Comprehensive Brief on Research Data Management Policies

April 7, 2015

Prepared by Kathleen Shearer, Consultant

Table of Contents

List of Acronyms

AHRC: Arts and Humanities Research Council (UK)

ANDS: Australian National Data Service

ARC: Australian Research Council

BBSRC Biotechnology and Biological Sciences Research Council (UK)

CARL: Canadian Association of Research Libraries

CARA: Canadian Association of Research Administrators

CFI: Canada Foundation for Innovation

CIHR: Canadian Institutes of Health Research

DCC: Digital Curation Centre (UK)

DMP: Data Management Plan

EC: European Commission

EPSRC: Engineering and Physical Sciences Research Council (UK)

ESRC: Economic and Social Research Council (UK)

IPY: International Polar Year

MRC: Medical Research Council (UK)

NERC: Natural Environment Research Council (UK)

NIH: National Institutes of Health (US)

NRC: National Research Council

NSERC: Natural Sciences and Engineering Research Council of Canada

NSF: National Science Foundation (UK)

OECD: Organization for Economic Cooperation and Development

RCUK: Research Councils UK

RDA: Research Data Alliance

RDC: Research Data Canada

RDM: Research Data Management and Sharing

REBs: Research Ethics Boards

SSHRC: Social Sciences and Humanities Research Council of Canada

STFC: Science and Technology Facilities Council (UK)

TCPS: Tri-Council Policy Statement on the Ethical Conduct for Research Involving Humans

1. Executive Summary

The volume of scientific data around the world is increasing at a phenomenal rate. According to a report of the Canadian Research Data Summit in 2011, "the way that we choose to manage our research data will directly impact our ability to undertake leading edge research and development in the future."[1] But managing data is about much more than supporting research excellence. "Digital data are the raw materials of the knowledge economy and are becoming increasingly important for all areas of society. Policies, services and infrastructure must be in place if we are to capitalize on this rising tide of data."[2]

This brief presents the results of an environmental scan undertaken in the fall/winter of 2014-2015. It provides an up-to-date and detailed overview of the policy environment for research data management and sharing in Canada and internationally. The report identifies some of the major challenges related to policy adoption and concludes with a readiness assessment for policy implementation in Canada.

The scan found that there are a growing number of research data management and sharing policies (herewith referred to as RDM policies) being adopted by funding agencies and institutions around the world. The objectives of these policies are to improve the efficiency of research, support the re-use of data for new insights and discoveries, foster collaboration, and facilitate greater transparency. To achieve these policy objectives, research data must be properly managed across the data lifecycle.

The jurisdictions with the most comprehensive policy environments are the United Kingdom, United States, Australia and European Union. Details of policies vary across regions, agencies and domains, but they also have a number of things in common. The most frequent policy components are requirements around standards and metadata, data sharing, and data retention and/or long-term preservation. Data management plans (DMPs) are usually required in the context of these policies, as they compel researchers to think about how they will manage their data in advance of the project, a key requisite for good data management practices. Policies also consistently contain provisions for the protection of confidentiality, intellectual property, and sensitive data.

There is significant diversity in terms of how policies are monitored and implemented. Research data management policies are new to many organizations and most are still working out how to administer them appropriately. In some cases, policies are adopted with little or no monitoring of compliance. In other instances, DMPs are attached to proposals and undergo a light review by peer review committees, but with little or no follow-up at the end of the project. Still in other cases, policy compliance is reviewed at the end of a project and there are consequences for non-compliance. Likewise, some organizations have chosen to phase in their policy. The European Commission, for example, has begun with a pilot project that requires 15% of their funded research projects to develop data management plans. The Engineering and Physical Sciences Research Council (EPSRC) in the UK has taken the distinctive approach of requiring universities (rather than researchers) to develop roadmaps that will ensure the Council's policy can be adhered to by funded researchers. Regardless of how policies are managed or implemented, support and guidance for researchers is essential to ensure compliance, since many researchers are not familiar with what is involved in good data management practices. It is also clear that full adherence to any policy will take time and will likely happen incrementally.

[1] http://www.rdc-drc.ca/wp-content/uploads/Report-of-the-Canadian-Research-Data-Summit1.pdf
[2] http://www.rdc-drc.ca/wp-content/uploads/Report-of-the-Canadian-Research-Data-Summit1.pdf

In Canada, the federal government has been increasing its interest and support for research data management and sharing through open government and open science initiatives. “Canada’s Science, Technology and Innovation Strategy 2014” includes a section promoting open science through the facilitation of “open access to publications and related data resulting from federally-funded research in order to accelerate research, drive innovation and benefit the economy”[3]. In February 2015, the agencies announced a new “Tri-Agency Open Access Policy on Publications” that requires research publications supported by public funds to be made openly available for the benefit of the community at large.[4] The TC3 are currently assessing how to move forward with research data management within this broader policy context.

Unquestionably, policies cannot be adopted in isolation. Good research data management practices will depend on multiple contributing factors including incentives, expertise, services and infrastructure, as well as appropriate funding mechanisms. This review found that the situation for RDM is improving and Canada has made significant progress since signing the “OECD Declaration on Access to Research Data from Public Funding” in 2004, with both bottom-up and top-down advances in RDM infrastructure, services and expertise.

That being said, Canada still lacks infrastructure, services and funding mechanisms to support widespread RDM. Infrastructure funding remains focused on domain-based solutions that support research excellence, rather than data sharing and preservation after the lifespan of the project. Portage, a national library-based network for managing research data (led by the Canadian Association of Research Libraries), and its collaborators (Research Data Canada and Compute Canada) are laying the foundation for more horizontal infrastructures and services, but this is a grassroots effort that will have difficulties expanding without external funding.

There are other challenges. Institutions and researchers still need to be convinced. While there are some research communities that have embraced a culture of data management and sharing, many stakeholders do not think RDM should be a priority for the research community. Both researchers and institutions are apprehensive about taking on greater responsibility for managing research data. Researchers are worried about the time, knowledge and resources involved in preparing data. Institutions are concerned about how they will fund data management support services and infrastructures. Parallel efforts must be made to increase acceptance of policy objectives within the various stakeholder communities, including the adoption of appropriate incentive schemes. Skills and expertise in the area of RDM must also be expanded.

Despite the challenges, it is clear that policies are an extremely powerful lever to push the community forward. They provide a framework that helps to guide best practices and without them it is unlikely that there will be widespread adoption of RDM in Canada. Countries that have chosen to move ahead with policy implementation have found that although full compliance cannot be expected immediately, policies can greatly assist in raising awareness of RDM. As noted in a 2013 TC3+ consultation document, “Canada now stands in direct competition with a host of other countries… in the race to develop an effective strategy for harnessing the digital wave.”[5] RDM policies are an important component of any such strategy.

[3] http://www.ic.gc.ca/eic/site/icgc.nsf/eng/07482.html#promoting
[4] http://www.science.gc.ca/default.asp?lang=En&n=F6765465-1
[5] http://www.sshrc-crsh.gc.ca/about-au_sujet/publications/digital_scholarship_consultation_e.pdf

2. Introduction

Over the past ten years we've witnessed the acceleration of a significant worldwide trend towards research data sharing and management. This trend is progressing in parallel with a movement calling for open access to research publications and can be situated within a broader effort to ensure that the results of publicly funded research are available to the public. In 2004, 34 countries including Canada signed the "OECD Declaration on Access to Research Data from Public Funding".[6] Through this declaration, signatories were recognizing that open access to, and unrestricted use of, data promotes scientific progress and contributes to new discoveries and innovation. In addition, they were acknowledging that data sharing maximizes the value derived from public investments and supports re-use of data across disciplinary and jurisdictional boundaries. Since that time, the momentum for data sharing and management has grown and investments in research data management policies and services have increased dramatically.

At the same time, in Canada the federal government has been increasing its commitment to open science. This originated with open government and open data initiatives, and has grown in importance with the adoption of the "Open Data Charter" in June 2013 at the G8 (now G7) Summit in Lough Erne, Northern Ireland. At this summit, Canada and all other G8 members agreed to implement a set of open data principles and best practices that would lay the foundation for the release and reuse of government data before December 31, 2015.[7] "Canada's Science, Technology and Innovation Strategy 2014" identified open science as a priority, as has the Open Government Initiative. In order to deliver on these commitments, the Government of Canada has prepared an Open Government Action Plan with three streams: open access, open data, and public engagement.

In parallel, the three federal granting agencies have been examining the issues around open access and research data management. In 2011, they commissioned an environmental scan of the policy context related to open access and research data management. The results were presented in a "Brief on Open Access to Publications and Research Data"[8], which provided background information about both the growing trend towards open access to publications as well as sharing of research data. In 2013 the TC3+ (CIHR, NSERC, SSHRC, CFI) and Genome Canada undertook a consultation with the stakeholder community to further engage and develop a strategy for a more comprehensive and coordinated approach to RDM in Canada. And most recently, in February 2015, the agencies released a "Tri-Agency Open Access Policy on Publications" that requires research publications supported by public funds to be made openly available for the benefit of the community at large.[9]

The aim of this brief is to inform the community about the state of RDM policy development internationally and in Canada. This brief presents the results of an environmental scan undertaken in the fall/winter 2014-2015. It provides an up-to-date and detailed overview of the RDM policy environment. The report discusses some of the major challenges related to policy adoption, and concludes with a gap analysis and readiness assessment for policy implementation in Canada.

