Research Data Management in Canada: A Backgrounder

July 18, 2019

Based on a report produced for Innovation, Science and Economic Development (ISED)

by the Canadian Association for Research Libraries (CARL), Consortia Advancing Standards in Research Administration Information (CASRAI), Leadership Council for Digital Research Infrastructure (LCDRI), Research Data Canada (RDC).

Contributors: David Baker, Donna Bourne-Tyson, Laura Gerlitz, Susan Haigh, Shahira Khair, Mark Leggott, Jeff Moon, Chantel Ridsdale, Robbin Tourangeau, Martha Whitehead

DOI: 10.5281/zenodo.3341596

Table of Contents

Introduction

Digital research infrastructure is transforming the practice of research by enabling the rapid creation of massive quantities of data at an explosive rate. We are only beginning to understand its total impact on the research process, but we recognize the changes it brings fundamentally reshape the speed at which our researchers work, the questions they can ask, and the results they can achieve. To fully comprehend the 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. This DRI must allow Canadian researchers to store, access, reuse, and build upon digital research data that 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 nonprofit and private sectors in helping them advance critical social and commercialization goals, though for it to be of any use it must be managed appropriately. The systematic management of research data is an essential component of the DRI ecosystem, and Canada has begun to lay the supportive foundations to enable researchers to thrive in this ecosystem through the development and delivery of high quality, equitable, and effective data management (DM) services and platforms. However, for Canada to become a global leader in data management and ensure we are continuing to produce world-class research in a changing research landscape, we must continue to build on those DM foundations.

This report provides an overview of the DM environment in Canada, and identifies key challenges to achieving this goal, and will be updated annually to communicate changes in the landscape.

Definitions

Advanced Research Computing provides researchers with digital technology and expertise to help them solve research issues that are either too large or too complex for them to undertake on their own. It includes services, advice, hardware and software, all supported by highly qualified personnel (HQP), to enable research activities with significant data or computation requirements, including data acquisition, simulation, experimentation, analysis, and exploration. It is known by a number of different names, depending on the audience. Examples include high performance computing (HPC), cyberinfrastructure, and supercomputing.[1]

Data[2] 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.

Data Management are the activities of 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.

Digital Research Infrastructure[3] are those layers that sit between base technology (a computer science concern) and discipline-specific science. The focus is on value-added systems and services that can be widely shared across scientific domains, both supporting and enabling large increases in multi- and interdisciplinary science while reducing duplication of effort and resources (e.g., including hardware, software, personnel, services and organizations). In Canada, the preferred term has become Digital Infrastructure to refer to what is also known as Cyber-Infrastructure or e-Research Infrastructure.

Research Data (RD) 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

[1] The Leadership Council for Digital Research Infrastructure, “Advanced Research Computing (ARC) Position Paper: For Innovation, Science, and Economic Development Canada,” The Leadership Council for Digital Research Infrastructure, August 31, 2017, 5-6.
[2] The following definitions were developed through consultation with the CASRAI IRiDiuM 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.
[3] This definition is taken from CASRAI's term for Digital Infrastructure; “Digital Infrastructure - CASRAI Dictionary,” CASRAI, August 13, 2015, https://dictionary.casrai.org/Digital_infrastructure.

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.

Research Data Management/Data Management (RDM/DM) refers to the storage, access and preservation of data produced from a given investigation. Data management practices cover the entire lifecycle of the data, from planning the investigation to conducting it, and from backing up data as it is created and used to long term preservation of data deliverables after the research investigation has concluded. Specific activities and issues that fall within the category of data management include: File naming (the proper way to name computer files); data quality control and quality assurance; data access; data documentation (including levels of uncertainty); metadata creation and controlled vocabularies; data storage; data archiving and preservation; data sharing and reuse; data integrity; data security; data privacy; data rights; notebook protocols (lab or field).

Metadata is literally "data about data". It is data that defines and describes the characteristics of other data, used to improve both business and technical understanding of data and data-related processes. Business metadata includes the names and business definitions of subject areas, entities and attributes, attribute data types and other attribute properties, range descriptions, valid domain values and their definitions. Technical metadata includes physical database table and column names, column properties, and the properties of other database objects, including how data is stored. Process metadata is data that defines and describes the characteristics of other system elements (processes, business rules, programs, jobs, tools, etc.). Data stewardship metadata is data about data stewards, stewardship processes and responsibility assignments.

Background to the Document

The original text for this document was based on the submission from the Leadership Council for Digital Research Infrastructure (LCDRI) to Innovation, Science and Economic Development Canada (ISED). Some of the material was removed (e.g. recommendations to ISED) and other sections were updated and expanded on in creating this version. This document was also reviewed and edited by a broad cross-section of the RDM stakeholder community, and was

done to reflect the input and interests of the full Canadian RDM community. The intention going forward is to update this document annually.

The Research Lifecycle and RDM Functions

Fig. 1. Data-Related Activities During the Research Process. (Created by the Leadership Council for Digital Research Infrastructure. In *Advanced Research Computing (ARC) Position Paper: For Innovation, Science, and Economic Development Canada*. Leadership Council for Digital Research Infrastructure. Unpublished manuscript. August 31, 2017, 5.)

The Research Lifecycle serves as a roadmap for researchers to understand what considerations they need to make for their data at every stage. Also at each stage are five high-level points in which DM actions fall: Policies, Standards and Protocols, Policies and Procedures, Leadership, Advice, Support and Training, and Tools and Platforms.

Plan

The Plan phase of the lifecycle is the stage at which the researcher organizes themselves and their data for discovery, reuse and archiving further along in time. Ideally they should acquaint

themselves with DM guidelines and mandates relevant to their funding or postsecondary institution or other organization, and identify appropriate standards and protocols that follow best practices for DM in their organization or domain. This includes the creation of a data management plan (DMP), and determining an appropriate repository for data storage and archiving.

This is the perfect time for the researcher to seek the assistance of RDM experts who can provide guidance in making these decisions. These experts can offer support in the form of DM training, clarify university (or other organization) processes that may intersect with national or domain mandates, and offer guidance on planning for the depositing, sharing and reuse of data.

Create

The Create stage involves the identification, acquisition and creation of research data and metadata. Unsurprisingly, it is at this stage that researchers must be aware of any institutional or domain-specific policies that define procedure for data collection. DM enters this phase in regards to best practices for data quality and integrity, versioning, and provenance; said practices at the local level often intersect with the international level. Best practices for metadata also must be observed at this stage to ensure interoperability and discovery, and occurs through the use of schemas and protocols. DM personnel and institutions can offer training and events in the realms of data quality and integrity, and focus on domain-specific approaches. Data can be shared and transformed during its creation using research software platforms such as Virtual Research Environments, Science Gateways and e-Science platforms.

Process

In the Process stage, data is prepared for analysis (checking, validating cleaning, describing, and so on); part of this process involves ensuring domain-specific ontologies are followed. Code also must be managed at this stage so that it may be discoverable and reused, often through platforms like GitHub and Jupyter. The standardization of workflows are another consideration, as well as documenting every process. DM experts can assist in identifying tools that can assist with all of these tasks (for example, tools that can be used for the reuse of workflows like Taverna, Galaxy, and Kepler), and help with domain-specific policies and procedures in (meta)data wrangling.

Analyze

Analysis of data naturally follows preparation and processing. At this stage, code and workflow management and process documentation is still important, along with the creation and promotion of domain policies that facilitate analysis, outputs, data linking, reproducibility and privacy. DM experts, again, can offer guidance in these areas, as well as training on the use of data software and modeling. Specialized computing resources such as high performance

computing and cloud services offered by Compute Canada, Amazon and Microsoft Azure may also be required by the researchers.

Disseminate

The Dissemination stage is the stage in which the majority of sharing occurs. Before data can be transferred to a repository deposit agreements, licensing, conditions for reuse (or access), and methods of discovery and preservation are considered. National policy framework is reflected in university policies (e.g. regarding ethics and privacy) and typically intersects with publisher policies and greater university strategies. Software code and codebooks, and other kinds of system details also need to be made available so that research may be reproduced.

Best practices to ensure sustainability, interoperability, discoverability and reuse must be followed. This includes the use of persistent identifiers for data, appropriate file formatting, and complying with international practices. Repositories and other data sharing platforms, such as Federated Research Data Repository (FRDR), Scholars Portal Dataverse and the Canadian Astronomy Data Centre are useful resources as they ensure metadata creation and quality assurance. DM experts may offer support to researchers by both creating and promoting best practices for sharing and reproducibility, and offering consultation on (meta)data curation.

Preserve

Preservation is the second last stage in the Research Lifecycle, and involves the process of moving data from an active to an archival state. In order to protect data, long-term university and national preservation policies (that often reflect or intersect with international guidelines) must be implemented. Data, metadata, documentation, coding and all back-up copies must be prepared for long-term access and reuse; in some cases data needs to be migrated to more preservation-friendly formats. Other kinds of digital preservation processing include file characterization and normalization. Trusted Data Repositories are excellent services that have undergone certification to prove to the research community the repository's digital infrastructure is trustworthy and sustainable, offering a platform in which data can be stored and accessed long term. At this level DM experts may offer training on best practices for archiving and digital preservation, and review and implement data deposit agreements or mandates.

Reuse

Reuse of research is the final stage in the Research Lifecycle, and, regarding data, involves ensuring discoverability and access to data so that it may be combined to form new datasets, and referenced or analysed by other researchers. At the highest level are national, international and domain policies and legislative frameworks focused on sharing and deposit. Following the FAIR[4] Principles at this stage allow for ease of data reuse, as do tools that allow the reuse of

[4] Findable, Accessible, Interoperable, Reusable. https://doi.org/10.1038%2Fsdata.2016.18

documentation (Lab Books) and software or coding (GitHub). DM Experts can continue to support with data wrangling and understanding policies surrounding attribution, provenance and licensing, as well as searching and secondary analysis.

