact can be recorded in a way that is reliable and comparable. However, we have yet to produce such information in a way that can be shared across platforms. This requires unique identifiers, defined information standards, agreed semantics, and open RMI processing methods: that is, the digital infrastructure for RMI.

In Canada, significant gaps exist in our ability to efficiently harness, understand, and exploit management information about the complete research lifecycle. There are also major gaps in the ability of various end-users to exchange management information efficiently and confidently using existing tools like web forms and PDF reports.

While the actual use of research management information is the responsibility of the various organizations in the research ecosystem (research-funding and research-performing organizations as well as publishers and other research service providers), we lack the cross-cutting, open standards infrastructure that would optimize productivity, efficiency, accountability, integrity, and transparency.

The practice of research management involves a wide spectrum of individuals, disciplines, jurisdictions (national, regional, international), organizations (research-performing, research-funding, research-publishing) and functional units within organizations (e.g. finance, research

---

27 Open standard: An agreed and documented specification that is made irrevocably available on a royalty-free basis with no constraints on reuse. Such standards are developed and maintained under the governance of a not-for-profit organization on the basis of an open decision-making procedure available to all affected parties. There are many kinds of standards (technical, business, qualitative, process, etc.), multiple approaches to how standards are developed and set (formal, informal, de facto, etc.), and many possible standards jurisdictions (national, international, industry/sector, etc.). A specific proprietary product or service may be inaccurately referred to as a "standard" due to its common and widespread use or adoption.
services, library). It is an increasingly important and distributed activity for ensuring the smooth, efficient and measurable functioning of the research lifecycle. Coordinated action among these diverse and independent stakeholders is required to enhance the capacity for transparent development and maintenance of open standards, to improve interoperability of systems and platforms, to strengthen international linkages, and to promote innovative implementations of RMI.

It is, however, also important to differentiate the practice of research management from the development and maintenance of the digital infrastructure for RMI that underpins reliable, comparable, and efficient information usage and exchange. Delivery and maintenance of that infrastructure has national and international implications, given that we need an open standard vocabulary that we can match to at a national and ideally at an international level.

The prime driver and “integrator” of the RMI infrastructure in Canada is CASRAI (Consortia Advancing Standards in Research Administration Information). It is a Canada-based, international non-profit initiative, currently with chapters in Canada, the UK and Europe, dedicated to adaptation and adoption of the principles and best practices of open standards and data governance, as applied to research management information (RMI). It leads and facilitates key RMI end-users and their suppliers in annual deliberations to develop “standard information agreements” that serve as bridges between RMI stakeholders.

CASRAI has recently been endorsed by the Board of the Canadian Association of Research Administrators (CARA) as being best positioned to lead and facilitate discussions around infrastructure and coordination requirements unique to RMI. In Canada, CASRAI has developed a platform and approach to solving the RMI challenges, and brought together the key Canadian partners including CARA, CAUBO (Canadian Association of University Business Officers), CAGS (Canadian Association of Graduate Studies), CARL (Canadian Association of Research Libraries), CAREB (Canadian Association of Research Ethics Boards), and Polytechnics Canada.

Recognizing that dialogue, learning, and exchange across different systems are important, and in addition to the spontaneous growth of CSARAI chapters in the UK and Europe, CASRAI links with other international organizations and communities of practice that deal with research RMI and RMI systems, including the following.

- ORCID International – the not-for-profit organization managing the unique identifier for researchers.
- euroCRIS, an international not-for-profit association that brings together experts on research information in general and current research information systems (CRIS) in particular.
- VIVO, a member-supported, open-source software and an ontology for representing scholarship. VIVO supports recording, editing, searching, browsing, and visualizing scholarly activity.
- CODATA (ICSU), the Committee on Data of the International Council for Science, which promotes global collaboration to improve the availability and usability of data for all areas of research.
- RDA, the Research Data Alliance dedicated to building the social and technical infrastructure to enable the open sharing of data.
4.2 The Vision for RMI in Canada

Our vision is for responsive services, founded on high-quality and comparable research management information, that enable Canada’s research and research management communities to optimize resources and productivity, enhancing the Canadian research enterprise.