[6] http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=157
[7] http://data.gc.ca/eng/g8-open-data-charter-canadas-action-plan
[8] http://science.gc.ca/default.asp?lang=en&n=2360F10C-1
[9] http://www.science.gc.ca/default.asp?lang=En&n=F6765465-1

3. Policy Environment

While certainly not ubiquitous, there are a growing number of research data management and/or sharing policies being adopted by funding agencies and institutions around the world. The purpose of these policies is to improve data management practices in order to allow data produced through research to be shared and re-used by others, enable the verification of research results, and ultimately fuel further innovation.

3.1 Policy Objectives and Principles

Research data management policies are situated in the context of a broader set of principles and objectives that guide their specific requirements. In general, policies will support several (or all) of the objectives listed below:

- Accelerate research
- Support new insights and discoveries
- Foster collaboration
- Improve efficiency of research
- Facilitate accountability

The 2003 OECD declaration, of which Canada was a signatory, outlines a comprehensive set of principles for RDM that have informed the principles adopted by many other organizations and remain relevant over a decade later[10]:

- **Openness:** balancing the interests of open access to data to increase the quality and efficiency of research and innovation with the need for restriction of access in some instances to protect social, scientific and economic interests

- **Transparency**: making information on data-producing organizations, documentation on the data they produce and specifications of conditions attached to the use of these data, available and accessible internationally

- **Legal conformity:** paying due attention, in the design of access regimes for digital research data, to national legal requirements concerning national security, privacy and trade secrets

- **Formal responsibility:** promoting explicit, formal institutional rules on the responsibilities of the various parties involved in data-related activities pertaining to authorship, producer credits, ownership, usage restrictions, financial arrangements, ethical rules, licensing terms, and liability

- **Professionalism:** building institutional rules for the management of digital research data based on the relevant professional standards and values embodied in the codes of conduct of the scientific communities involved

- **Protection of intellectual property:** describing ways to obtain open access under the different legal regimes of copyright or other intellectual property law applicable to databases as well as trade secrets

[10] http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=157

- **Interoperability:** paying due attention to the relevant international standard requirements for use in multiple ways, in co-operation with other international organizations.

- **Quality and security:** describing good practices for methods, techniques and instruments employed in the collection, dissemination and accessible archiving of data to enable quality control by peer review and other means of safeguarding authenticity, originality, integrity, security and establishing liability

- **Efficiency:** promoting further cost effectiveness within the global science system by describing good practices in data management and specialized support services

- **Accountability:** evaluating the performance of data access regimes to maximize the support for open access among the scientific community and society at large

3.2 Typical Policy Requirements

The features of a given RDM policy will reflect the particular objectives and principles on which it is based. Therefore, while many policies contain similar elements, there may be greater emphasis on some requirements over others. For example, a policy based on the principle of data sharing will likely concentrate on key practices needed for providing access to the data, while a policy based on data stewardship will focus on the roles and responsibilities involved in managing data.

The most common elements of the RDM policies reviewed for this scan are outlined in the table below:

Table 1: Common elements of a RDM policy

Policy requirements	
Data quality and standards	Investigators are required to adhere to international standards to enable access and reuse. Data documentation and metadata must accompany data so that the data are understandable by others.
Data access and sharing	Investigators are required to make data available to be shared (usually upon publication of results or shortly thereafter, although some agencies do allow embargo periods). Requirements for deposit of metadata into a local or national catalogue.
Data retention and preservation	Data are required to be retained for a certain minimum time period. Where possible, investigators must deposit their data in a long-term archive to ensure the preservation of their data.
Data management plans	Research proposals must include a Data Management Plan.

Common provisions to policies	
Privacy	The rights and privacy of individuals who participate in research must be protected at all times. Thus, data made available for broader use should be free of identifiers that would permit linkages to individual research participants and variables that could lead to deductive disclosure of the identity of individual subjects.
Traditional knowledge	Where local and traditional knowledge is concerned, rights of the knowledge holders shall not be compromised.
Data of a sensitive nature	Where data release may cause harm, specific aspects of the data may need to be protected (for example, locations of nests of endangered birds or locations of sacred sites).
Intellectual property/Data ownership	It may be necessary on occasion to delay publication for a short period to allow time for applications to be drafted.
Other aspects	
Principles	Data policies adhere to a set of overarching principles that articulate their value.
Scope/Coverage of Policy	Describe the scope of data covered by the policy.
Roles and responsibilities	The policy identifies the various parties responsible for managing data across the different stages of the lifecycle.
Monitoring and enforcement	The means by which policies will be monitored or enforced are outlined in the policy.

3.3 International

The jurisdictions that are most advanced in terms of research funders' policies are the United Kingdom and the United States. The different agency policies vary across organizations in terms of their strength, coverage, roles and responsibilities, and requirements. A written summary of the major funders' policies follows.

United Kingdom: In 2011, Research Councils UK (RCUK) issued a set of "Common Principles on Data Policy". The principles call for data to be made openly available with as few restrictions as possible in a timely and responsible manner. Since then, each of the seven councils in the Research Council UK has implemented a policy on access to research data, as has the Wellcome Trust (a major charitable organization that funds biomedical research). The UK funders' policies range in terms of their requirements and details, but they are generally aligned with the Common Principles. According to an overview published by the University of Bath[11], policies typically cover the following elements:

- Types of data covered by the policy

[11] http://www.bath.ac.uk/research/data/policy/funder-data-policies.html

- Expectations for data sharing including access and timescales
- Minimum data retention periods
- Use of metadata and documentation standards
- Justified exemptions to data sharing
- Costs associated with data management that may be paid for through grants
- Requirements for submission of data management plans with grant applications
- Acknowledgement of data creators

The following table of UK funders' requirements was developed by the (DCC) Centre in the UK and provides a useful overview:

Table 2: DCC's Overview of UK Funders' Data Policies[12]

(A list of acronyms is provided at the beginning of the report)

	Policy Coverage		Policy Stipulations					Support Provided			
Research Funders	Published outputs	Data	Time limits	Data plan	Access/ sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	◐	○	●	○	◐	◐
BBSRC	●	●	●	●	●	●	●	●	●	◐	●
CRUK	●	●	●	●	●	●	●	◐	●	○	○
EPSRC	●	●	●	◐	●	●	●	◐	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	◐
MRC	●	●	●	●	●	●	○	◐	●	○	◐
NERC	●	●	●	●	●	●	●	●	●	●	◐
STFC	●	●	●	●	●	●	●	◐	●	◐	◐
Wellcome Trust	●	●	●	●	●	●	●	●	●	◐	●

The table illustrates that most of the UK funders require researchers to complete a data management plan; ensure that they are using appropriate metadata and standards; and retain data or deposit into a repository when available. For example, the Economic and Social Research Council requires researchers to prepare a data management plan and stipulates, "[t]he data must be made available for re-use or archiving with the ESRC data service providers within three months of the end of the grant."[13]

Alternatively, the Arts and Humanities Research Council (AHRC) requires a technical plan in cases where digital outputs or digital technologies are an essential part to the planned research outcomes. The plan should give a summary of those outputs, explain the technical methodology, technical support and relevant experience, and address preservation, sustainability and use. The AHRC also requires that significant electronic

[12] http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies
[13] http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx

resources or datasets be made available in an accessible depository for at least three years after the end of the grant.[14]

The Engineering and Physical Sciences Research Council (EPSRC) has taken a somewhat different approach from the other councils. In its policy, EPSRC has set out clear expectations for institutions they fund, which include a requirement that institutions develop of an institutional 'Roadmap'. This is discussed in more detail below, in the section entitled Approaches to Policy Implementation.

***United States*:** In order to improve the management of research data produced through publicly funded research, the White House's Office of Science and Technology Policy (OSTP) published a policy memorandum that directed all federal agencies with more than $100M in R&D expenditures to require researchers to better account for and manage the digital data resulting from federally funded scientific research. Each of the 22 agencies subject to these requirements was required to develop a plan that outlines how they will adhere to this policy. Plans for fulfilling this directive are being developed by all agencies and are beginning to be made publicly available. Several agencies including the Department of Defence, Department of Energy, NASA, National Science Foundation (NSF), and the National Institute of Standards and Technology have all released plans (or draft plans) that will require funded researchers to develop data management plans.[15]

Both the National Institutes of Health (NIH) and the NSF implemented RDM policies before the OSTP directive. Adopted in 2001, NIH was one of the first funding agencies to have a policy about research data sharing. The policy states "[d]ata should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data."[16] In addition, investigators submitting a research application requesting $500,000 or more of direct costs in any single year to NIH are expected to include a plan to explain how they will share their data, or explain why data sharing is not possible.

The NSF policy states, "[i]nvestigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing."[17] In 2011, the NSF introduced a new requirement that all proposals must include a supplementary document of no more than two pages labeled "Data Management Plan". The plan should describe how research teams will conform to the policy. On March 2015, NSF released their plan for pubic access to NSF funded research as requested by OSTP, which underscored their ongoing support for data management plans.[18] The plan also encourages researchers to cite their data in the context of publications and proposals.