Store

The storage of data differs depending on the Research Lifecycle, and whether or not it is in an active state. Regardless of the state, the data, or at the very least the metadata, should be made to some extent accessible to other researchers. Therefore, storage also includes considerations into the deposit and retrieval of data (often facilitated by open standards such as SWIFT and ORE) into online storage platforms (and when appropriate, physical media). Archival use requires thought towards long term access and protecting the digital integrity of the content, as well as dissemination. University, national and domain policies regarding privacy and security and data-sharing must be taken into account when determining access.

In order for (meta)data to be accessible long term storage platforms should be open and sustainable, and there currently exists a number of options: OpenStack, FRDR Globus, Centre for Open Science Open Science Framework, and other domain services). DM experts can provide knowledge into domain-specific services, and in defining appropriate storage timelines as per local, national, international and domain data-governance practices. Experts should also consider the integration of desktop environments and processes like file synchronization to ease the task of data storage for researchers.

Discover

Researchers should strive to make their data discoverable at all stage of the Research Lifecycle, with ‘discover’ in this context relating not only to searching for data, but to the mobilization, location, interpretation, and assessment of it. This, in turn, allows fellow researchers to compile and create new (meta)data. There are a number of best practices that should be considered that will lead to high quality discoverability later in the Research Lifecycle: deciding upon appropriate metadata schemas and ontologies and understanding potential cross walks, considering relevant harvesting protocols for all types of metadata (OAI-ORE), and adopting PIDs (DOI, ORCID). Understanding the repository in which the data will be deposited is also an important step, as this will enable the researcher to be prepared to follow the repository’s standards (such as the SHARE data model), or take into account specialized discovery layers, such as registries that facilitate a federated approach to discover.

Following the FAIR Principles reinforces accessibility and discoverability to metadata from all stages of the Research Lifecycle, and including that derived from research data and information. DM experts can provide guidance on FAIR as well as training in different kinds of discovery services and approaches. Experts should also consider the development of services based on broad use cases.

Document and Curate

Documentation and curation of (meta)data should be planned out early, and occur throughout the Research Lifecycle for maximum interoperability and discoverability. This involves describing the context and workflow surrounding the data - coding and other materials, for example - and using appropriate metadata standards (found through resources such as FAIRsharing.org) to provide rich descriptions at appropriate levels. Of course, the data itself should also be described, identified and explained for preservation purposes.

FAIR and appropriate national data management policies, as well as journal and domain-specific policies should be understood by the researcher, but if they are unfamiliar they can seek out their institution's DM experts or external resources such as the CASRAI RDC RDM glossary to facilitate training. Format policy registries (e.g. PRONOM, RDA registries) also provide valuable standardization services.

Secure

Consent around sharing of data is another aspect of DM that needs to be considered; whether consent or anonymization of data is required, how much or how little of the data can be shared, and ensuring the legal and ethical conditions on the use of the data are followed and integrity and provenance are maintained. Researchers should be prepare to guard against unintended disclosure while also allowing appropriate access to data. Understanding ethics policies at all levels is necessary, and, in many cases, a researcher has domain-specific or international best practices (e.g. Health Insurance Portability and Accountability Act, Federal Information Security Act) that can offer direction. Other standards such as W3C security standards can be helpful in guiding the researcher to success in securing their data. Proven security platforms like RedCap and DataSHIELD, or use of secure facilities like Canadian Research Data Centres Network (CRDCN), are resources that should be highlighted at this stage.

It is particularly important to understand security changes with the Research Lifecycle, and that the researcher adopt best practices and procedures that reflect these changes. Privacy offices, university IT security services, and communities in which privacy and access are of particular importance can provide advice and support in the security of research data.

The Continuum of Research Data Storage

Data repositories are an important bridge between active and archival storage. As illustrated in Figure 1, the three primary types of storage are: active, repository, and archival storage. ARC has a key role in providing the active storage component of this continuum, while DM is more focused on the provision of repository and archival storage. The primary purpose of repository storage is to support data dissemination by ensuring that research data is stored securely and

can be discovered and accessed appropriately. Repositories also serve as an important pathway to archival storage.

Archival storage is intended to preserve a 'copy-of-last-resort' for the long-term. Specifically, archival storage is the function of the archival system which manages the long-term storage and maintenance of the content under the stewardship of the archive. In addition to storage, regular maintenance activities such as format migrations, media refreshment, error checking, and disaster recovery plans are a very important part of this service which help to enable long-term access.

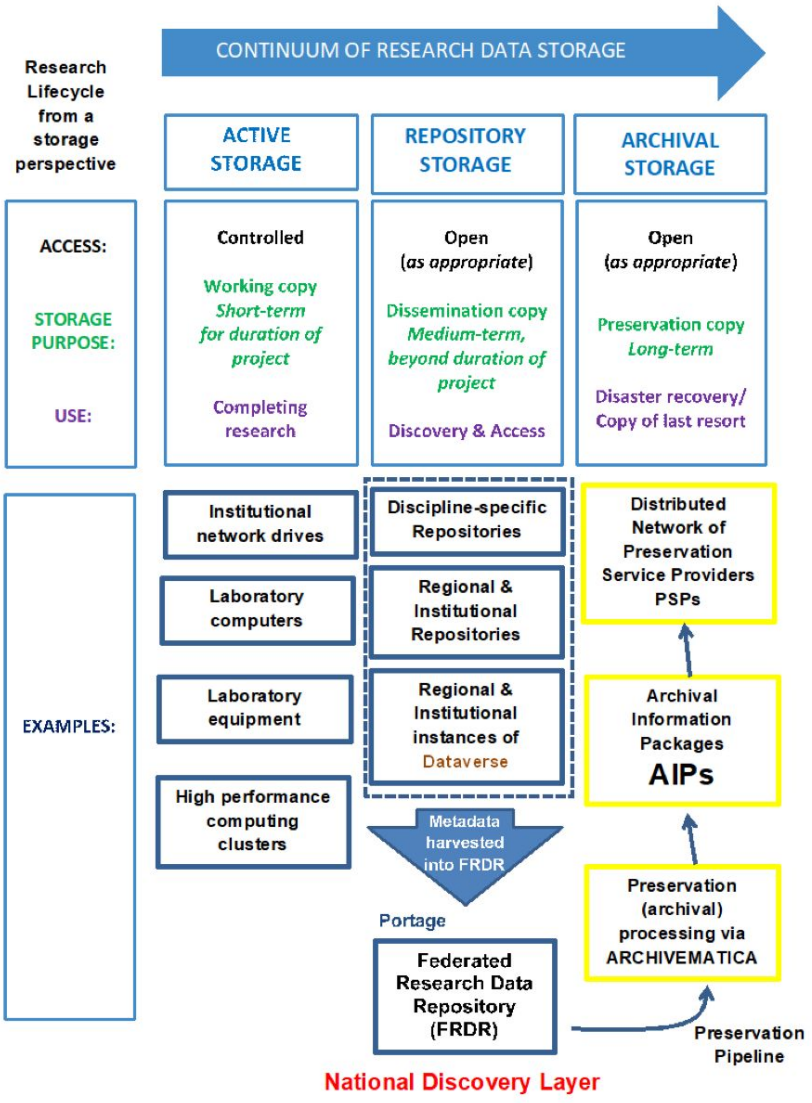

Fig. 2. Continuum of Research Data Storage. (Based on an image created by CARL Portage Network. In *ISED Follow-up Questions on DM*. CARL Portage Network. Unpublished manuscript. 2018, 6.)

A number of factors need to be considered and challenges addressed when implementing a network of repositories.There are different types of repository services and this reflects the diversity of communities using or providing them. Any national framework for a network of repositories must recognize this reality. It must also ensure that repositories are interoperable and federated across Canada, and, ideally, internationally. In addition, the framework must situate repositories as part of a broader national storage strategy that provides access to a continuum of research data storage that is optimized for the specific use case. This would enable shared discovery of Canadian research, while maintaining rich domain-specific views of research data in various disciplines.

Canada currently faces a number of challenges in relation to the development of this essential network of repositories. First, there is a lack of repository storage for researchers to use. For instance, Isaac Tamblyn's Computational Laboratory for Energy and Nanoscience at the University of Ontario Institute of Technology creates 50 TB of data related to molecular structures annually, which is then used as input to machine learning algorithms. The data is stored in a traditional database using Compute Canada resources, but that storage is only accessible on a maximum 3-year allocation: the hunt for funding to maintain the growing data repository illustrates why this is such a critical gap in so many research labs. How does a researcher like Dr. Tamblyn provide open access to this data beyond a 3-year allocation and in a form that can be accessible to researchers and machine processes alike?

Second, in cases where data repositories do exist (e.g., the University of British Columbia's Open Science Framework implementation, McGill's C-BIG repository at the local level; the Ontario Council of University Libraries' Scholars Portal Dataverse platform and the Atlantic Research Data Repository at the regional/provincial level; the Canadian Astronomy Data Centre at the national level; and the Canadian Cryospheric Information Network's Polar Data Catalogue at the international level), they serve their communities well, but operate largely in isolation from one another, causing data to be siloed. One exception to this is the emerging effort to create Dataverse North, which is working "to develop a community of practice that brings together Dataverse providers and libraries to coordinate and discuss local and national training, support services, outreach strategies, promotions, and infrastructure development and needs."[5] Further, without national cost-sharing support and coordination, they are only able to meet the needs of a small percentage of Canadian researchers.