4.3 The Future We Seek

In 10 years, with appropriate strategic planning, stakeholder engagement, and investment, we see a future in which the productivity, efficiency, accountability, integrity and transparency of the research lifecycle is increased through system-wide use of high-quality and comparable research management information grounded in principles and lessons learned from good data/information governance and open standards. The Canadian DRI enterprise utilizes and supports the activities and outputs of CASRAI and her Canadian and international partners.

4.4 Principles

The following principles are vital to achieving the vision and desired future state of RMI-enabled services and platforms, developed and delivered locally, regionally, and nationally, and aligned with international efforts:

- **Responsive** to the needs of end-users: researchers, research-performing organizations, research funders, and their service providers; research publishers and systems suppliers;
- **Adaptive** to changes in research and research management;
- **Innovative** through the continuous evolution of RMI open standards that integrate national and international best practices;
- **Collaborative and cooperative** working relationships in the utilization of RMI among the stakeholders in the research enterprise;
- **Well-governed** to be accountable to the research community, to those providing services and resources, and to those funding it; and
- **Sustainable** to allow its ongoing operation.

The LCDRI also supports the RDM principles that have been developed by RDC. They can be found in Appendix B.

4.5 Goals

To achieve the vision and desired future state, the following goals must be pursued collectively and rigorously by stakeholders (e.g. research-performing organizations, research funders, science & technology policy organizations, research publishers) within the Canadian research enterprise:

1. Strengthen Canadian capacity to develop and maintain open RMI standards and to contribute to and learn from international efforts.
• Identify and secure sustained and predictable funding for the operational requirements of the standards development body, including the ongoing activities of the national coordinating body.
• Partner with entities that are developing and implementing RMI services and platforms.
• Coordinate Canadian open RMI standards activity with that of other nations.

2. Extend utilization of open RMI standards.
• Demonstrate the benefits of effective integration of RMI at university and inter-university/inter-sectoral levels.
• Promote widespread adoption of open RMI standards.

3. Extend the community of practice involved in the development, evolution, and implementation of open RMI standards.
• Increase the capacity of research-performing and funding organizations to integrate best practices in the use and collective governance of RMI.
• Engage Canadian NGO and government agencies as contributing partners in this community of practice.
• Facilitate and encourage Canadian participation in international forums.

4.6 Building and Sustaining RMI infrastructure

While the responsibility for investment in the use of RMI lies with the users themselves, development and maintenance of the infrastructure, including pilot demonstrations and convening the Canadian community, are a shared responsibility. Table 3 details a notional budget for the Canadian component of the RMI; this represents 50% of the projected RMI infrastructure costs for Canada (i.e., 1:1 matching) requested from the federal government.

<table>
<thead>
<tr>
<th>Table 3. Proposed Five-Year Federal Contribution to RMI Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>FY1</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Staffing (3 PY) – Canadian component</td>
</tr>
<tr>
<td>Communications, online services</td>
</tr>
<tr>
<td>Convening and standards development/maintenance</td>
</tr>
<tr>
<td>Implementation pilots and demonstrations</td>
</tr>
<tr>
<td>International liaison</td>
</tr>
<tr>
<td>SUBTOTAL</td>
</tr>
</tbody>
</table>
For the purposes of this report, we focus on the business of building and maintaining the RMI digital infrastructure, rather than the various uses of it. Figure 5 captures the major elements of the process.

5. Conclusion

In conclusion, data management is an essential component of Canada’s DRI ecosystem and the LCDRI believes that investment in it is critical to the success of Canada’s researchers in today’s digital environment. We further believe there could be potentially serious consequences to inaction in this area including:

- a failure to fully realize the full potential of investments in Canadian research, as data that is produced would not be broadly accessible;
- inefficiencies in the delivery of DM that create unnecessary barriers for researchers in finding, accessing, and reusing the data of others;
- Diminished capacity to build a critical shared national culture of data management among researchers;
- lost opportunity for ground-breaking discovery; and
• a significant risk to the ongoing capacity of Canadian researchers to participate in international research collaborations or for their work to benefit from the data collected by their colleagues elsewhere in the world.