***European Commission*:** In Horizon 2020, the European Commission's (EC) financial instrument for funding research from 2014-2020, a pilot action on open access to research data will be implemented with the aim of improving and maximizing access to

[14] http://www.dcc.ac.uk/resources/policy-and-legal/research-funding-policies/ahrc

[15] http://science.energy.gov/funding-opportunities/digital-data-management/
[16] http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm
[17] http://www.nsf.gov/bfa/dias/policy/dmp.jsp
[18] www.nsf.gov/news/special_reports/public_access/

and re-use of research data generated by funded projects. The "Pilot on Open Research Data" will be monitored throughout Horizon 2020 with a view to further developing EC policy on open research. As part of the pilot, projects that fall into 7 pre-selected research areas[19] will be required to develop a data management plan outlining how they will manage and provide access to their data. Pilot areas represent about 20% of the overall research funding budget of Horizon 2020. Projects participating in the pilot must also deposit data into a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — both the data needed to validate results and any other data generated in the project.[20] The EC does not currently specify what this would entail, but presumably they will be asking researchers to assign re-use licenses to their data when depositing. More information about this project is described in a later section, Approaches to Policy Implementation.

***Australia*:** Developed jointly by the National Health and Medical Research Council, the Australian Research Council and Universities Australia in 2007, "The Code for the Responsible Conduct of Research" contains a section on the "Management of Research Data and Primary Materials". Primary materials are defined as objects (physical or virtual) acquired through a process of scholarly investigation from which Research Data may be derived.[21] The Code outlines the roles and responsibilities of different stakeholders in the research process, and assigns responsibility for the management and retention of research data to both researchers and institutions. It requires institutions to have policies on the retention and management of materials and research data, stating: "It is important that institutions acknowledge their continuing role in the management of research material and data. The institutional policy must be consistent with practices in the discipline, relevant legislation, codes and guidelines." It goes on to say, "[i]nstitutions must provide facilities for the safe and secure storage of research data and for maintaining records of where research data are stored."[22] The Code directs researchers to manage their research data and primary materials according to their institutional policy. Although the Code is not stringently applied nor enforced by ARC, it has been an incentive for improving RDM practices in Australia and a number of universities have developed RDM policies.

Despite the lack of strong policies, Australia is still considered one of the leaders in RDM in that it has made major investments in its services and infrastructure. In 2007, the Australian Government through the National Collaborative Research Infrastructure Strategy Program created the Australian National Data Service (ANDS). ANDS invests in and hosts a number of local and centralized services, including Research Data Australia, a national discovery service to promote visibility of research data collections.

***Germany*:** The German Research Foundation (DFG), one of the major scientific funding agencies in Germany includes a section about "data handling" in all grant proposals. This section asks researchers to describe if and how data will be made available for future reuse by other researchers.[23] Researchers can request funding for making research data available for future reuse, but they must also describe how the institutions

[19] http://europa.eu/rapid/press-release_IP-13-1257_en.htm
[20] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
[21] http://ppl.app.uq.edu.au/content/4.20.06-research-data-management
[22] http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/r39.pdf
[23] http://www.dfg.de/formulare/54_01/54_01_en.pdf

participating in the project will contribute to data and information management."[24] In addition, in 2010, the Alliance of German Science Organizations adopted a set of principles for the handling of research data[25].

***Others*:** Funders and research organizations in Chile[26] and Finland[27] are also in the process of developing RDM policies in their jurisdictions. The Fundação para a Ciência y a Tecnologia in Portugal recommends, but does not (yet) require, that researchers applying for funding share the primary data and other materials produced in projects financed by FCT with other researchers and include data management plans in their funding applications.[28]

3.4 Canada

In 2010, the three federal granting agencies released a set of principles relating to access to research outputs, which states the following:

"CIHR, NSERC and SSHRC are committed to developing a shared approach for improving access to publicly funded research in keeping with internationally recognized best practices, standards and policies for funding and conducting research. The following principles will guide the agencies in promoting access to research results:

- Advance Knowledge: The advancement of knowledge depends upon peer review to ensure excellence, as well as long-term preservation to ensure that research can be built upon by others
- Minimize Research Duplication: Broad dissemination increases the effectiveness of public investments in research by reducing the potential for unnecessary duplication
- Maximize Research Benefits: Publicly funded research should be as accessible as possible in order to maximize the economic, social, cultural and health benefits for Canadians
- Promote Research Accomplishments: Improving access to research results will better promote the accomplishments of Canadian researchers throughout the world"[29]

These principles are consistent with other international statements about research data management and sharing such as the 2004 OECD Declaration and the 2013, G8 Science Ministers' Statement[30] addressing open scientific research data.

As part of these broader policy objectives to improve access to the results of research and to increase the dissemination and exchange of research results, the TC3 announced a new "Tri-Agency Open Access Policy on Publications"[31] in early 2015. This policy requires all grant recipients to ensure that any peer-reviewed journal publications

[24] Ibid
[25] www.allianzinitiative.de/en/core_activities/research_data/principles
[26] http://datoscientificos.cl/files/manual-2014.pdf
[27] http://www.aka.fi/en-GB/A/Funding-and-guidance/How-to-apply/Appendices/Research-plan
[28] https://www.fct.pt/documentos/PoliticaAcessoAberto_Dados.pdf (translated in English by automated translating tool)
[29] http://www.science.gc.ca/default.asp?Lang=En&n=9990CB6B-1
[30] https://www.gov.uk/government/news/g8-science-ministers-statement
[31] http://www.science.gc.ca/default.asp?lang=En&n=F6765465-1

arising from Agency-supported research are freely accessible within 12 months of publication.

In terms of data management, the 2011 "Tri-Agencies Framework: Responsible Conduct of Research"[32] outlines the appropriate conduct of research. Section 2.1.2 of the framework states that at minimum, researchers are responsible for:

a. Using a high level of rigour in proposing and performing research; in recording, analyzing, and interpreting data; and in reporting and publishing data and findings

b. Keeping complete and accurate records of data, methodologies and findings, including graphs and images, in accordance with the applicable funding agreement, institutional policies and/or laws, regulations, and professional or disciplinary standards in a manner that will allow verification or replication of the work by others

Several Canadian research funders have also adopted more explicit RDM policies. CIHR requires all grant recipients to retain original data sets arising from CIHR-funded research for a minimum of five years after the end of the grant. This applies to all data, whether published or not. In addition, for research in some areas (bioinformatics, atomic, and molecular coordinate data), data must be deposited into public repositories.[33]

SSHRC has had a "Research Data Archiving Policy" in place since 1990. The policy states, "[a]ll research data collected with the use of SSHRC funds must be preserved and made available for use by others within a reasonable period of time. SSHRC considers 'a reasonable period' to be within two years of the completion of the research project for which the data was collected." However, SSHRC has not actively enforced the policy and few SSHRC funded researchers are aware of it.

In 2008, Genome Canada adopted a "Data Research and Resource Sharing Policy". The policy "expects data to be released and shared no later than the original publication date of the main findings from any datasets generated by that project."[34] In addition, applicants must provide a Data and Resource Sharing Plan as part of their application.

NSERC does not have an RDM policy, but it has implemented requirements in the context of specific programs including International Polar Year and the Discovery Frontiers (jointly with Genome, CIHR, and CFI). The International Polar Year (IPY) was a large scientific program focused on research in the Arctic and the Antarctic regions from March 2007 to March 2009. With a budget of $1.2 billion US, IPY involved more than 60 countries, over 200 international research networks, and thousands of researchers. In order to meet its objectives of interdisciplinary and international collaboration and to ensure a lasting legacy, IPY committed to ensuring full, free, and open access to IPY data as described in the IPY Data Policy.[35] This was one of the most comprehensive policies for research data at that time and remains so to date.

The Heart and Stroke Foundation, which provided $38 million in funding to 1,500 researchers in 2013, has a data policy that requires grant recipients to deposit

[32] http://www.rcr.ethics.gc.ca/eng/policy-politique/framework-cadre/
[33] http://www.cihr-irsc.gc.ca/e/46068.html#5.1.2
[34] http://www.genomecanada.ca/medias/PDF/EN/DataReleaseandResourceSharingPolicy.pdf
[35] http://www.api-ipy.gc.ca/pg_IPYAPI_050-eng.html

bioinformatic, atomic, and molecular coordinate data into the appropriate public database immediately upon publication of research results.[36]

The following table compares the different elements of research data management policies at some of the major agencies reviewed for this scan.

Table 3: Comparison Table of Funders' Policies

Agency	Domains	Coverage	Timing for data sharing	Monitoring	DMPs	Specified Repository	Funds available for RDM
Australia-ARC	All	All	Not specified	No	No	No	Not specified
Canada-CIHR	Health	All	Upon publication of research results	No	No	bioinformatics, atomic, and molecular coordinate data into public databases	No
Genome Canada	Genomics	All	Upon publication	Yes	Yes	Yes	Not specified
Canada-Heart and Stroke Foundation	Life sciences	All	Upon publication	?	No	Yes- when a repository is available in a discipline	Not specified
Canada-SSHRC	Humanities and social sciences	All	Within two years of the completion of the research project for which the data was collected	No	No	Institutional or domain repository	Not specified
European Commission-Horizon 2020	All	Selected areas	Not Specified	Yes	Pilot project with specific disciplines - opt out available	Available repository	Yes
Portugal-	All	All, but voluntary	Not specified	No	Yes	No	Not specified
UK- AHRC	Arts and humanities	All	At least within three years after the end of the grant	No	Yes	No	Yes

[36] http://www.hsf.ca/research/en/hsf-open-access-research-outputs-policy-guidelines

UK- BBSRC	Biotechnology and biological sciences	All	No later than the release of main findings through publication	Yes	Yes	In specific disciplines	Yes
UK- CRUK	Health (cancer research)	All	no later than the acceptance for publication of the main findings	Yes	Yes	No	No
Agency	Domains	Coverage	Timing for data sharing	Monitoring	DMPs	Specified Repository	Funds available for RDM
UK- EPSRC	Engineering and physical sciences	All	Not specified	Institutions must develop a roadmap	No	Yes- Institutional responsibility	No
UK- ESRC	Economic and Social Research	All	At or around the time of publication	Yes	Yes	Yes-UK Data service	Yes
UK- MRC	Medical	All	Not specified	Yes	Yes	No	Yes
UK- NERC	Environmental science	All	No later than of two years from the end of data collection	Yes	Yes	Yes-NERC data centres	Yes
UK- STFC	Science and technology	All	Not specified	Review of DMPs	Yes	No	Not specified
UK- Wellcome Trust	Biomedical sciences	data holding significant value as a resource for the wider research community	Upon publication of their research	End of grant report	Yes	Yes- discipline and institutional repositories	Yes
US- NIH	Health	All	No later than the acceptance for publication of the main findings from the final data set	Yes	For grants that exceed 500k/year	No	Yes
US- NSF	Sciences and Engineering	All	within a reasonable time	Yes	Yes	No	Yes

3.5 Other Stakeholders

Funders are not the only actors developing policies related to research data management and sharing. Journals, projects and institutions are also adopting policies. Typical policy elements of these actors mirror those of funders, with variations based on community standards and practices, as well as the availability of repositories.