Third, there are also a number of specialized repositories developed by Canadian researchers that offer, or have the potential to offer, access to invaluable collections of data. However, these repositories may be vulnerable because they have been established and maintained by individual researchers without a strong background in RDM best practices, such as proper documentation, backups, and adherence to standards. These collections of data, typically, have no plans in place for long term sustainability, such as "data rescue" initiatives that would migrate the data to more stable, shared platforms.

[5] https://portagenetwork.ca/network-of-experts/dataverse-north-working-group/

The CARL Portage and Compute Canada FRDR system is an example of a recent effort to try and tackle these challenges by aggregating metadata from multiple repositories, facilitating the transfer of 'big data', and ensuring long-term data preservation through Archivematica. While this effort represents an important step in the development of a national federated approach that can be used to fulfill both repository and archival storage needs in a seamless and transparent fashion, stable long-term funding for this initiative is needed if it is to fully achieve its potential for Canada's research community.

The funding that is proposed in the DM position paper would be used to address these challenges by supporting and coordinating the following activities across the country:

1. develop a coordinated and sustainable approach to storage throughout the research lifecycle that recognizes the need for all three types of data storage;
2. develop mutually-accepted standards and protocols to facilitate the flow of data and metadata;
3. lease/license infrastructure, technology platforms, and services necessary to support RDM nationally (e.g. license Globus services - current information technology makes it possible for a national organization, to develop, coordinate, and support comprehensive RDM services in Canada without having to directly own and operate them);
4. implement succession policies to ensure that all repositories are backed up in the event a repository is forced to cease operating;
5. support "data rescue";
6. train researchers on the use of data repositories and related tools; and,
7. given the constant change in the world of standards and storage, provide funds to spur innovation and ensure that the national federated network of repositories is able to evolve with international communities of practice.

In addition, some funding may flow to repositories that provide services to researchers at universities or research centres that do not have a local data repository of their own - not every institution will need a data repository. To maximize efficiencies, we envision an environment in which repository infrastructure and services are shared across institutions and discipline-specific communities, with the potential for targeted repositories eventually taking on different domain specializations.

Additionally, a coordinated approach to the acquisition of storage (cloud-based and local) could benefit from economies of scale. Cloud computing services can be leased from multiple platform providers in customizable configurations, providing a viable and effective means of offering national RDM services. Infrastructure providers can be institutional, regional, national, or international, as well as based on commercial and open source frameworks. 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 and maintain their own tools and platforms. Where needed, it allows institutions to

redirect local efforts towards secure local storage for highly sensitive datasets, where access via the internet is an unacceptable risk.

A number of key, federally-funded data repositories are at risk. These repositories are currently funded through research grants awarded to researchers for specific research projects, leaving data in these repositories vulnerable as operational funding for them disappears at the end of a granting cycle. Stabilizing these repositories by making their current funding sustainable is critical to ensuring that the significant investment that has been made in them is not lost and that important data are protected and available for current and future researchers. Sustainable funds to facilitate the maintenance of these repositories is an important consideration. This would help to ensure that these key repositories are integrated as part of a broader network/community of practice, operating according to international standards and best practices. It would also send an important signal to the international community that Canada is a leader in good public stewardship of data - in some cases, these key repositories are high-profile, internationally-recognized, and essential to maintaining Canada's research reputation on an international stage.

Examples of such Canadian repositories abound. Many are indexed in the international Registry of Research Data Repositories (re3data.org). Notable examples include: CBrain (McGill Centre for Integrative Neuroscience), Canadian Astronomy Data Centre, Polar Data Catalogue, Ocean Networks Canada, Genome Canada, and the Canadian Writing Research Collaboratory.

Canada's research community is very diverse, with different needs and practices. Any solution proposed for archival storage must recognize this reality and develop an approach that will respond to diverse researcher requirements and in which they all can have confidence. An example of this diversity would be the physics or genomics communities in which data storage is part of a broader set of global archival storage practices and systems. It would be very difficult to implement centralized archival storage in this context. In addition, certain communities such as those working in health research, have highly specialized security requirements that may make centralized archival storage difficult to realize. Lastly, researchers are often most comfortable adopting new approaches when the infrastructure and services are close to them. Given that a DM culture is relatively new in Canada, it will be important to establish an environment that is conducive to increasing researcher adoption and confidence. Having the flexibility of local/regional and discipline-specific archival storage solutions is important to realizing this goal.

Cost Effectiveness

Avoiding the prohibitive costs of establishing, operating, and maintaining a centralized national data archive is one of the key benefits of a decentralized-federated approach. Technology and expertise need not be centralized to be efficient and cost-effective, and this is particularly true in the context of data archiving. This view is supported by the Portage Preservation Expert Group's (PEG) White Paper on data archiving. This report highlights that overseeing the

provision of active, repository, and archival storage as part of a federated national storage strategy will introduce substantial efficiencies.[6]

To achieve these efficiencies, it would be desirable for a national coordination of archival storage functions among new and existing organizations with capabilities and the associated expertise and experience to perform these functions effectively. So, rather than imposing homogeneity in archival storage architecture, coordination of a strategic and diverse network of preservation service providers (PSPs), who are federated through a national strategy that identifies gaps and areas of overlap in the delivery of their services, and that defines a set of 'best practice' requirements, while leaving the operation and maintenance of individual PSPs to their host institutions. This approach would leverage existing institutional and organizational capacity, expertise, and investment in support of the broader archival storage strategy for the country. It would also help recognize PSPs that represent best practices in specific domains, which may provide the same services to an international community of researchers. This recognition may also speak to the need for a more sustainable level of support for all Canadian PSPs, national and international.

Risk Mitigation

A national strategy for archival storage would need to recognize the importance of risk mitigation as part of responsible public stewardship. One of the core value propositions of archival storage is its ability to respond to this requirement - a single centralized storage option is antithetical to this. Decentralization fulfils the "lots of copies keep stuff safe" (or LOCKSS) rule and the best practice of using geographically distributed storage locations to ensure data recovery in the event of a disaster. That the proposed decentralized network of archival storage implicitly achieves geographic dispersion of archived data is an important argument in favour of such an approach.

Regional Talent Development

A decentralized-federated approach would help leverage and grow regional and institutional highly qualified personnel (HQP). As discussed above, researchers are most comfortable adopting new tools and approaches when the infrastructure and services are close to them. In this context, having well-trained and knowledgeable front-line service providers to help them with RDM is essential. A strategy that recognizes the need for regional talent development in the area of archival and repository storage would serve three purposes. First, it would help to ensure that researchers have equitable access to RDM expertise across the country, regardless of the location or size of their institution. Second, it would support first-hand knowledge development and skills training for these RDM experts who would work directly with both RDM

[6] Umar Qasim, Corey Davis, Alex Garnett, Steve Marks, and Michael Moosberger, "Research Data Preservation in Canada : A White Paper," R. UBC Community and Partner Publications, Last modified April 30, 2018, https://dx.doi.org/10.14288/1.0371946.

infrastructure and researchers. And third, it would provide an important training and skills development opportunity for undergraduate and graduate students in this growth area.

The Impact of Good Research Data Management

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. Good RDM practices also contribute substantially to the integration of the FAIR Principles into standard research practice.

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.[7]

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.

[7] "Tri-Agency Statement of Principles on Digital Data Management," Government of Canada, last modified December 21, 2016, http://www.science.gc.ca/eic/site/063.nsf/eng/h_83F7624E.html

Growth in Data Production

Science in all areas has become a major producer and consumer of data. Whereas in the past researchers had to contend with the issue of data scarcity, the information age has brought on a data deluge, with data now being generated in unprecedented volumes and variety, and at increasing velocity.[8] Particularly in the physical sciences, satellites and remote sensing tools are being deployed on a global scale that dwarf traditional sampling methods.[9] Additionally, new methods in science such as microarrays, combinatorial chemistry, and sensor networks produce new classes of born-digital data such as workflows, ontologies, supporting code, and other lab materials.[10] Well-funded RDM initiatives that are supported by a model of national coordination will be essential to meeting the myriad of challenges posed by increasingly data-driven research.

In this context, it is important to note that currently only a small percentage of data sets produced by researchers are made available and preserved over the long-term. Therefore, we need to build capacity to both manage what is currently being produced and to address future growth in data production.

Trends Driving RDM Initiatives

Drivers affecting the need for a RDM are both top down, with journal and funder policies requiring data to be made discoverable and, where possible, openly available, and bottom up, as researchers are increasingly identifying the advantages of being able to integrate data from a variety of sources across disciplines. In addition, given the proliferation of fraudulent research and research journals, it has become increasingly important for data to be preserved in order to be able to test outcomes and reproduce results.

Journal publishers

Journal publisher policies are increasingly recognizing the value of data as a standalone research output,[11] with the number of dedicated “data journals” on the rise.[12] The number of

---

[8] Mark Costello, and Edward Vanden Berghe, “‘Ocean Biodiversity Informatics’: A New Era in Marine Biology Research and Management,” *Marine Ecology Progress Series*, 316 (2006): 203-214. https://doi.org/10.3354/meps316203; Whorisky, Fred, and Kes Mortin, “Canadian Integrated Ocean Observing System,” Marine Environmental Observation, Prediction and Response Network, November 30, 2017, http://meopar.ca/uploads/IE_Report_-_Observations_and_Data.pdf.

[9] Costello, and Vanden Berghe, “Ocean Biodiversity Informatics,” 203.

[10] Ross Harvey, *Digital Curation: A How-To-Do-It Manual* (New York: Neal-Schuman Publishers, Inc., 2010), quoted in Alex H. Poole, “How Has Your Science Data Grown? Digital Curation and the Human Factor: A Critical Literature Review,” *Archival Science* 15, no. 2 (June 1, 2015): 102, https://doi.org/10.1007/s10502-014-9236-y.