The federal contribution that we have proposed to support sustainable and broader-based development of DM in Canada would maximize our collective investments by enabling us to coordinate and leverage the significant funding commitments and talents of Canada’s universities in this area and ensure that all researchers across Canada have access to DM services and platforms, regardless of their discipline, geographical location, or size of their institution.
Appendix A: Core Functions Based Upon Data-Related Activities

Plan

Activities: identifying resources, expertise, and services required to develop, manage, and share high-quality data; planning for resource, time, and cost management; preparing, arranging, identifying, and planning for the deposit of data so that it is discoverable in a data repository and reusable; developing and implementing policies at universities and within domains; reviewing policies for compliance at project level; determining data management requirements from the research design; identifying data management practices and workflows; planning consent for sharing.

- Policies: guidelines and mandates for DM, and data collection and deposit;
- Standards and Protocols: international standards for consistent and machine-actionable data management plans (DMPs);
- Processes and Procedures: university processes that intersect with national and domain mandates;
- Leadership, Advice, Support, and Training: DMP tool training, DM best practices;
- Tools and Platforms: DMP tools (Portage DMP Assistant, DFO DMP Template), university services and resources.

Create

Activities: identifying, acquiring and creating (i.e., experiment, observe, and measure) data for the research project; generating corresponding metadata; consulting policies that structure and define procedures for collection data, such as ethics procedures and approval.

- Policies: intersection of local DM practice with international and domain expectations;
- Standards and Protocols: file formats, metadata schemas, interoperability services and protocols;
- Processes and Procedures: data quality and integrity, multiple copies/versions of appropriate data (data provenance);
- Leadership, Advice, Support, and Training: data quality and integrity training, domain-specific approaches, training and events;
- Tools and Platforms: research software (i.e., Virtual Research Environments (VREs), Science Gateways, e-Science platforms), data sharing and transformation during creation.

Process

Activities: preparing research data for analysis; transforming and manipulating outputs; documenting workflow and processing procedures; consulting expertise in data wrangling and statistics, as well as policy and procedures in domain-specific instances.

- Policies: articulating and promoting domain policies to inform data processing;
- Standards and Protocols: accessing requirements for the use of tools further down the lifecycle; adherence to domain-specific ontologies; standardizing workflows;
Processes and Procedures: check, validate, clean, describe, transform, aggregate, manipulate, reduce, and anonymize;

Leadership, Advice, Support, and Training: best practices for data sharing and tools;

Tools and Platforms: re-usable workflows (e.g. Taverna, Galileo, Kepler), code-reuse and management (e.g., GitHub, Jupyter).

Analyze

Activities: describing, comparing, interpreting, and modelling patterns within the data; deriving variables and data; consulting general and domain expertise for statistical or research methodologies; accessing policies and procedures in domain practices.

Policies: Creation and promotion of domain policies to facilitate analysis, outputs, data linking, reproducibility, and privacy;

Standards and Protocols: code and process documentation (e.g. GitHub, Jupyter, domain code repositories);

Processes and Procedures: code reuse; software containers and reproducibility; re-usable workflows;

Leadership, Advice, Support, and Training: promotion and training on use of data software, data modeling;

Tools and Platforms: access to specialized computing resources via HPC/cloud services (e.g., Compute Canada, Microsoft Azure).

Disseminate

Activities: sharing data; transferring data from project to repository; consulting policy for data deposit agreements, licenses, access (conditions for reuse), and preservation; promotion and discovery; consulting expertise in data curation, and submission information package management (i.e., formats, identification of files and types of files); ensuring complete metadata.

Policies: national policy framework as reflected in university policies, ethics/privacy, publisher policies; intersections with university strategies and policies;

Standards and Protocols: consistent use of persistent identifiers (PIDs), file format best practices, ensuring sustainable and international interoperability;

Processes and Procedures: inclusion of all components of reproducibility (data codebooks, software code, system details and containers);

Leadership, Advice, Support, and Training: creating and promoting best practices for data sharing and reproducibility;

Tools and Platforms: repositories and data sharing platforms (e.g., FRDR, Dataverse, Canadian Astronomy Data Centre), metadata creation and quality assurance.

Preserve

Activities: preparing, enhancing, and storing data, metadata, documentation, and code for long-term access and reuse; migrating data to best-practice formats and media; storing data and defining backup strategies; applying best practices for digital preservation processing (e.g., file characterization and normalization); developing and promoting policies for trusted digital repositories (TDRs) for research data; promoting and providing expertise for digitization, digital preservation, and metadata development.