Some journals, mainly in the life sciences, require researchers to deposit any data available related to their articles into a repository for verification. These policies typically require authors to make their data available to others, and where public repositories exist, that authors deposit their data into these repositories. There is some indication that journal data policies may soon expand beyond the life sciences. PLOS, for example, a major publisher in the sciences and medicine, has recently implemented a policy for all of its journals requiring "authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception."[37] That being said, a recent review of the current state of journal sharing policies found that there is still a large percentage of journals that do not have policies on data sharing, and where policies exist, they are often vague and rarely enforced.[38]

A growing number of research projects are also implementing RDM policies. These policies are most common in the context of international projects, in which it is important to establish common approaches across jurisdictions. The International Polar Year is one example, as is the International Barcode of Life (iBOL) project. iBOL has a data release and resources sharing policy, which "seeks to accelerate the timely development of products that will benefit humankind by providing rapid access to the primary outputs from iBOL: DNA sequences associated with high quality meta-data including taxonomic assignments. The working philosophy of the iBOL project is full release of data within 18 months of a sequence being produced. There is the expectation that this 18-month time period will be reduced as the project progresses, and from the outset more rapid data release is encouraged whenever practical."[39]

Ocean Networks Canada has a data policy that states, "[d]ata collected by Ocean Networks Canada are for research and education purposes (not for profit), and are generally open access and free to anyone. Ownership of the data lies with the instrument owner, which in most cases is Ocean Networks Canada."[40] The policy does contain exceptions to availability for preliminary data or in cases where data will be used for commercial purposes.

Universities have also begun to implement RDM policies. Policies are most prolific in the UK[41] and Australia[42], likely because both countries have funding agencies that have placed some responsibility for RDM with the institution. According to the Digital Curation Centre, there are currently 20 UK universities with an RDM policy, and 6 that have draft

[37] http://www.plosone.org/static/policies.action#sharing
[38] Sturges, Paul and Bamkin, Marianne and Anders, Jane H.S. and Hubbard, Bill and Hussain, Azhar and Heeley, Melanie (2014) *Research data sharing: developing a stakeholder-driven model for journal policies.* Journal of the Association for Information Science and Technology. ISSN 2330-1643 (Available at: http://eprints.nottingham.ac.uk/3185/)
[39] http://ibol.org/resources/data-release-policy/
[40] http://www.oceannetworks.ca/data-tools/data-help/data-policy
[41] http://www.dcc.ac.uk/resources/policy-and-legal/epsrc-institutional-roadmaps
[42] http://www.ands.org.au/datamanagement/policy.html

policies in development.[43] The Australian National Data Service currently lists 5 institutional policies in Australia.[44]

University RDM policies tend to contain similar elements as those found in funders' policies, while also defining the specific role for the institution. The University of Edinburgh provides an example of a very comprehensive university policy that could act as an exemplar for other institutions.[45] Through this policy, the institution accepts responsibility for tracking, preserving and supporting researchers in managing their data. The policy requires faculty to develop data management plans. It also commits the university to providing support, training and "mechanisms" for storage and sharing. The university acknowledges that it is an aspirational policy and will take some years to implement fully. Clearly, even the most advanced universities in terms of RDM services and infrastructure would not assume that all researchers are able to adhere to requirements, and policies must still be phased-in.

Many institutions have policies on ethical conduct that require the proper handling of data by the researchers including accurate presentation and retention of data. Two Canadian universities also have more specific policy requirements addressing data management practices. The University of Alberta promotes the use of data management plans (DMPs) in the context of the University's policy on the stewardship of research records. The University's "Research Records Stewardship Guidance Procedure" makes two references to Data Management Plans: "The articulation of the primary stewardship responsibilities for all parties throughout the research lifecycle should be made at the very beginning of a research project in a Data Management Plan," and "[w]ith regard to human participant research generally, records do not have to be destroyed, provided the researcher's Data Management Plan has a clear statement about appropriate records management, storage and retention."[46]

The University of PEI has a policy on "Open Access and Dissemination of Research Output", which encourages the deposit of research data into the UPEI Virtual Research Environment (VRE). It asserts that, "research data be made accessible in a fashion and timeline deemed appropriate by the researcher/research group. Where possible, research data would be made publicly accessible on publication of results of the research. Where privacy rights of human subjects conflicts with full public access, the researcher/research group will aim for the most public access possible and consistent with privacy, for example by providing anonymized data, or providing full access to data to other research groups that can demonstrate having met acceptable research ethics guidelines for handling such private information."[47]

[43] http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies
[44] http://www.ands.org.au/datamanagement/policy.html
[45] http://www.ed.ac.uk/schools-departments/information-services/about/policies-and-regulations/research-data-policy
[46] https://policiesonline.ualberta.ca/PoliciesProcedures/Procedures/Research-Records-Stewardship-Guidance-Procedure.pdf
[47] https://cab.upei.ca/sites/default/files/attachments/OpenAccessandDisseminationofResearchOutput.pdf

4. Data Management Plans

Increasingly, many research funders require a data management plan (DMP) as a component of funding applications. DMPs are formal documents that "typically state what data will be created and how, and outline the plans for sharing and preservation, noting what is appropriate given the nature of the data and any restrictions that may need to be applied."[48] DMPs are seen as a way of improving data management practices during the research process and they compel researchers to establish how they plan manage their data in advance of a project. Writing a DMP helps organize the research process and provides consistent guidelines for handling data, making the research process more efficient. In addition, DMPs can reduce the costs of research, as early planning for research data management has been shown to significantly reduce costs of data management over the long term.[49]

In all the examples reviewed for this scan, DMPs represent only one element of a broader research data management policy. The policy provides the requirements and the DMP outlines how requirements will be adhered to. From a policy perspective, DMPs are an important tool for ensuring that researchers are aware and have a plan to adhere to policy requirements in advance of starting to collect their data.

Common requirements for DMPs have been outlined in a checklist developed by the DCC in the UK and are documented below. The emphasis on different elements will vary and often depend on the focus of the policy requiring the DMP. For example, a policy focus on data sharing may emphasize the data sharing elements of the policy over other elements.

Table 4: DCC Checklist for Data Management Plans

Data Collection	What data will be collected or created? How will the data be collected or created?
Documentation and Metadata	What standards, documentation and metadata will accompany the data?
Ethics and Legal Compliance	How will ethical issues be managed? How will copyright and Intellectual Property Rights (IPR) issues be managed?
Storage and Backup	How will the data be stored and backed up during the research? How will access and security be managed?
Retention and Preservation	Which data should be retained and/or preserved? What is the long-term preservation plan for the data?
Data Sharing	How will the data be shared? Are any restrictions on data sharing required?

[48] http://www.dcc.ac.uk/resources/data-management-plans
[49] http://ukdataservice.ac.uk/manage-data/plan/costing.aspx

Responsibilities and Resources	Who will be responsible for data management? What resources will be required to deliver the data management plan?

In practice, DMP requirements vary significantly across organizations. Most of the differences lie in the level of detail required in the plan and the types of guidance and support offered. Some notable areas of variation are described below.

***Level of detail*:** Some organizations require very comprehensive descriptions across all elements listed above, while others ask for much less detail. In addition, some organizations are very prescriptive about how the DMP should be formatted, while others leave it up the individual researchers to include whatever information they think is relevant. The trend seems to be towards greater specificity of requirements, as this helps to guide researchers in terms of what a data management plan should entail.

***Timing and versioning of DMPs*:** In the majority of cases, DMPs must be completed and attached to the funding proposal in advance of the project launching. However, there are a few exceptions. In the Horizon2020 Pilot Project for example, the European Commission expects the first version of the DMP to be delivered within the first 6 months of the project, with more elaborated versions delivered at later stages of the project. The EPSRC in the UK requires data management plans, but does not review them. DMPs may change over time, and some organizations (like Horizon2020) ask researchers to update their DMPs regularly if there are changes to the original plan.

***Scope*:** The scope of the data to be addressed in the DMP also varies across funders. Some policies target only the research data that underpins the publications. Others, such as the Wellcome Trust, require a DMP only when the research "involves the generation of datasets that have clear scope for wider research use and hold significant long-term value."[50] Still others require that all data produced in a project be made available, for example, the NIH requests plans for "all data from funded research that can be shared without compromising individual subjects' rights and privacy, regardless of whether the data have been used in a publication."[51]

4.1 Sample Data Management Plans

In order to demonstrate the different approaches to DMPs, a number of examples are offered below.