[11] Buckland notes: “The potentially useful record of science is increasingly not the written reports but (mainly non-textual) digital data sets of many kinds: the raw material, the operations upon it and progressively more refined derivations can be beneficially shared and built upon by other researchers.’’ “Data management as bibliography,” *Bulletin of the American Society for Information Science and Technology* 37, no.6 (2011): 35, https://doi.org/10.1002/bult.2011.1720370611.

journals requiring researchers to deposit data and related materials alongside publications is also on the rise, with some journals, such as PLOS One, refusing to accept submissions that are not supported by the accompanying data. While data reuse is still an emerging paradigm, it can be expected to rise as more research data becomes widely available. One of the challenges with this new approach is that academic journals, many of which are, or are published by, commercial entities, will become the owners of publicly-funded data that they will then be able to use as a very profitable source of revenue generation. This effectively means that public institutions and researchers will need to purchase back publicly-funded data at a potentially high cost.

Funding Agencies

Over the past decade, there has been an international trend among funding agencies and governments toward the development of national RDM policies, and the data services that necessarily follow, in recognition of the need for publicly-funded data to be made discoverable and accessible.[13] Within this global context, in 2016, the Canadian Tri-Agencies released a Statement of Principles on Digital Data Management that outlines funder expectations for RDM, and the responsibilities of researchers, research communities, institutions, and funders in meeting these expectations. The Tri-Agencies are now engaged in a consultation process around a draft data management policy (expected for final release in 2019) that could require "all research data and code that support journal publications, pre-prints and other research outputs that arise from agency-supported research [...] to be deposited in an appropriate public repository or other platform that will ensure safe storage, preservation, curation, and (if applicable) access to the data."[14] Funder mandates for the sharing of research data are expected to strongly impact the demand for robust RDM infrastructure and services.

Researchers

Researchers are also increasingly discovering the need for and advantages of being able to find, access, and reuse data. This, in turn, is driving a requirement for interoperable, sustainable, and agile RDM infrastructure and services. Usage statistics from Canadian repositories underscore this demand. For example, the Scholars Portal Dataverse reported over 100,000 research data downloads[15] over the 5 years of its existence, and UBC's Abacus Dataverse Network reported over 45,306 downloads.[16] According to a recent report by the Portage Data Discovery Expert Group, there are roughly 180 data repositories in operation

---

[12] See Nature's *Scientific Data* journal.

[13] Kathleen Shearer, "Comprehensive Brief on Research Data Management Policies," Portage Network, April 7, 2015, https://portagenetwork.ca/wp-content/uploads/2016/03/Comprehensive-Brief-on-Research-Data-Management-Policies-2015.pdf. Examples include the US National Data Service, the Australian National Data Service, and the UK Data Service.

[14] Matthew Lucas, "Tri-Agency Data Management Policy Initiative," presented at the RDA 10th Plenary Collocated Event, Montreal, Quebec, September 18, 2017, 14, https://portagenetwork.ca/wp-content/uploads/2017/09/TC3-Data-Management-Policy-Initiative.pdf.

[15] As of May 29 2019.

[16] As of May 29 2019.

across Canada.[17] While these repositories are able to serve some of the needs of their respective researcher communities, support for national RDM coordination is needed to create the shared services and infrastructure required to ensure that data do not become siloed across many repository platforms that do not interoperate or conform to international standards and best practices.[18] A recent survey of Canadian researchers found that the majority of engineering/science researchers who participated were already depositing research data in external data repositories, but less than 40% felt their data were sufficiently documented for someone outside their lab to use.[19]

The National Digital Stewardship Alliance (NDSA) recently surveyed 133 institutions engaged in digital presentation activities to investigate how these organizations staffed and organized their digital preservation functions, and to identify any changes since the NDSA's 2012 survey.[20] They found that data holdings had grown since the 2012 survey, and while the majority of respondents expected less than 25% growth of repository holdings over the next year, respondents reported that they require nearly twice as many FTE staff to properly manage current holdings. A recent survey conducted by Portage examining staffing levels for RDM activities in CARL institutions found an average of only 1.42 FTEs directly supporting RDM functions,[21] far below the NDSA average of 13.6 FTEs. Comparing the Canadian FTE levels to the NSDA survey, which looked at organizations primarily in the US, but also internationally, one could conclude that increased investment in RDM initiatives is needed to properly manage existing data holdings, without taking into consideration expected increases in research data deposits from the drivers outlined above.

Another trend driving RDM initiatives is the need to be able to re-access data in order to reproduce and verify results. In general, this is important for good public stewardship of publicly-funded research, but, as mentioned above, this has become increasingly important in the context of a growing proliferation of fraudulent research and research journals.

In much the same way as it would have been difficult for early proponents of the web to quantify its profound impact on all of us, one of the challenges that we have in answering this question is, of course, that Canada's RDM practice is still nascent and our infrastructure/services are not yet fully developed. As such, it is hard to quantify impact in the absence of the metrics that a more mature RDM platform would generate.

---

[17] Berenice Vejvoda, Alison Ambi, Eugene Barsky, Kevin Lindstrom, Heather MacDonald, Kathleen Matthews, Michael Moosberger, et al. "Portage Data Discovery Expert Group - Collections Development Working Group: Phase One Report," July 2017, https://doi.org/10.14288/1.0351978.

[18] See our response to Question 3 for more information on the proposed network of repositories and Appendix 1 for more information on FRDR.

[19] Cristina Sewerin, Eugene Barsky, Dylanne Dearborn, Angela Henshilwood, Christina Hwang, Sandra Keys, Marjorie Mitchell, Michelle Spence, Kathy Szigeti, and Tatiana Zaraiskaya. "From Coast to Coast: Canadian Collaboration in a Changing RDM Seascape." (paper presented at the International Association of University Libraries Conference, Halifax, Nova Scotia, 2016), https://tspace.library.utoronto.ca/handle/1807/72802.

[20] Winston Atkins, Kussmann, Carol, Kim, Katherine, and Reich, Aliya. "Staffing for Effective Digital Preservation 2017: An NDSA Report," September 13, 2017, https://doi.org/10.17605/OSF.IO/3RCQK; Carol Kussmann. "Staffing for Effective Digital Preservation 2013: An NDSA Report," October 19, 2017, https://osf.io/5vpxt/.

[21] Jeff Moon, personal communication. 2018.

What we do know, however, is that RDM is essential to realizing the benefits of our new data-rich environment — if this powerful resource is not managed properly and researchers are not supported in their use of it, we could find ourselves data rich, but information poor, losing all of the promise and potential that these data offer. RDM is critical to good public stewardship of the significant investment in research that governments make on behalf of Canadians. It enables appropriate dissemination of the data that underpins evidence-based research, which in turn supports reproducibility and validation of research and the sharing of data within and between disciplines.

We can also point to some quantitative examples from early adopter communities to demonstrate impact. At the 2017 Research Data Canada National Data Services Framework Summit, one of the most compelling demonstrations of the impact and value of RDM and data sharing came from the Canadian Astronomy Data Centre. The speaker, David Schade, showed a slide illustrating data-sharing behaviour in the astronomy community. This slide shows not only growth in numbers of academic papers published by 'first data owner' but, most importantly, growth in the number of papers published by others who found and used these data.

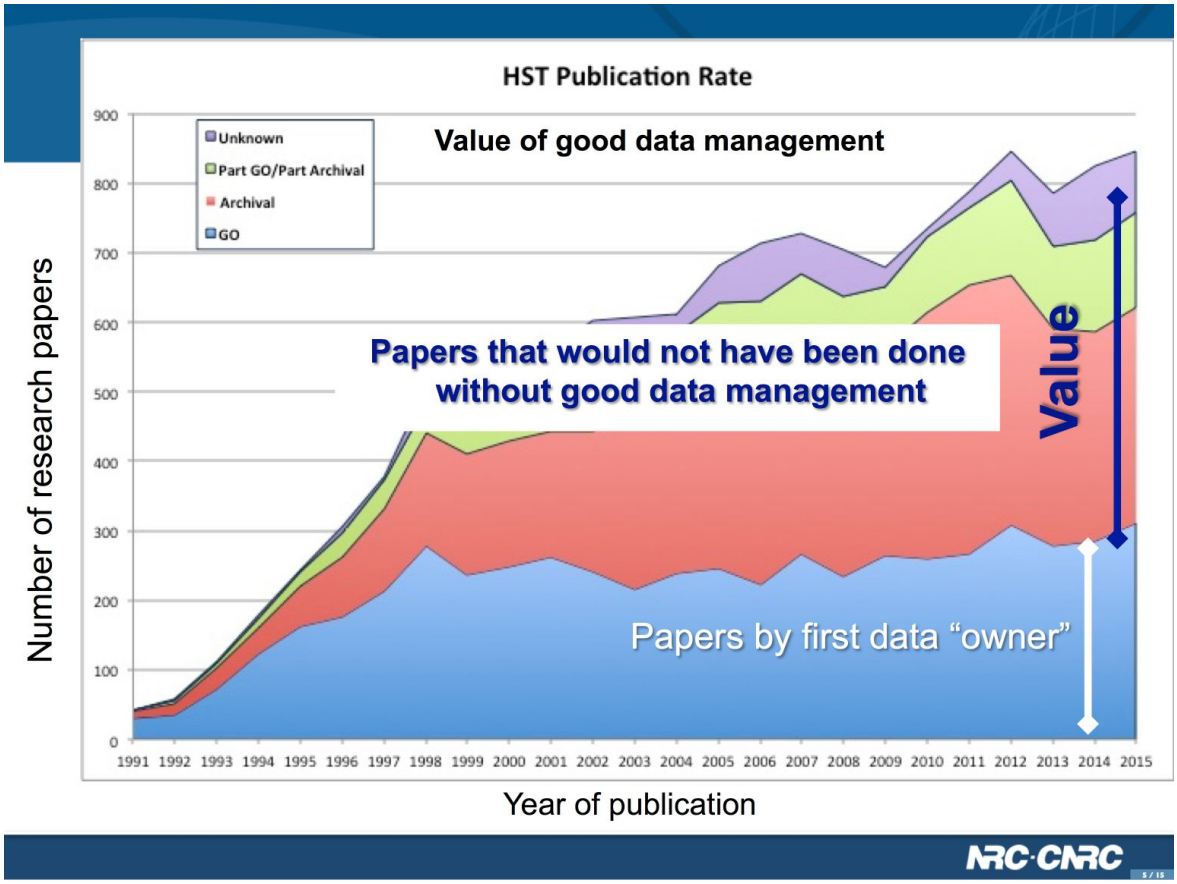

Fig. 3. HST Publication Rates. (Created by David Schade. In *HST Publication Statistics*. [n.d.]. http://archive.stsci.edu/hst/bibliography/pubstat.html.)