Policies: implementing university and national preservation policies as well as TDRs, reviewing and maintaining deposit agreements/mandates;
- **Standards and Protocols**: defining policies for multiple/federated copies, Open Archives Initiative-Object Reuse and Exchange (OAI-ORE), consistent use of PIDs;
- **Processes and Procedures**: policy-based storage; file transformation and normalization; integrity checks; tracking reuse;
- **Leadership, Advice, Support, and Training**: training on best practices; promoting and mandating data preservation;
- **Tools and Platforms**: use of university, national, international, and domain platforms; federating to a national data platform.

### Reuse

**Activities**: combining existing datasets to create new data for research and analysis; ensuring data discovery, access, and reuse of existing data (secondary data analysis); resources to perform reuse; consulting policies concerning attribution, provenance, and licensing; consulting expertise on data wrangling, search skills, and secondary analysis skills.

- **Policies**: national and domain data-sharing and deposit policies; emerging policies in reproducibility; legislative frameworks, including data-deposit legislation similar to book deposit;
- **Standards and Protocols**: endorsing and using FAIR Principles (findable, accessible, interoperable, and reusable); open peer review;
- **Processes and Procedures**: ensuring consistent and detailed documentation throughout lifecycle; use of open Lab Books/Open Science platforms;
- **Leadership, Advice, Support, and Training**: coding packaging practices;
- **Tools and Platforms**: use of software repositories such as GitHub (including encouraging forking and reuse); reusability platforms and repositories.

### Store

**Activities**: reading (retrieval) and writing (access/placing/putting) digital content to a variety of physical media; making content available for different purposes across the life cycle: copies for active use (use of digital content within research currently being conducted), archival use (long-term access by protecting the digital integrity of content for current and future uses), and dissemination.

- **Policies**: university privacy and security policies; national and domain data-sharing/deposit policies; appropriate authentication and authorization policies; appropriate policies for active and archival storage (including machine-actionable policies);
- **Standards and Protocols**: open standards for data storage and retrieval (e.g., SWIFT, ORE, other accessible APIs);
- **Processes and Procedures**: integration with desktop environments and where researchers work (e.g., mobile platforms), file synchronization;
- **Leadership, Advice, Support, and Training**: defining and promoting appropriate storage timelines as per local, national, international, and domain data-governance practices;
- **Tools and Platforms**: use of open and sustainable storage platforms (e.g., OpenStack, FRDR Globus, Centre for Open Science Open Science Framework, and various domain services).
Discover

Activities: mobilizing, locating, interpreting, searching, assessing, and enabling new and existing data and materials.
- Policies: attention to FAIR (findable, accessible, interoperable, and reusable), especially to ensure open accessibility to metadata from all stages of the research lifecycle, and including that derived from research data and information;
- Standards and Protocols: rich metadata schemas and ontologies; cross walking between different schemas and a common core schema; adherence to international repository standards for metadata (e.g., SHARE data model); relevant harvesting protocols (OAI-ORE); adoption of PIDs (e.g., Digital Object Identifier (DOI), Open Researcher Contributor ID (ORCID), Geospatial Repository for DM Systems (GRID)); ensuring harvest of all types of metadata (e.g., descriptive, administrative, provenance, rights);
- Processes and Procedures: intersection with regional, national, international, and domain repositories to facilitate broad discovery; use of registries to facilitate a federated approach;
- Leadership, Advice, Support, and Training: development of services based on broad use cases; development of training modules highlighting discovery services and approaches;
- Tools and Platforms: SHARE and other Open Science Foundation platforms; various Canadian platforms (e.g. Abacus, Scholar’s Portal Dataverse, Federated Research Data Repository (FRDR)); specialized discovery layers.