Economic and Social Research Council (UK): [52] All ESRC grant applicants planning to create data during their research have to include a data management plan with their application. A data management plan helps to decide how research data will be managed throughout the research cycle and will be available for sharing afterwards. Most research data can be successfully archived and shared.

ESRC expects award holders to consider all issues related to confidentiality, ethics, security and copyright before initiating the research. Any challenges to data sharing (e.g. copyright or data confidentiality) should be critically considered in a plan, with possible solutions discussed to optimize data sharing.

[50] http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/index.htm#_B._When_is_a%20data%20management%20and%20sh
[51] http://grants.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm#901
[52] http://ukdataservice.ac.uk/manage-data/plan/dmp-esrc.aspx

A data management plan includes the following topics:

- Assessment of existing data
- Information on new data
- Quality assurance of data
- Backup and security of data
- Expected difficulties in data sharing
- Copyright/Intellectual Property Rights
- Responsibilities
- Preparation of data for sharing and archiving

Detailed guidance about preparing data management plans is provided by ESRC through the UK Data Archive.

Engineering and Physical Science Research Council (UK): The EPSRC does not require researchers to submit data management or sharing plans in grant applications. However, it does expect policies and plans to be in place.[53] In clarifying its policy, the EPSRC states, "[i]t is suggested that research offices ensure appropriate provision for research data management is included in a research proposal before it is submitted to EPSRC. In particular: a) does a data management plan (DMP) exist? (EPSRC does not require DMPs with research grant applications, but our research data principles include that "…project specific data management policies and plans… …should exist for all data')"[54]

Medical Research Council (UK): [55] All applicants submitting funding proposals to the MRC are required to include a DMP as an integral part of the application. The council asserts that everyone in a research team should have a clear sense of their responsibilities. Specific elements of the data management plan outlined in a template provided by MRC are as follows:

- Description of the data
- Data collection / generation
- Data management, documentation and curation
- Data security and confidentiality of potentially disclosive information
- Data sharing and access

Detailed guidance is also available to researchers if needed.

National Institutes of Health (US): Starting with the October 1, 2003 receipt date, investigators submitting an NIH application seeking $500,000 or more in direct costs in any single year are expected to include a plan for data sharing or state why data sharing is not possible. NIH is less prescriptive about the contents of data management plans, stating "[t]he precise content and level of detail to be included in a data-sharing plan depends on several factors, such as whether or not the investigator is planning to share data, the size and complexity of the dataset, and the like."[56]

It then provides a number of examples of different types of data management plans to assist researchers in developing their own. However, NIH states that the plans should ideally cover the following elements:[57]

[53] http://www.dcc.ac.uk/resources/policy-and-legal/research-funding-policies/epsrc
[54] http://www.epsrc.ac.uk/files/aboutus/standards/clarificationsofexpectationsresearchdatamanagement/
[55] http://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/data-management-plans/
[56] http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#ex
[57] http://grants.nih.gov/grants/sharing_key_elements_data_sharing_plan.pdf

- What data will be shared
- Who will have access to the data
- Where will the data be available
- When will the data be shared
- How will researchers locate and access the data

National Science Foundation (US): Each proposal must include a supplementary document of no more than two pages labeled "Data Management Plan". This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results.

The data management plan should include the following information:[58]

1. Products of the Research: The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project
2. Data Formats: The standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies)
3. Access to Data and Data Sharing Practices and Policies: Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements
4. Policies for Re-Use, Re-Distribution, and Production of Derivatives
5. Archiving of Data: Plans for archiving data, samples, and other research products, and for preservation of access to them
6. Certain directorates within the NSF, however, provide explicit guidelines and advice on forming data management plans, which may require more details[59]

5. Administering Policies

Funders with RDM policies have taken a variety of approaches to administering them. Some of the key issues and current practices in terms of policy administration, such as guidance, evaluating DMPs, monitoring compliance, confidentiality and IP, and dealing with multiple policies are discussed here.

5.1 Policy Guidance

Clear and detailed guidance on how to adhere to a data policy is essential to ensure compliance. However, given the intricacies of research data management, guidance can also become so complex that it is confusing for users. University of Edinburgh was one of the first UK universities to provide online research data management guidance in 2009. The resource aims to assist university researchers in complying with the increasingly demanding requirements of both external funding bodies and the university, and direct them to appropriate sources of support. Although well received, the guide was considered overly complex and had to be re-vamped. The new, much briefer version

[58] http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#IIC2j
[59] http://www.arl.org/focus-areas/e-research/data-access-management-and-sharing/nsf-data-sharing-policy/243-resources-for-data-management-planning#.VGTXGYd7Sqk

consists of eight pages covering the essential topics researchers must understand before embarking on a research project.[60]

In the US context, both NIH and NSF have recognized that guidance about adhering to their policy is critical, but also a challenge because practices and infrastructures vary significantly across disciplines and sub-disciplines. In order to navigate this, NSF has taken the approach of allowing the disciplines themselves to define best practices around managing data, and relies heavily on input from these communities for guidance and peer-review. It has been reported that, at least in the initial stages of this process, some disciplines are beginning to develop valuable resources to which researchers can turn for guidance.[61]

Given the complexities of RDM, some organizations have opted to provide access to experts who can give individualized support for RDM to researchers. For example, the Medical Research Council in the UK is developing a Data Support Service to facilitate and support data sharing for population and patient studies, in order to optimize the long-term use of rich data assets for new science. The project works closely with MRC data managers to coordinate and promote work on data sharing tools and standards and to promote the exchange of good practice. The Digital Curation Centre also provides numerous resources as well as consulting services to institutions across the UK that are developing RDM support services for their communities.

5.2 Evaluating DMPs

Funders have also taken different approaches to evaluating DMPs. According to the literature, most of the UK funding councils are assessing DMPs during the initial peer review process. The Economic and Social Research Council (ESRC), for example, says that it "will seek an assessment of data management plans via its peer review and assessment processes. Although the application will first and foremost be assessed on grounds of its scientific merit, nonetheless, an assessment of the data management and sharing plan will be included in the general assessment of the application. "[62] A poorly prepared DMP, it goes on to explain, may have a detrimental effect on an otherwise strong application.

The Biotechnology and Biological Sciences Research Council (BBSRC) evaluates DMPs separately from the scientific excellence of the proposed research, "however, an application's credibility will suffer if peer review agrees the statement is inappropriate. In the case where a highly rated proposal has an inappropriate Data Management Plan, Committees and Panels may choose to offer conditional awards and/or provide specific feedback to the applicants."[63]

Genome Canada states that staff and review committees "will review the applicant's proposed data and resource sharing plan to verify that it conforms to the Genome Canada policy and funds will not flow until an acceptable plan has been approved and incorporated into the terms of award."[64]

Other funders have chosen not to include the DMP as a part of the proposal evaluation at all. The NIH states that reviewers will not factor the proposed data-sharing plan into

[60] http://www.ijdc.net/index.php/ijdc/article/view/8.2.194/327
[61] http://www.asis.org/Bulletin/Aug-14/AugSep14_Kozlowski.html
[62] http://www.esrc.ac.uk/_images/Research_Data_Policy_2010_tcm8-4595.pdf
[63] http://www.bbsrc.ac.uk/web/FILES/Policies/data-sharing-policy.pdf
[64] http://www.genomecanada.ca/medias/PDF/EN/DataReleaseandResourceSharingPolicy.pdf

the determination of scientific merit or priority score. However, program staff members are responsible for overseeing the data-sharing policy and assessing the appropriateness of the plan. In other words, NIH staff will evaluate the content of the plan based on whether it provides comprehensive information about how researchers will manage the data. Any concerns must be resolved prior to making any award.[65] Presumably, the NIH has developed the expertise internally taking the responsibility off of the peer-review committees.

In general there are four potential options for evaluating DMPs in the proposal stage:

1) The DMP is reviewed as part of the excellence review. Assessment is a full-weight component of the excellence assessment and can impact the adjudication
2) The DMP is reviewed separately from the excellence review, but with an impact on acceptance of proposal
3) The DMP is reviewed separately from the excellence review and has no impact on acceptance of proposal
4) No review process

Table 5 highlights the approaches of the different funders where information is available.

Table 5: Funders approaches to assessing DMPs

Funder	Assessed as part of peer review process	Assessed separately, impact on proposal	Assessed separately, no impact on proposal	Not assessed during peer review process
Genome Canada		**X**		
EC- Horizon 2020				**X**
UK- BBSRC		**X**		
UK- CRUK	**X**			
UK- ESRC	**X**			
UK- MRC	**X**			
UK- STFC	**X**			
US- NIH			**X**	
US- NSF	**X**			

It can be challenging to assess data management plans as part of a funder's peer review process, depending on the discipline and the expertise of committee members. In the scenario where the peer review committees are reviewing the plans, they may not have sufficient knowledge to be able to determine quality. It has been reported, for example, that reviewers of NSF proposals (from disciplinary peer-review committees) who rely on the general policy guidelines have found it difficult to identify the components of a good

[65] http://grants.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm#901

data management plan.[66] To address this, some organizations have developed guidance for peer reviewers. John Hopkins University Library developed a checklist to assist NSF proposal reviewers[67] and both the MRC and ESRC in the UK have published guidance documents to assist their peer review committees in evaluating the quality of DMPs.