There are other international examples of organizations that have tried to quantify the financial/efficiency impact of RDM by using open data as a measure. For example, in a 2016 review of the open data made available by the European Bioinformatics Institute (EMBL-EBI)[22], the authors noted the following impacts from making molecular data and services public and accessible to anyone:

1. A contribution to the wider realization of future research impacts worth £920 million every year; and,
2. Annual direct efficiency impact estimated at between £1 billion and £5 billion per annum.

In addition, in a recent report, the Open Data Institute suggests[23] that making all forms of data openly accessible (research data, as well as government and industrial data) would stimulate research and innovation by approximately 0.5% of GDP. For Canada that represents a $90 billion impact annually.

The Canadian Landscape

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:

1. Libraries (i.e., librarians/data librarians, academic IT personnel, data management committees);
2. Research Offices (i.e., vice-presidents research, research office staff, university research policy committees);

[22] Neil Beagrie, and John Houghton. “The Value and Impact of the European Bioinformatics Institute: Executive Summary,” EMBL-EBI, 2016, https://beagrie.com/static/resource/EBI-impact-summary.pdf.
[23] Open Data Institute, “Permission Granted: The Economic Value of Data Assets under Alternative Policy Regimes,” Open Data Institute, March 2016, https://theodi.org/article/research-the-economic-value-of-open-versus-paid-data/.

3. Information Technology Departments (IT) or Computing Services (i.e. Chief Information Officers (CIO), university data governance committees); and
4. 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:

1. Provincial Research Funding Agencies/Ministries (e.g.,Ontario Ministry of Research and Innovation, les Fonds de recherche du Québec, le Ministère de l'Économie 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,
2. 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).[24]

[24] 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.

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 to 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 CARL Portage and Research Data Canada.

1. CARL Portage Network. Launched in 2015, the Portage Network is a national initiative of the Canadian Association of Research Libraries (CARL), with the goal of promoting shared stewardship of research data and building research data management (RDM) capacity in Canada through a network of over 100 in-kind experts in a growing community of practice. The core aim of Portage is to coordinate and expand existing expertise, services, tools, and platforms so that all researchers in Canada have access to the support they need for research data management. To that end, Portage has identified and addressed a number of specific gaps in Canada's national research data-management infrastructure. CARL's leadership and investments, and in-kind expert contributions, have established a solid foundation for national RDM services and platforms, and helped raise the national profile of RDM more generally. In particular, Portage has:
    1. established and grown a distributed network of over 100 in-kind experts working on over a dozen working and expert groups, and forming the basis of a growing community of practice;
    2. developed, in collaboration with Compute Canada, a Federated Research Data Repository platform (FRDR), that, in concert with other national repository options, will help serve the data deposit needs of Canadian researchers;
    3. signed memoranda of understanding with all four Canadian regional academic library consortia;
    4. launched, in partnership with the University of Alberta, the Data Management Planning (DMP) Assistant, a national, online, bilingual, data management planning tool, along with associated guidance documentation and a growing number of DMP templates;

5. addressed core training needs through the development of online training modules, including a series for health researchers through CIHR, guides, primers, and face-to-face workshops;
6. launched a website that provides access to RDM platforms, resources, and services.
7. worked closely with the Tri-Agency funders and other stakeholders to promote these practical solutions and socialize RDM best practices among researchers and institutions.

2. 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 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.

There are a number of other organizations and associations that are engaged in supporting the delivery of research data management services in Canada.

1. University-Based Organizations
    a. Examples of Research Institutions: Marine Environmental Observation Prediction and Response Network (MEOPAR), CBrain (McGill Neurological Institute), Ocean Networks Canada (ONC), Polar Data Catalogue.
    b. Examples of University Consortia/Organizations: 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).
2. Discipline-Specific Communities of Practice
    a. Examples of Multi-disciplinary: Federation for the Humanities and Social Sciences, Royal Society of Canada, and
    b. Examples of Single-domain: Canadian Society of Microbiologists, Canadian Association of Physicists.

Traditional Knowledge and Ways of Knowing

Canada's Indigenous population is made up of First Nations, Inuit, and Métis[25] people. First Nations, Inuit, and Métis people make up 2.6%, 0.2%, and 1.4% of the Canadian population respectively, accounting for 1,400,685 people as of 2011. Within the Indigenous population, First Nations, Inuit, and Métis people make up 60.8%, 4.2%, and 32.3% respectively.

Canada's national funding agencies have defined various policies and support for research involving the indigenous community, including Chapter 9 of the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans 2 (TCPS 2)*[26], which refers to the ethical conduct of research, and Article 12 of CIHR's Guidelines for Health Research Involving Aboriginal People. The Social Sciences and Humanities Research Council (SSHRC) has a number of resources related to indigenous research.

A key resource developed by First Nations, the OCAP® Principles[27], are a set of standards that establish how First Nations data should be collected, protected, used, or shared. The Principles of OCAP was established in 1998 by Cathryn George, a member of the National Steering Committee (NSC). The NSC evolved into the First Nations Information Governance Committee and in 2010, became First Nations Information Governance Centre (FNIGC). OCAP stands for Ownership, Control, Access and Possession, and asserts that First Nations have control over data collection in their communities, and that they own and control how that information can be used.

Ownership refers to the relationship of First Nations to their cultural knowledge, data, and information. This principle states that a community or group owns information collectively in the same way that an individual owns his or her personal information.

Control affirms that First Nations, their communities, and representative bodies are within their rights in seeking control over all aspects of research and information management that impact them, and throughout the entire research life cycle.

Access refers to the fact that First Nations must have access to information and data about themselves and their communities regardless of where it is held, and have the right to manage and make decisions regarding access to their collective information.

Possession refers to the physical control of data, and the mechanism by which ownership can be asserted and protected.

The OCAP Principles are highlighted in many Canadian organizational policies and guidelines for conducting research, and are also referenced in some international frameworks. The

---

[25] Statistics Canada, "Aboriginal Peoples in Canada: First Nations People, Métis and Inuit," Statistics Canada, July 25 2018, http://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-011-x/99-011-x2011001-eng.cfm.

[26] Government of Canada, Interagency Advisory Panel on Research Ethics, "Research Involving the First Nations, Inuit and Métis Peoples of Canada," Panel on Research Ethics, February 5 2016, www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/chapter9-chapitre9/.

[27] First Nations Information Governance Centre / Le Centre de gouvernance de l'information des Premières Nations, "The First Nations Principles of OCAP®," First Nations Information Governance Centre / Le Centre de gouvernance de l'information des Premières Nations, 2019, https://fnigc.ca/ocap.html.

Principles are important when considering data governance, collection, storage, sharing, and preservation. It is important to note that the OCAP Principles, while created and intended for the use of First Nation's data, may be applicable to any indigenous group.

Two other guiding principles exist in Canada: the Ownership, Control, Access, and Stewardship (OCAS) Principles and Inuit Qaujimajatuqangit (IQ).[28] The OCAS principles govern Métis data whereas IQ governs Inuit data. IQ means "that which has long been known by Inuit" and is focused on integrating traditional Inuit culture into current governance structures, and lessening the disempowerment of Inuit peoples. These two principles are not as widely referenced or disseminated as OCAP, but are an important part of the conversation in Canada. Jodi Bruhn's report *Identifying Useful Approaches to the Governance of Indigenous Data*[29], has some useful suggestions for working with data from indigenous communities. A few other organizations that can provide guidance in this context are the National Collaborating Centre for Aboriginal Health (NCCAH), the British Columbia First Nations' Data Governance Initiative (BCFNDGI), OpenNorth, and their efforts under the Indigenous Data Sovereignty umbrella, the First Nations Health Authority, the Native Council of PEI Research Advisory Committee.

The International Landscape

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:

1. Government Organizations (e.g., the Organization for Economic Co-operation and Development (OECD), the Global Science Forum, the European Commission, the Belmont Forum)
2. Communities of Practice
    a. RDM-focused (e.g. World Data Systems (WDS), WDS-International Technology Office (WDS-ITO), 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,

[28] University of Manitoba First Nations, Metis and Inuit Health, "Framework for Research Engagement with First Nation, Metis, and Inuit Peoples." *University of Manitoba*, n.d., https://umanitoba.ca/faculties/health_sciences/medicine/media/UofM_Framework_Report_web.pdf.
[29] Jodi Bruhn, "Identifying Useful Approaches to the Governance of Indigenous Data," *The International Indigenous Policy Journal,* 5(2), 2014, https://ir.lib.uwo.ca/iipj/vol5/iss2/5/.

3. 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.[30] 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."[31] 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.

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 infrastructure and service offerings, and duplicating this effort makes little economic sense. The

[30] "Data Policies | Scientific Data," Scientific Data, Accessed February 21, 2019, https://www.nature.com/sdata/policies/data-policies.
[31] "Data Availability," PLOS One, Accessed February 21, 2019, https://journals.plos.org/plosone/s/data-availability.

increasingly collaborative and interdisciplinary nature of most research makes this level of coordination and facilitation necessary. Some examples[32] are highlighted below.