Document and Curate

Activities: describing the context, workflow, and data; using appropriate metadata standards to provide rich descriptions; applying appropriate levels of metadata that can be applied/described, (e.g., study-level description, sample description, variable description, instrument description); describing, identifying, and explaining the data, materials, and workflow to preserve.
- Policies: FAIR and appropriate national data-management policies, as well as journal and domain-specific policies;
- Standards and Protocols: adherence to best-practice metadata standards (e.g., FAIRsharing.org); coding best practices (e.g., GitHub, W3C PROV provenance specifications);
- Processes and Procedures: ensuring availability of metadata enrichment (automated and human) at all stages of the lifecycle; use of artificial intelligence, machine learning, and deep learning to enhance metadata;
- Leadership, Advice, Support, and Training: inclusion of DM personnel in university and research project contexts; integration of standard terminology, such as CASRAI RDC RDM glossary to facilitate training;
- Tools and Platforms: integration of best practices into all tools and platforms and across lifecycle; use of standard services like a format policy registry (e.g., PRONOM, RDA registries, CARL Portage PREFER project).
Secure

Activities: honouring and anticipating ethical constraints; protecting data in appropriate environments; guarding against unintended disclosure; ensuring provenance and integrity; and providing appropriate access to data.

- **Policies:** university, privacy, and ethics policies; national and international legislative frameworks; deployment and review of university data-management policies;
- **Standards and Protocols:** adherence to international best practices (e.g. Health Insurance Portability and Accountability Act (HIPAA), Federal Information Security Act (FISMA), Shibboleth and Canadian Access Federation (CAF); adoption of other standards such as W3C security standards;
- **Processes and Procedures:** adoption of dynamic security policies and procedures to reflect changes throughout the lifecycle;
- **Leadership, Advice, Support, and Training:** ensuring intersection with privacy offices, university IT security services, and aboriginal communities;
- **Tools and Platforms:** adoption of proven security platforms (e.g. RedCap and Data SHIELD); use of secure facilities such as Canadian Research Data Centres Network (CRDCN).
Appendix B: RDC Principles

As part of their dedication to research excellence and knowledge dissemination, Canadian research universities are committed to meeting the challenges and opportunities involved in the effective management of research data.

1. The Importance of Data for Research: Canadian universities recognize the central role of data in 21st century research. Data are both the product of research and a foundation for future research. The appropriate management of research data and facilitating appropriate access to that data are foundations of modern scholarship and scientific discovery.

2. National and International Collaboration: Canadian universities will aim to meet RDM challenges and opportunities collaboratively and in alignment with international activities in RDM, including the development and setting of standards such that Canadian research data are interoperable with those of global research partners.

3. Access: Research data generated by publicly-supported research are a public good, produced in the public interest, and will normally be made openly available in a timely and responsible manner with as few restrictions as possible. Rights to publicly supported research data will not normally be assigned to others without ensuring that the data remain accessible and available for reuse.

4. Ethical, Legal, and Privacy Issues: Canadian universities recognize that there are privacy considerations, legal concerns, ethical issues, and commercial interests reflected in contractual requirements that may constrain the release of research data. University policies will be developed to accommodate these complexities.

5. Privileged Use: Those who conduct research are entitled to a limited period of privileged use of the data they have collected and generated, for example to enable them to publish the results of their research. Such limited periods may vary in length depending upon the academic discipline involved.

6. Recognition of Intellectual Contributions: University policies and reward systems will recognize the intellectual contributions of researchers who generate, preserve, and share research data. Users of already generated research data are obliged to acknowledge the source of their data and abide by the terms and conditions under which they are accessed.

7. A Public Trust: As research data constitute a public good, appropriate management of such data constitutes a public trust. It is recognized that while RDM may be the primary responsibility of the principal investigator during the life of the research project, long term preservation is only possible with appropriate funding from funding agencies and governments. Resources will be commensurate with the original support provided to the research, including training and expertise.

---

8. Data Management Plans: University and project-specific data management plans typically follow recognized, relevant international standards and community best practices. Such plans should recognize that data may be of potential long-term value, sometimes for purposes distinct from those for which the data were created, and will require plans and resources for preservation and access. Decisions about the length of time for data preservation should be based on policies which recognize the potential long-term value of research data.

9. Metadata and Discoverability: Metadata will normally be recorded and made openly available in an internationally recognized standard. This will enable research data to be discoverable and effectively reused by others. Published results will include information on how to access the data on which the results are based. If the data cannot be or is not yet available (see principles 4 and 5), the metadata may be published in order to alert potential users to the existence of such data.

10. Multilingual Access: Canadian researchers and universities may employ tools for data management in either French or English, ideally both.