5.3 Monitoring Compliance

A variety of approaches are also being used to monitor adherence to the policies. Most commonly, researchers/projects are required to provide a written report about how they have adhered to the policy requirements in their final reports.

According to the DCC, several of the UK funding councils are actively monitoring compliance with their policies via the final report process. The Engineering and Physical Sciences Research Council (EPSRC) is monitoring progress and compliance on a "case-by-case basis". Both the Economic and Social Research Council (ESRC) and the Natural Environment Research Council (NERC) state that they are prepared to withhold the final grant payment if data are not properly managed and offered for deposit. However, the extent to which such penalties are applied is unclear."[68] In addition, ESRC expects grant holders to report about "the on-going implementation of the data management and sharing plan through annual reporting to ESRC".[69] It is unclear to what degree and through which methods the US funders and other agencies are monitoring compliance.

Ultimately, comprehensive monitoring will require that data sets can be tracked. There are a number of tools that are emerging to help improve the discoverability of datasets. Data citation using permanent identifiers, such as Digital Object Identifiers (DOIs) provides a permanent ID for datasets, which is helpful for tracking data that has been deposited into a public repository. In addition, there are initiatives in both Australia and the UK to improve the visibility and discoverability of data sets, both to support access and re-use. ANDS has developed a dataset registry, called Australian Research Data Commons, to make better use of Australia's research data outputs.[70] Similarly, Jisc (a membership organization that supports the use of digital technologies in UK education and research community) and the DCC are currently working on a UK registry, which will "provide a coherent point of access to discoverable, searchable, browsable and actionable descriptions of given datasets and how to access them, and so showcase the wealth of UK research data".[71] These registries not only assist policy makers in tracking datasets, but also contribute to the ultimate aims of RDM policies by improving the discoverability of research data and supporting re-use. They also contribute to a system where data can be cited.

5.4 Confidentiality and Intellectual Property

Almost all RDM policies contain clauses addressing the issues of privacy and intellectual property, and some also have clauses dealing with other types of sensitive data.

[66] http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/aag_6.jsp#VID4
[67] http://dmp.data.jhu.edu/resources/grant-reviewers-guide/
[68] http://www.dcc.ac.uk/sites/default/files/documents/RC%20policy%20overview%20v2.2.pdf
[69] http://www.esrc.ac.uk/_images/Research_Data_Policy_2010_tcm8-4595.pdf
[70] http://researchdata.ands.org.au
[71] http://www.dcc.ac.uk/projects/research-data-registry-pilot

Referred to as “ethical open access” or “intelligent open access”, policies must aim to strike a balance between the rights and interests of investigators, study participants, and the public. This balance is not always easy to achieve.

In Canada, university research ethics boards are very concerned with ensuring privacy of study participants and often enforce stringent practices restrict researchers’ abilities to share data. They base their approach on the “Tri-Council Policy Statement on the Ethical Conduct for Research Involving Humans” (TCPS), which sets out privacy and confidentiality requirements for researchers working with human participants, including for secondary use of research data. The policy statement emphasizes that respect for privacy in research is an internationally recognized norm and ethical standard. These codes can conflict with data sharing policies, if applied to stringently. For example, research ethics boards may request a plan for data disposal, which could conflict with a funding agency policy that requires data retention and sharing.

There are, however, established best practices to ensure that the confidentiality of study participants is protected. In cases where the data cannot be modified to protect confidentiality without significantly compromising the research potential of the data, data is restricted and confidentiality safeguards are imposed. Many funding agencies therefore require practices such as anonymization to be adopted by researchers before releasing data. In cases where anonymization is not possible, then researchers must explain why in their DMP or grant application. For example, the Wellcome Trust deals with confidentiality in the following way:

> “In designing studies, researchers must ensure that they have appropriate systems to protect the confidentiality and security of data pertaining to human subjects, and minimise any risks of identification by data users. This can be achieved through the use of appropriate anonymisation procedures and managed access processes. Such systems should be sufficient to safeguard participants, but proportionate to the level of sensitivity of the data and associated risk. They should not unduly inhibit responsible data sharing for legitimate research uses.”[72]

In terms of intellectual property, many policies have exceptions for data that have potential commercial value. In these cases, they try to strike a balance between the value of broad data sharing and deriving any commercial benefits from the research. Approaches to protecting IP usually involve implementing embargo periods to allow for patent applications. For example, the NSF policy states:

> “It is NSF’s strong expectation that investigators will share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. However, it is also necessary to protect intellectual property rights and potential commercial value. The Data Management Plan should describe the proposed approach, which will then be subject to peer review and program management. (For example, research use of sensitive data is often allowed through reasonable binding agreements that contain confidentiality provisions.)”[73]

The NIH “recognizes that the investigators who collected the data have a legitimate interest in benefiting from their investment of time and effort. NIH continues to expect

[72] http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/index.htm#five
[73] http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp

that the initial investigators may benefit from first and continuing use but not from prolonged exclusive use."[74]

In the UK, IP related to research falls under the auspices of the institution. Therefore, the UK funding councils tend to rely on institutional policies to set requirements for IP issues around data. In Canada, at most universities it is the researcher/creator who owns the IP, although there are some exceptions to this.[75]

A few policies also include exceptions for "sensitive" data. This is particularly important in research areas dealing with indigenous knowledge and national security. "The Statement of Principles and Practices for Arctic Data Management" published in April 2013 by the International Arctic Science Committee (IASC) provides a framework for handling data collected from indigenous communities, (as does guidance contained in the TCPS2):

> "In the context of research involving Indigenous knowledge, data management principles based on the concepts of respect, reciprocity, and responsibility should be observed. This includes appropriate engagement of Indigenous people, communities or organizations throughout the entire data life cycle, formal attribution of contributed knowledge, establishment of informed consent for use of knowledge and derived products, and the maintenance of contributor control of data and information resources. Required institutional ethics review processes (e.g. Institutional Review Boards, Research Ethics Boards etc.) will guide data management, however Indigenous communities or organizations may have specific practices or requirements in place. It is the responsibility of researchers to familiarize themselves with and adhere to these practices and requirements." [76]

5.5 Multiple Policies

Co-funding is becoming a common practice and many researchers may find themselves in a position where they are subject to the requirements of two or more funding agencies, in addition to the requirements or their institution and journal. This can be a particularly problematic issue for researchers that are co-funded by government and private industry, since companies may seeks to protect intellectual property resulting from research, which could include the data. In these cases, researchers could be asked to create interagency agreements concerning data management that would be shared and approved by all funders.

A few funders have developed guidance for researchers about how to adhere to policies in the case of multiple funders with differing requirements. The NIH advises grantees as follows, "[t]he NIH recognizes that there may be circumstances where a co-funder has requested restrictions on data sharing as a condition of funding. These restrictions should be identified in the application and a proposal made about how data from the co-funded project will be shared. Should you believe that you are unable to share any of the data, your justification will be considered by NIH program staff."[77]

---

[74] http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#time
[75] http://blogs.sfu.ca/departments/cprost/wp-content/uploads/2012/10/IP-Policy-Introduction-January-2010FINALCombined.pdf
[76]http://www.innovation.ca/sites/default/files/Rome2013/files/IASC%20Statement%20on%20Arctic%20Data%20Management_2013.pdf
[77] http://www.data-archive.ac.uk/media/247429/costingtool.pdf

Harmonization of approaches and policies across agencies (both in Canada and internationally) can go a long way to help address some of these issues that arise when dealing with multiple funders.

5.6 Costs

Adhering to research data management policies can incur extra costs. A crucial factor is the type of data created. The larger and more complex the data being managed, the greater the effort required and the greater the potential costs.[78] Most funders reviewed in this scan consider data management activities as being eligible expenses within a project budget.

In addition, there are funders that fund and maintain data repositories to support data archiving, although the scope of these domain repositories do not cover all datasets produced through their funded research.

There are costs for data management that fall across the entire data lifecycle. The UK Data Service has compiled a detailed list of the costs associated with research data management that are incurred during the life of the project. [79]

Table 6: Costs for Research Data Management

Data storage	Operationalizing data
Data transfer and access	Data description
Data backup	Data cleaning
Data security	Data documentation
Consent for data sharing	Metadata
Transcription	Formatting and organizing
Anonymization	Digitization
Data sharing	Data format

The real costs for any given project will depend to a large extent on the nature of the data collected. More information about potential costs is provided in the Implementation Challenges section of this report.

6. Approaches to Policy Implementation

This section describes the various approaches taken to implementing RDM policies in different jurisdictions. Regardless of approach, experiences of others demonstrate that full adherence to policies takes time.

6.1 Engineering and Physical Sciences Research Council (UK)

[78] http://www.ed.ac.uk/schools-departments/information-services/research-support/data-management/how-manage-data
[79] http://www.data-archive.ac.uk/media/247429/costingtool.pdf

In the UK, the funding agencies have taken a strong stance through comprehensive policy adoption across all agencies based on a set of common principles. In 2011, the seven RCUK councils adopted a set of common principles for research data management. Each council was expected to develop a policy that adheres to those principles within a certain time frame.