1. 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."[33]
2. 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 a 3 years pilot, the Research Data Shared Services program, which will provides a new "a la carte" RDM platform, as well as an RDM services and consultancy initiative to support researchers and their organizations. This pilot ends December 31, 2018.[34]
3. 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
4. The Data Curation Network is a multi-institutional effort to support researchers in making their data findable, accessible, interoperable and reusable through repositories.
5. 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.
6. 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.
7. 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

[32] 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.; RDA NDS Interest Group, "Summary of Country Reports: National Data Services," Research Data Alliance, January 30, 2018, https://docs.google.com/document/u/1/d/17iUyJ2icY0gFzMZGPWJyY5E0tUoukAtI4BFeRonefv4/edit?usp=embed_facebook.

[33] Australian National Data Service, "The ANDS, Nectar and RDS Partnership," ANDS, Accessed February 21, 2019, https://www.ands.org.au/about-us/ands-nectar-rds.

[34] Jisc, "Research Data Shared Service," Jisc, Accessed February 21, 2019, https://www.jisc.ac.uk/rd/projects/research-data-shared-service.

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.

8. In the Netherlands, the national networking (SURFnet), and HPC entities (SURFsara) have formed one organization (SURF), 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 frameworks, and has been providing long-term archival storage for the nation's researchers since its inception.
9. 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 Czech Academy of Sciences.
10. 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[35], 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

[35] Comprehensive brief on Research Data Management Policies. 2015. https://portagenetwork.ca/wp-content/uploads/2016/03/Comprehensive-Brief-on-Research-Data-Management-Policies-2015.pdf

Data: Principles and Implementation Guidelines."[36] 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 Commission[37] and other agencies that recognize the value of RDM and interoperability on an international scale.

1. 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,"[38] 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."
2. 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, and in a timely manner. 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."[39]
3. 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

[36] RDA-CODATA Legal Interoperability Interest Group, "Legal Interoperability of Research Data: Principles and Implementation Guidelines," Zenodo, October 20, 2016, https://doi.org/10.5281/zenodo.162241.
[37] European Union, "Commission Implementing Decision (EU) 2017/1358 of 20 July 2017 on the Identification of ICT Technical Specifications for Referencing in Public Procurement (Text with EEA Relevance.)," EUR-Lex, July 20, 2017, http://data.europa.eu/eli/dec_impl/2017/1358/oj.
[38] Higher Education Funding Coucil for England, Research Councils UK, Universities UK, and Wellcome Trust, "Concordat on Open Research Data," Research Councils UK, July 28, 2016, https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/.
[39] National Science and Technology Council Committee on Science Subcommittee on International Issues Interagency Working Group on Open Data Sharing Policy, "Principles for Promoting Access to Federal Government-Supported Scientific Data and Research Findings Through International Scientific Cooperation," Executive Office of the President of the United States, December 2016, https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/NSTC/iwgodsp_principles_0.pdf.

guidelines.[40] Researchers also must comply with requirements to produce research data management plans.

4. 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"[41] 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,"[42] 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.

Challenges and Opportunities

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.

Current Opportunities for RDM in Canada

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

[40] European Commission, "Open Research Data in Horizon 2020," European Commission, 2016, http://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf.
[41] Australian National Data Service, "Institutional Policies and Procedures," Australian National Data Service, February 3, 2017, https://www.ands.org.au/__data/assets/pdf_file/0008/738782/Institutional-policies-and-procedures.pdf.
[42] "Policy on Data, Software and Materials Management and Sharing | Wellcome," Wellcome Trust, July 10, 2017, https://wellcome.ac.uk/funding/guidance/policy-data-software-materials-management-and-sharing.

previously, the federal government's Tri-Agency Statement of Principles on Digital Data Management, draft RDM Policy[43], 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.

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., Scholars Portal Dataverse), 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., 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.

1. **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.

[43] Draft Tri-Agency Research Data Management Policy (For Consultation), May 25, 2018, http://www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html

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 proactive, ongoing preservation activities to ensure the long-term stewardship, access and usability of data remain under-developed and under-resourced.

2. **Clarity on roles and responsibilities.** Given the diverse and 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, *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.[44] 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."[45] This challenge is being addressed with recent efforts by the federal government to fund the creation of a new organization, that will have a mandate to support high performance computing, storage, research data management, and research software. The new organization will be launched later in 2019. RDC's National Data Services Framework Summits (held in 2017[46] and 2019[47,48]), have gathered a wide group of stakeholders together to document the RDM landscape in Canada, including gaps and opportunities. will need to be addressed if Canada is to develop a strong and effective RDM environment.

3. **Common policies, standards, and protocols.** Common policies, standards, and protocols are critical building blocks for excellence in RDM. They allow researchers to

[44] Shearer, "Comprehensive Brief on Research Data Management Policies," 39-40.
[45] Ibid., 40.
[46] Mark Leggott, "Canadian National Data Services Framework Summit, Sep 22, 2017: Summary", October 25, 2017. https://zenodo.org/record/1035843.
[47] Research Data Canada, "Kanata Declaration", March 17, 2019. https://zenodo.org/record/3234815.
[48] Mark Leggott and Laura Gerlitz, "National Data Services Framework Summit 2019: Notes and Resources", March 7, 2019. https://zenodo.org/record/2584260.

share their data with others and to find, access, and reuse data that has been generated by their colleagues, both domestically and internationally. The promise of a rich semantic repository that represents the full granularity of all types of (meta)data, while facilitating the creation of less-detailed metadata, is one possible way of approaching this complex issue. While RDA and the Portage Network of Expertise have begun to look at how to tackle this issue, and some disciplines such as genomics, astronomy, and ocean science have developed protocols to guide their own RDM practices, this work is still nascent in many other disciplines.[49]

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 creating and accessing the research management information it needs to support structured and efficient RDM planning. One area where this is improving, is in the availability of the Portage DMP Assistant[50], which provides a solid baseline for moving ahead in this aspect of the research lifecycle. 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 only more recently become 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.

4. **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

[49] Research Data Canada, "Submission to Industry Canada," Research Data Canada, February 2014, 3, https://www.rdc-drc.ca/?wpdmdl=671.
[50] https://assistant.portagenetwork.ca/en

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 general lack 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."[51]

Once again, Portage is helping to address these challenges (e.g. campus RDM Day participation, training resources, in-kind experts developing their expertise through participation in the Portage Network), 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 be achieved at all universities across Canada.

5. **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.

6. **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 them for access to this data, publicly

[51] The Leadership Council, "Canadian DI Environmental Scan: A Supplement to the Background Précis Document Provided to DI Summit 2012," The Leadership Council, 2014, 7, https://digitalleadership.ca/wp-content/uploads/2017/05/Canadian-DI-Environmental-Scan.pdf.

owned and managed data repositories and archival storage are essential.

7. **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 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:
    a. attitudinal barriers – researchers have indicated that they are concerned about issues such as a loss of control, being scooped, and privacy)
    b. technical challenges – a Dutch survey across 15 international jurisdictions concluded that technical challenges such as obsolete software were an issue;
    c. a lack of professional expertise and formal training in data management;
    d. a lack of formalized and standardized procedures;
    e. insufficient peer support for and awareness of the importance of RDM; and
    f. insufficient incentives and rewards.[52]

    These barriers will need to be addressed if Canada is to build a culture of and commitment to RDM among universities and researchers.

8. **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.[53] 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.

    Some of these gaps are starting to be addressed with programs like CANARIE's inaugural RDM funding call, which resulted in funding for nine projects, both new developments efforts, and enhancements to existing projects.[54] These projects were selected for their focus on support for the FAIR Principles, and the development of platforms that would provide services at a national level to all Canadian researchers, or

[52] Shearer, "Comprehensive Brief on Research Data Management Policies", 32-33.
[53] Research Data Canada, "Submission to Industry Canada," 2.
[54] Mark Leggott and Laura Gerlitz, "2019 CANARIE RDM Workshop Notes and Resources", January 23, 2019. https://zenodo.org/record/2584274.

all Canadian researchers in a specific domain.[55]

9. **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. 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 RMI to support collaboration among universities, funding agencies, and disciplines. As well, shared processes and tools for migrating research data stored in obsolete software, or 'data rescue', 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[56] 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

[55] Mark Leggott and Laura Gerlitz, "RDM Workshop - Jan 23, 2019 Discussion Document" January 23, 2019. https://zenodo.org/record/2584179.
[56] Jill Kowalchuk, "Storage Requirements for the Long Tail of Researchers in Canada," April 24, 2015, http://bit.ly/RDC_Long_Tail_Storage.

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.

10. **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.[57] 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."[58] 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.[59]

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.

---

[57] Research Data Canada, "Submission to Industry Canada," 2.
[58] Ibid., 3.
[59] Ibid., 3.

A National Vision

In ten years, with appropriate strategic planning, coordination, and investment, we see a future in which RDM in Canada will be transformed.[60]

At the community level, the transformation will result in:

1. a nationally coordinated RDM community that provides researchers with robust RDM services and resources in both official languages;
2. 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;
3. a RDM community and delivery system that is recognized internationally as a leader;
4. innovation for local and domain-specific RDM services and resources that is supported and encouraged nationally;
5. libraries adopting the long-term preservation and discovery of research data as a core responsibility;
6. 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;
7. Canadian data repositories being developed and interoperating collaboratively in a federated environment, nationally and internationally;
8. international standards being used to advance the certification of research data repositories in Canada;
9. 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;
10. 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;
11. the unique data interests of Indigenous Peoples and Traditional Knowledge being an integral part of the RDM community;
12. 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;
13. research data and metadata (that includes clear and appropriate metadata) being integrated into scholarly communications; and
14. a sustainable and transparent system for coordinating RDM within Canada's digital research infrastructure being firmly established.