The policies across the 7 agencies differ and reflect the specific disciplinary context served by each council. 6 of the councils have requirements for data management plans, expect research projects to manage their data according to certain standards, and require them to share data at the end of a project. The exception is the EPSRC, which has taken a different approach by placing some of the responsibility for adherence to the policy on the institutions. In April 2011, all UK Vice Chancellors received a letter from the EPSRC, that sets out 9 "expectations" of organisations in receipt of EPSRC research funding which are summarized below: [80]

1. Research organisations will promote internal awareness of these principles and policies
2. Published research papers should include a short statement describing how and on what terms any supporting research data may be accessed
3. All of their researchers or research students funded by EPSRC will be required to comply with research organization policies in this area or, in exceptional circumstances, to provide justification of why this is not possible
4. Publicly-funded research data that is not generated in digital format will be stored in a manner that facilitates it being shared should a valid request for access to the data be received
5. Research organisations will ensure that appropriately structured metadata describing the research data they hold is published (normally within 12 months of the data being generated) and made freely accessible on the Internet. Where the research data referred to in the metadata is a digital object it is expected that the metadata will include use of a robust digital object identifier (e.g. DOI from DataCite)
6. Where access to the data is restricted the published metadata should also give the reason and summarise the conditions
7. Research organisations will ensure that EPSRC-funded research data is securely preserved for a minimum of 10 years
8. Research organisations will ensure that effective data curation is provided throughout the full data lifecycle
9. Research organisations will ensure adequate resources are provided to support the curation of publicly-funded research data

The framework also requires organizations (rather than researchers) to identify and map out the steps they will take to achieve full compliance to the Roadmap. The EPSRC has given the universities a deadline of May 1st, 2015 to achieve full compliance.

The EPSRC has stated that it may request to see individual roadmaps on a case-by-case basis and could require evidence of activity to achieve compliance at any time. If, after May 2015, an institution is found to be deliberately obstructing the sharing of

[80] http://www.epsrc.ac.uk/about/standards/researchdata/expectations/

research data or otherwise seriously failing to comply with the ESPRC's expectations, then the EPSRC may ultimately withdraw its funding.[81]

This framework has been the impetus for a number of UK universities to invest in research data management, including services, infrastructure and storage and the DCC now lists 10 institutional roadmaps that have been made public with more in development.

According to a report published by the Society for Research into Higher Education in March 2013,

> "This move on the part of the EPSRC […] is pushing universities to review their research data management practices, develop Research Data Management policies, and investigate the resource and infrastructure implications. Some universities have already introduced data preservation and sharing policies requiring their researchers to address, at the outset of their projects, the question of data management and sharing (e.g. Universities of Edinburgh and Oxford); and some have developed institutional data repositories in which academics and PhD students are encouraged to deposit their data (e.g. 'Edinburgh DataShare'). One of the ways in which these new requirements are being institutionalised by universities is by defining data sharing as 'good research practice' and incorporating data management and sharing into university ethics and research governance regulations and procedures."[82]

6.2 Pilot on Open Data (European Commission)

The European Commission (EC) has chosen to introduce an RDM policy incrementally beginning with a pilot project, rather than implementing across-the-board requirements for all EC-funded research projects. The "Pilot on Open Research Data" targets specific research areas (listed below):[83]

- Future and Emerging Technologies
- Research infrastructures – part e-Infrastructures
- Leadership in enabling and industrial technologies – Information and Communication Technologies
- Societal Challenge: Secure, Clean and Efficient Energy – part Smart cities and communities
- Societal Challenge: Climate Action, Environment, Resource Efficiency and Raw materials – with the exception of topics in the area of raw materials
- Societal Challenge: Europe in a changing world – inclusive, innovative and reflective Societies
- Science with and for Society

The pilot areas correspond to about €3 billion or 20% of the overall Horizon 2020 budget in 2014 and 2015.

The aim of the Pilot is to give the Commission a better understanding of what supporting infrastructure is needed and of the impact of limiting factors such as security, privacy or data protection or other reasons for projects opting out of sharing. It will also contribute

[81] http://www.bath.ac.uk/rdso/assets/pdf/University-of-Bath-Roadmap-for-EPSRC.pdf
[82] http://www.srhe.ac.uk/downloads/MauthnerScopingReport.pdf
[83] http://europa.eu/rapid/press-release_IP-13-1257_en.htm

insights into how best to create incentives for researchers to manage and share their data. The pilot will be monitored throughout Horizon 2020 with a view to developing future policy and EU research funding programs.

6.3 Code of Conduct and the Australian National Data Service

In Australia, the "Australian Code for Responsible Conduct of Research"[84] places the onus of responsibility on the universities for managing and preserving data. As discussed earlier, the code requires institutions to retain research data, provide secure data storage, identify ownership, and ensure security and confidentiality of research data. Although the Code is not applied in a strict manner, it has been an inducement for Australian universities to develop RDM services, resulting in more robust repositories and services at Australian institutions than in many other jurisdictions.

Also contributing to a more robust RDM environment is the fact that the Australian government has invested significantly in data management via the Australian National Data Service (ANDS).[85] Research Data Australia is the flagship service of ANDS, which provides a comprehensive window into the Australian Research Data Commons enabling Internet-based discovery to Australia's data, projects, researchers and institutions.[86] Currently, the service aggregates metadata from over 90 collections across Australia, including 22 universities and numerous domain research data centres. ANDS also invests in the development of domain and institutional repositories as well as training and support for RDM.

6.4 Office of Science and Technology Policy (United States)

In February 2013, The White House's Office of Science and Technology Policy (OSTP) released a policy memorandum that directed all 22 federal agencies with more than $100M in R&D expenditures (including the NSF and NIH) to develop plans to make the published results of federally funded research freely available to the public within one year of publication and requiring researchers to better account for and manage the digital data resulting from federally funded scientific research.

In a March 2014 letter to the House and Senate Appropriations Committees, the director of the Office of Science and Technology Policy (OSTP) reported that "all agencies subject to the requirements in the memorandum have now submitted draft plans" and "are currently revising their plans to address OSTP and OMB [Office of Management and Budget] comments and ensure compliance with all of the requirements laid out in the OSTP memorandum."[87]

Although the agencies are working together to try to align their policies in terms of data management, it is likely that there will be significant variations in plans across these agencies, making for a confusing environment for researchers who will be required to adhere to policies.

Little to no extra support and funding has been provided to support the implementation of policies.

---

[84] http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/r39.pdf
[85] http://ands.org.au/resource/ands-business-plan-2013-14.pdf
[86] http://researchdata.ands.org.au/home/about
[87] http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/ostp.html

7. Implementation Challenges

A 2013 survey of over 300 Canadian researchers from across disciplines undertaken by Susan Mowers et. al. provides some indication that many researchers do not use repositories to share their data. The survey found only 4% of respondents shared their data through a "curated digital data repository" [88], and another 14% used an institution repository or a "public domain archive" 81% of respondents indicated that they stored data on their local hard drives.[89]

7.1 Disciplinary Contexts

Implementation challenges must be viewed through a disciplinary lens. Across domains, disciplines and sub-disciplines, the types of data produced and used are extremely diverse; standards differ significantly, as does the availability of infrastructure. In some fields researchers already have a well-established culture of data sharing, there are well-established practices, and support and infrastructures to allow data sharing. In other fields no such mechanisms exist, and others fall in between these two extremes. Many of the challenges with implementing RDM must also be viewed through the particular disciplinary context. "Barriers to effective data sharing and preservation are deeply rooted in the practices and culture of the research process as well as the researchers themselves."[90]

7.2 Researcher Preparedness

Researchers' perspectives towards data sharing are very discipline specific. Surveys and interviews undertaken over the last decade have articulated a wide range of opinions on the topic which cannot be easily generalized into a single statement about researchers' attitudes. Typical objections to data sharing include data ownership and fears of being scooped; the time and skills involved with managing data; and issues of privacy involving data about human participants. A review of the literature across 15 international jurisdictions undertaken in the Netherlands found that "although there are major differences in the way disciplines conduct their research, they also have a number of factors in common when it comes to data storage and access. They all encounter both technical barriers, for example the use of obsolete software, and non-technical ones, such as fear of competition, lack of trust, lack of incentives, and lack of control."[91]

Expertise in the research community is also an important barrier. A survey in the US of researchers at five different institutions found that none of the researchers interviewed had received formal training in data management practices. "None of the scholars interviewed during this study expressed satisfaction with their level of expertise in data management, and few had access to individuals who could provide knowledgeable guidance. On the contrary, most participants reported feeling adrift when establishing

[88] A curated archive refers to a repository where there is active management and appraisal of data over the lifecycle of scholarly and scientific materials - http://digitalcuration.blogspot.ca/2009/08/curated-databases-and-data-curation.html

[89] http://gsg.uottawa.ca/data/open/aa-interim-survey-report/20130801-en.pdf

[90] Tenopir (2011) et. al. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101

[91] SURF Foundation: What Researchers Want. www.surffoundation.nl/en/publicaties/Pages/Whatresearcherswant.aspx

protocols for managing their data and added that they lacked the resources to determine best practices, let alone to implement them."[92]

Many disciplines still lack formalized and standardized procedures for managing research data. There are also large gaps in terms of training for data management. According to a report published by Knowledge Exchange, a multi-national, European co-operative effort that supports the use and development of information and communications technologies (ICT) infrastructure for higher education and research, the most important challenge to data sharing is that it is not yet very common among scholars and is not yet seen as a regular activity among scientists.[93] In interviews they conducted, the main hurdle in data sharing is the individual scientist who is reluctant to put effort into data sharing. "This is mainly for cultural reasons: ownership of the data, workload to properly curate the data making them available for others, and lack of career-reward for making this type of effort."[94] Ongoing promotion and education of researchers will be needed to address these barriers.