[60] The 2019 NDSF Summit resulted in the Kanata Declaration, which reflects a similar vision from a broad stakeholder community.

For researchers, the transformation will result in:

1. easy and intuitive access to RDM expertise, training, and resources across a full spectrum of user needs and all disciplines;
2. 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;
3. efficiencies in RDM, whether through automated services or increased support resources, saving time and resources;
4. recognition and reward for the curation and sharing of research data;
5. increased domestic and international collaboration by allowing greater interoperability and accessibility to research data;
6. new connections among researchers that are enabled by improved RDM practices;
7. an increase in interdisciplinary research that results in leading edge outcomes through new and exciting knowledge creation and innovation;
8. early career researchers having greater access to research data and a richer research environment, enabling them to achieve earlier and more significant results; and
9. a culture of data sharing and recognition of the importance of excellence in data stewardship as the norm across all disciplines;
10. increase in data quality due to adherence to standards and attention to RDM best practices.

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

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

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.

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.

1. Responsive to the needs of the research community.
2. Adaptive to changes in research and data management.
3. Innovative in developing local and domain-specific solutions that are scalable.
4. Coordinated among key local, regional, national, and where possible international, research service providers.
5. Integrative by design to facilitate easy incorporation of new and interoperable services and platforms.
6. 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.
7. Distributed to maximize stakeholder cooperation and to leverage their engagement in a community of practice.
8. Accountable to Canada's research community, to those funding it, and to the wider citizenry benefitting from research.
9. Sustainable to allow its ongoing operation.
10. Diverse by valuing and respecting the importance of diversity.

Goals

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

1. Build innovative services and resources that are distributed across universities, nationally coordinated, internationally recognized, sustainable, and necessary for scientific implementation and integral to knowledge development. These services and resources should respect researcher, discipline-specific, national, and university data stewardship policies, and be based on best-practice standards and protocols.
    a. 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.
    b. 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.
    c. Coordinate local and domain-specific resources and services to align with international RDM standards and protocols; and
    d. Develop university plans and procedures for RDM in all Canadian higher education establishments in accordance with federal granting agencies and other funder requirements.

2. 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.
    a. 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.
    b. Develop and ensure widespread adoption of RDM definitions, taxonomies, and unique identifiers by the broader research community.
    c. Integrate metadata production into RDM tools at the project level and throughout the research lifecycle to automate its capture and reuse.
3. 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.
    a. Increase the capacity of Canada's higher education sector to support RDM services and resources for researchers and organizations supporting research.
    b. This includes encouraging collaborative relationships among institutional units, such as libraries, offices of research, research ethics boards, IT services, and external RDM organizations.
    c. Continue to strengthen the capacity of libraries to provide front-line support for RDM as key service points and loci of expertise on campus.
    d. Support, at international, national and local levels, the development of a variety of RDM training resources, such as online courses, webinars, guides, presentations, and in-person workshops.
    e. Engage Canadian non-governmental organizations and government RDM agencies as contributing partners in this community of practice.
    f. Facilitate and encourage Canadian participation in international RDM organizations and the adoption of best practices recognized by the international community.

Selective Bibliography

Atkins, Winston, Kussmann, Carol, Kim, Katherine, and Reich, Aliya. “Staffing for Effective Digital Preservation 2017: An NDSA Report,” September 13, 2017. https://doi.org/10.17605/OSF.IO/3RCQK.

Australian National Data Service. “Institutional Policies and Procedures.” Australian National Data Service, February 3, 2017. https://www.ands.org.au/__data/assets/pdf_file/0008/738782/Institutional-policies-and-procedures.pdf.

———. “The ANDS, Nectar and RDS Partnership.” ANDS. Accessed February 21, 2019. https://www.ands.org.au/about-us/ands-nectar-rds.

Beagrie, Neil, and John Houghton. “The Value and Impact of the European Bioinformatics Institute: Executive Summary.” EMBL-EBI, 2016. https://beagrie.com/static/resource/EBI-impact-summary.pdf.

Bruhn, Jodi. “Identifying Useful Approaches to the Governance of Indigenous Data.” *The International Indigenous Policy Journal* 5, no. 2 (April 23, 2014). https://doi.org/10.18584/iipj.2014.5.2.5.

Buckland, Michael. “Data Management as Bibliography.” *Bulletin of the American Society for Information Science and Technology* 37, no. 6 (2011): 34–37. https://doi.org/10.1002/bult.2011.1720370611.

Costello, Mark, and Edward Vanden Berghe. “‘Ocean Biodiversity Informatics’: A New Era in Marine Biology Research and Management.” *Marine Ecology Progress Series* 316 (2006): 203–14. https://doi.org/10.3354/meps316203.

“Data Availability.” PLOS One. Accessed February 21, 2019. https://journals.plos.org/plosone/s/data-availability.

“Data Policies | Scientific Data.” Scientific Data. Accessed February 21, 2019. https://www.nature.com/sdata/policies/data-policies.

“Digital Infrastructure - CASRAI Dictionary.” CASRAI, August 13, 2015. https://dictionary.casrai.org/Digital_infrastructure.

European Commission. “Open Research Data in Horizon 2020.” European Commission, 2016. http://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf.

European Union. “Commission Implementing Decision (EU) 2017/1358 of 20 July 2017 on the Identification of ICT Technical Specifications for Referencing in Public Procurement (Text

with EEA Relevance.).” EUR-Lex, July 20, 2017. http://data.europa.eu/eli/dec_impl/2017/1358/oj.

Government of Canada. “Tri-Agency Statement of Principles on Digital Data Management.” Government of Canada, 2016. http://www.science.gc.ca/eic/site/063.nsf/eng/h_83F7624E.html.

Government of Canada, Interagency Advisory Panel on Research Ethics. “Research Involving the First Nations, Inuit and Métis Peoples of Canada.” Panel on Research Ethics, February 5, 2016. http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/chapter9-chapitre9/.

Harvey, Ross. *Digital Curation: A How-To-Do-It Manual*. New York: Neal-Schuman Publishers, Inc., 2010.

Higher Education Funding Coucil for England, Research Councils UK, Universities UK, and Wellcome Trust. “Concordat on Open Research Data.” Research Councils UK, July 28, 2016. https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/.

Jisc. “Research Data Shared Service.” Jisc. Accessed February 21, 2019. https://www.jisc.ac.uk/rd/projects/research-data-shared-service.

Kowalchuk, Jill. “Storage Requirements for the Long Tail of Researchers in Canada,” April 24, 2015. http://bit.ly/RDC_Long_Tail_Storage.

Kussmann, Carol. “Staffing for Effective Digital Preservation 2013: An NDSA Report,” October 19, 2017. https://osf.io/5vpxt/.

Lucas, Matthew. “Tri-Agency Data Management Policy Initiative.” presented at the RDA 10th Plenary Collocated Event, Montreal, Quebec, September 18, 2017. https://portagenetwork.ca/wp-content/uploads/2017/09/TC3-Data-Management-Policy-Initiative.pdf.

National Science and Technology Council Committee on Science Subcommittee on International Issues Interagency Working Group on Open Data Sharing Policy. “Principles for Promoting Access to Federal Government-Supported Scientific Data and Research Findings Through International Scientific Cooperation.” Executive Office of the President of the United States, December 2016. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/NSTC/iwgodsp_principles_0.pdf.

Open Data Institute. “Permission Granted: The Economic Value of Data Assets under Alternative Policy Regimes.” Open Data Institute, March 2016. https://theodi.org/article/research-the-economic-value-of-open-versus-paid-data/.

“Policy on Data, Software and Materials Management and Sharing | Wellcome.” Wellcome Trust, July 10, 2017.

https://wellcome.ac.uk/funding/guidance/policy-data-software-materials-management-and-sharing.

Poole, Alex H. “How Has Your Science Data Grown? Digital Curation and the Human Factor: A Critical Literature Review.” *Archival Science* 15, no. 2 (June 1, 2015): 101–39. https://doi.org/10.1007/s10502-014-9236-y.

Qasim, Umar, Corey Davis, Alex Garnett, Steve Marks, and Michael Moosberger. “Research Data Preservation in Canada: A White Paper.” R. UBC Community and Partner Publications, Last modified April 30, 2018, https://dx.doi.org/10.14288/1.0371946.

RDA NDS Interest Group. “Summary of Country Reports: National Data Services.” Research Data Alliance, January 30, 2018. https://docs.google.com/document/u/1/d/17iUyJ2icY0gFzMZGPWJyY5E0tUoukAtI4BFeRonefv4/edit?usp=embed_facebook.

RDA-CODATA Legal Interoperability Interest Group. “Legal Interoperability of Research Data: Principles and Implementation Guidelines.” Zenodo, October 20, 2016. https://doi.org/10.5281/zenodo.162241.

Research Data Canada. “Submission to Industry Canada.” Research Data Canada, February 2014. https://www.rdc-drc.ca/?wpdmdl=671.

Schade, David. “HST Publication Statistics.” HST Publication Statistics. Accessed February 15, 2019. http://archive.stsci.edu/hst/bibliography/pubstat.html.

Sewerin, Cristina, Eugene Barsky, Dylanne Dearborn, Angela Henshilwood, Christina Hwang, Sandra Keys, Marjorie Mitchell, Michelle Spence, Kathy Szigeti, and Tatiana Zaraiskaya. “From Coast to Coast: Canadian Collaboration in a Changing RDM Seascape.” Halifax, Nova Scotia, 2016. https://tspace.library.utoronto.ca/handle/1807/72802.

Shearer, Kathleen. “Comprehensive Brief on Research Data Management Policies.” Portage Network, April 7, 2015. https://portagenetwork.ca/wp-content/uploads/2016/03/Comprehensive-Brief-on-Research-Data-Management-Policies-2015.pdf.

Statistics Canada. “Aboriginal Peoples in Canada: First Nations People, Métis and Inuit.” Statistics Canada, July 25, 2018. https://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-011-x/99-011-x2011001-eng.cfm.

The First Nations Information Governance Centre / Le Centre de Gouvernance de L’information des Premières Nations. “First Nations Information Governance Centre / Le Centre de Gouvernance de l’information Des Premières Nations.” First Nations Information Governance Centre / Le Centre de gouvernance de l’information des Premières Nations, 2019. http://fnigc.ca/ocap.html.

The Leadership Council for Digital Research Infrastructure. “Advanced Research Computing (ARC) Position Paper: For Innovation, Science, and Economic Development Canada.” The Leadership Council for Digital Research Infrastructure, August 31, 2017.

———. “Canadian DI Environmental Scan: A Supplement to the Background Précis Document Provided to DI Summit 2012.” The Leadership Council for Digital Research Infrastructure, 2014. https://digitalleadership.ca/wp-content/uploads/2017/05/Canadian-DI-Environmental-Scan.pdf.

University of Manitoba First Nation, Metis and Inuit Health. “Framework for Research Engagement with First Nation, Metis, and Inuit Peoples.” Winnipeg, Manitoba: University of Manitoba. Accessed March 7, 2019. https://umanitoba.ca/faculties/health_sciences/medicine/media/UofM_Framework_Report_web.pdf.

Vejvoda, Berenica, Alison Ambi, Eugene Barsky, Kevin Lindstrom, Heather MacDonald, Kathleen Matthews, Michael Moosberger, et al. “Portage Data Discovery Expert Group - Collections Development Working Group : Phase One Report,” July 2017. https://doi.org/10.14288/1.0351978.

Canadian

Barsky, Eugene. “Getting Your Hands Dirty with Research Data Management (RDM) : ACCOLEDS 2016.” November 23, 2016. https://open.library.ubc.ca/cIRcle/collections/ubclibraryandarchives/494/items/1.0339936.

———. “Research Data Management DataGuide.” UBC, August 21, 2015. http://researchdata.library.ubc.ca/files/2015/08/RDM_DataGuide_V02.8_20150821.pdf.

Barsky, Eugene, Brosz, John, and Leahey, Amber. “Research Data Discovery and the Scholarly Ecosystem in Canada,” July 2016. https://portagenetwork.ca/wp-content/uploads/2016/07/Portage_discovery_white_paper_EN.pdf.

CARL. “CARL Response to the Tri-Agency Statement of Principles on Digital Data Management,” 2015. https://portagenetwork.ca/wp-content/uploads/2016/04/2015_Portage_response_DDM_EN.pdf.

CARL Portage Network. “ISED Follow-up Questions on DM.” CARL Portage Network, 2018.

Daniels, Morgan, Ana Van Gulick, Sarah Pickle, Scout Calvert, Stephanie Simms, Susan Wells Parham, Patricia Hswe, Brian Westra, and Amanda Whitmire. *Training up and Reaching out : Library Strategies to Coordinate Research Data Management on Campus; Data*

Management Plans as a Research Tool, 2016. https://open.library.ubc.ca/cIRcle/collections/55474/items/1.0220814.

Davis, Corey. “Recommendations Report for the COPPUL Digital Preservation Network,” February 19, 2016. http://www.coppul.ca/sites/default/files/uploads/RecommendationsReportfortheCOPPULDigitalPreservationNetwork_0.pdf.

Government of Canada. “Tri-Agency Statement of Principles Data Management,” June 2016. http://www.science.gc.ca/eic/site/063.nsf/eng/h_83F7624E.html?OpenDocument.

Ishida, Mayu. “The New England Collaborative Data Management Curriculum Pilot at the University of Manitoba: A Canadian Experience.” *Journal of EScience Librarianship* 3, no. 1 (December 31, 2014). https://doi.org/10.7191/jeslib.2014.1061.

Moon, Jeff. “Developing a Research Data Management Service – a Case Study.” *Partnership: The Canadian Journal of Library and Information Practice and Research* 9, no. 1 (May 29, 2014). https://doi.org/10.21083/partnership.v9i1.2988.

Portage DMP Customization working Group. “DMP Assistant Institutional Customization Guide: Version 1,” July 13, 2016. https://portagenetwork.ca/wp-content/uploads/2016/07/DMPAsssistant_Customization_Guide_1.0_EN.pdf.

Sewerin, Cristina, Eugene Barsky, Dylanne Dearborn, Angela Henshilwood, Christina Hwang, Sandra Keys, Marjorie Mitchell, Michelle Spence, Kathy Szigeti, and Tatiana Zaraiskaya. “From Coast to Coast: Canadian Collaboration in a Changing RDM Seascape,” n.d. https://open.library.ubc.ca/cIRcle/collections/ubclibraryandarchives/494/items/1.0305007.

Steeleworthy, Michael. “Research Data Management and the Canadian Academic Library: An Organizational Consideration of Data Management and Data Stewardship.” *Partnership: The Canadian Journal of Library and Information Practice and Research* 9, no. 1 (June 5, 2014). https://doi.org/10.21083/partnership.v9i1.2990.

Tenopir, Carol, Robert J. Sandusky, Suzie Allard, and Ben Birch. “Research Data Management Services in Academic Research Libraries and Perceptions of Librarians.” *Library & Information Science Research* 36, no. 2 (April 2014): 84–90. https://doi.org/10.1016/j.lisr.2013.11.003.

Whitehead, Martha, and Bourne-Tyson, Donna. “Multi-Stakeholder Engagement in Research Data Management,” June 2016. https://portagenetwork.ca/wp-content/uploads/2016/06/IATUL2016_Multi_Stakeholder_Engagement_in_RDM.pdf.

International

"Best Practices | DataONE." DataONE. Accessed July 22, 2016. https://www.dataone.org/best-practices.

European Commission. *Open Innovation, Open Science, Open to the World: A Vision for Europe*. Luxembourg: European Union, 2016. file:///C:/Users/chantel.ridsdale/Downloads/KI0416263ENN_002.pdf.

Russell, Fiona. *Delivering Research Data Management Services: Fundamentals of Good Practice*. Vol. 45, 2014. http://dx.doi.org/10.1080/00048623.2014.974239.

"State of the Art Report on Open Access Publishing of Research Data in the Humanities." Accessed September 14, 2016. https://halshs.archives-ouvertes.fr/halshs-01357208/document.

United Kingdom

Ball, Alex. "Research Data Management: Policy, Planning and Practice." May 22, 2013. http://opus.bath.ac.uk/35329/1/ball_rdm_aslib_slides.pdf.

DCC. "Checklist for a Data Management Plan v4.0." DCC, 2013. http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf.

Donnelly, Martin. "The DCC's Institutional Engagements: Raising Research Data Management Capacity in UK Higher Education." *Bulletin of the American Society for Information Science and Technology* 39, no. 6 (August 1, 2013): 37–40. https://doi.org/10.1002/bult.2013.1720390612.

Van den Eynden, Veerle, Louise Corti, Matthew Woolard, Libby Bishop, and Laurence Horton. *Managing and Sharing Data: UK Data Archive*. Third. University of Essex, UK: UK Data Archive, 2011. http://www.data-archive.ac.uk/media/2894/managingsharing.pdf.

United States

Borghi, John. "Building a User-Friendly RDM Maturity Model." *Data Pub* (blog), September 12, 2016. https://datapub.cdlib.org/2016/09/12/building-a-user-friendly-rdm-maturity-model/.

Borgman, Christine. "The Conundrum of Sharing Research Data." *Journal of the American Society for Information Science and Technology*, 2011, 1–40. http://dx.doi.org/10.2139/ssrn.1869155.

Henderson, Margaret, Regina Raboin, Yasmeen Shorish, and Steve Van Tuyl. "Research Data Management on a Shoestring Budget." *Bulletin of the Association for Information Science*

and Technology 40, no. 6 (August 1, 2014): 14–17. https://doi.org/10.1002/bult.2014.1720400606.

Katz, John. “Data Publication Consensus and Controversies [Version 3; Referees: 3 Approved].” F1000 Research, October 16, 2014. http://f1000researchdata.s3.amazonaws.com/manuscripts/5878/397d927b-269a-4f45-826d-27fef7106eaf_3979_-_john_kratz_v3.pdf.

Kevin Ashley, Sarah Jones, Daniel Mietchen, Stephanie Simms, and Angus Whyte. “Workshop on Active DMPs: Making DMPs Actionable and Public.” Conference Website, June 29, 2016. https://indico.cern.ch/event/520120/contributions/2177900/attachments/1299358/1938982/DMPs_-_actionable_and_public.pdf.

Rolando, Lizzy, Jake Carlson, Patricia Hswe, Susan Wells Parham, Brian Westra, and Amanda L. Whitmire. “Data Management Plans as a Research Tool.” *Bulletin of the Association for Information Science and Technology* 41, no. 5 (June 1, 2015): 43–45. https://doi.org/10.1002/bult.2015.1720410510.

Wiley, Christie. “Metadata Use in Research Data Management.” *Bulletin of the Association for Information Science and Technology* 40, no. 6 (August 1, 2014): 38–40. https://doi.org/10.1002/bult.2014.1720400612.

Wright, Stephanie, Amanda Whitmire, Lisa Zilinski, and David Minor. “Collaboration and Tension between Institutions and Units Providing Data Management Support.” *Bulletin of the Association for Information Science and Technology* 40, no. 6 (August 1, 2014): 18–21. https://doi.org/10.1002/bult.2014.1720400607.