7.3 Incentives

Equally significant is the lack of incentives and rewards for data management and sharing. Incentives would greatly accelerate the adoption of RDM practices in the research community and these rewards should be part of the formal evaluation processes at funding agencies and institutions. While far from common, we are beginning to see this happen in some contexts. In 2013, for example, the NSF began to allow datasets to be cited as relevant work products in biographical sketches independent of the related publication(s).[95] The journals that require data sharing can also be an important incentive for researchers, as they ask that researchers deposit their data in order to have their articles published. Services that offer a permanent URL for a dataset, such as DataCite's DOIs, enable citations that can then be used to track and acknowledge the re-use of data. These mechanisms are "paving the way for new metrics and publication models that recognize and reward data sharing" but without actually developing any of these indicators. "The availability of proper metrics can help researchers to make their data work more visible. This may subsequently act as an incentive for more data sharing and in this way a virtuous circle may be set in motion."[96]

7.4 Costs

Costs for data management are often divided into two aspects: the costs of managing and preparing data during the research project; and the costs of providing access and preservation to data once the project is over.

In terms of project-based data management activities, the UK Data Service has developed a costing tool to help researchers anticipate costs of RDM[97]. The tool identifies costs across the data lifecycle, which will be unique to each project depending on the size and complexity of the datasets collected. The costing tool was developed with input from researchers, who were asked to estimate the time or cost needed for

[92] http://www.clir.org/pubs/reports/pub154/problem-of-data
[93] Ibid
[94] Ibid
[95] http://www.asis.org/Bulletin/Aug-14/AugSep14_Kozlowski.html
[96] http://www.knowledge-exchange.info/datametrics
[97] Ibid

activities related to: data collection, data entry and transcription, data validation and documentation and the cost of preparing data for archiving and re-use. The researchers participating in the process "found it hard to cost data management activities, as many activities are an integral part of standard research activities and data analysis."

While the costs of managing data will differ depending on volume and complexity of data, one general rule of thumb suggested via the Jisc Research Data Management listserve is that 5% of the project costs will be for data management activities in cases where "data have high re-use potential, and the data have at least some features (anonymization, complex documentation, size) that might make data preparation more costly." [98] In other cases, where data is not subject to such complicated preparation, the costs could be significantly lower. These costs reflect the costs of data management during the lifespan of the project.

There is a second part of the cost scenario, which represents the costs of preserving and providing access to data after the project is over. These generally fall under the costs of maintaining a data repository. These costs can also vary significantly depending on what level of curatorial service is attached to the repository. In 2012, the Royal Society in the UK undertook an analysis of the costs involved in the long-term management of data.[99] They categorized 4 tiers of data management: Tier 1 and 2 are represented by major international data initiatives that have well defined protocols for the selection and incorporation of new data and access to them (e.g. genomics data) and data centres and resources managed by national bodies (such as UK Research Councils or prominent research funders such as the Wellcome Trust). Tier 3 is curation at the level of individual universities and research institutes, or groupings of them; and Tier 4 is when the individual researcher or research group collates and stores its own data, often making it available via a website to collaborators or for public access. The analysis found that the costs of running a curated data archive[100] ranged from $350,000 US per year for the Dryad repository to $6-7 million for the Worldwide Protein Data Bank. In terms of institutional data repositories, costs were generally lower, but real costs were difficult to determine because institutional data repositories often share infrastructure and staff cross other positions in the same institution.

7.5 Institutional Role

Universities have not traditionally seen research data management as part of their mandate and have been relatively slow to become engaged. This is beginning to change. There has been gradual growth in RDM services provided by universities, usually through the libraries, and it is likely that this trend will continue and expand. In the UK and Australia, many of the larger institutions have already implemented fairly robust services. Funding agency requirements that place some responsibility for data management on the institution have been the impetus for adoption of RDM services. In the US and Canada, several universities are providing support services for researchers, and a few are managing data repositories.

---

[98] RESEARCH-DATAMAN@jiscmail.ac.uk discussion list" <RESEARCH-DATAMAN@JISCMAIL.AC.UK, October 8, 2014

[99] https://royalsociety.org/policy/projects/science-public-enterprise/digital-repositories/

[100] A curated archive refers to one where there is active management and appraisal of data over the life-cycle of scholarly and scientific materials (http://digitalcuration.blogspot.ca/2009/08/curated-databases-and-data-curation.html)

Universities are important stakeholders in the area of research data management activities. They have direct access to researchers (unlike funders) and can raise awareness of the benefits of RDM, provide support and guidance, and collect data in local repositories. A 2012 UK report argued "[u]niversities and research institutes should play a major role in supporting an open data culture by [...] developing a data strategy and their own capacity to curate their own knowledge resources and support the data needs of researchers."[101]

In 2013, the German Rectors Conference, an association of 268 universities in Germany issued a resolution[102] about research data management that urged universities to:

- Agree on guidelines about how to handle digital research data
- Collaborate beyond the boundaries of the university
- Improve information skills
- Develop institutional infrastructures for research data management

The role of the institution was also underscored in the 2011 report of Canada's Research Data Summit, which stated that institutions should:

- Maintain sustainable research data repositories
- Support the implementation and enforcement of funding agency data policies
- Provide support on campus for data management activities through employment of trained data scientists
- Implement rewards for data management and include these in promotion and tenure processes

One of the reasons that institutions have been reluctant to get involved in research data management is that they must find and justify resources to devote to this area. Most Canadian institutions that are providing RDM services have redirected some of their budget from other services, rather than obtaining new monies from other funding streams, although a few have received funding via other sources.

8. Current State of RDM in Canada

In Canada, over the past fifteen years, there have been numerous consultations and meetings discussing the state of research data management in the country and proposing various solutions. The more recent events, summarized from the comprehensive account written by Chuck Humphrey[103], include the National Data Archive Consultation in 2002, which produced a report calling for the adoption of a national data archive service to collect and preserve the research data produced in Canada.[104] In 2004, there was a National Consultation on Access to Scientific Research Data (NCASRD), which aimed to address the issues of data access in the physical and life sciences. The report[105] called for the establishment of a national steering body to help coordinate data management and preservation services. In 2008, the Research Data Strategy Working Group was launched to bring together the major stakeholder communities and develop strategies for improving the situation of RDM in Canada. They

[101] http://royalsociety.org/policy/projects/science-publicenterprise/report/
[102] From a presentation by Jochen Schirrwagen at RDA Long Tail of Research Data, Amsterdam, September 23, 2014
[103] http://preservingresearchdataincanada.net/category/introduction/
[104] http://www.sshrc-crsh.gc.ca/about-au_sujet/publications/da_phase1_e.pdf
[105] https://datalib.library.ualberta.ca/data/NCASRDReport_e.pdf

hosted the National Data Summit in 2011, which brought together over 160 senior managers. The final report published by the group, "Mapping the Data Landscape: Report of the 2011 Canadian Research Data Summit", included a set of recommendations to develop stronger community involvement in research data management and preservation. This led to the launch of Research Data Canada, an organization that is working to move forward on the recommendations in the report. And, in 2014, the Digital Infrastructure Leadership Council hosted a national meeting to discuss better coordination around Canada's ecosystem, including the management of research data.

In 2013 the TC3+ (CIHR, NSERC, SSHRC, CFI) and Genome Canada published a document that proposed "changes to their funding policy frameworks that promote excellence in data management practices, thereby advancing digital scholarship and Canada's digital infrastructure ecosystem to the benefit of Canadians". This document identified three important next steps for the members of TC3+:

1. Define the core elements of an agency-based and focused data stewardship plan
2. Work with other organizations and working groups to ensure ongoing consultation and coordination with all stakeholders, including the provinces, in the development of Canada's national digital infrastructure for research
3. Collaborate in the development of a coordinated plan to encourage the establishment of new and/or the enhancement and sustained operation of existing world-class centres specializing in data management

This was followed by a consultation with the stakeholder community. While none of these initiatives have resulted in an immediate or profound change in the research data management environment yet, they have contributed to a slow but steady increase in the visibility of research data management as an issue.

8.1 Gap Analysis

RDM policies cannot be adopted in isolation. Good research data management practices depend on multiple factors being in place – including incentives, skills and expertise, services, infrastructure, funding, and policies. These factors create a setting that supports RDM across the lifecycle and research domains. This gap analysis reviews the current situation in Canada across four axes representing the key factors that will contribute to the successful implementation of an RDM policy:

- Funding for RDM across the data lifecycle
- Infrastructure and services for RDM
- Expertise and support for the proper management of data
- A shared understanding of roles and responsibilities of the different players

Funding: Funding in Canada for data management activities is not consistently available across the lifecycle and disciplines. As discussed earlier in the document, the costs of managing research data can be significant depending on the types of data produced. Costing tools developed in Australia, UK and the US illustrate a variety of costs across the entire data management lifecycle including production, dissemination, sharing, and preservation.

Most funding for research data management in Canada is available during the lifespan of the project. For example, the costs of data management in the collection and analysis phase of the research project are generally considered eligible expenses in most grant programs, as are the costs for the development of databases and the storage of that data. However, once the project is over, data must be archived and handed over to a long-term data repository that provides access and preservation services in order to enable its further re-use.

While there are many possible models for funding RDM services and infrastructures, Canada currently has only a few select mechanisms to support data management beyond the lifespan of the project. There is funding for some large domain data centres through CFI and direct government funding, as well as the indirect costs of research, however, these do not cover all domains nor, in some cases, do they support long term access. The lack of funding models in Canada was identified as an important issue at the Digital Infrastructure Summit held in January 2014. The summit's report states, "there are few vehicles of support for the system-wide elements of RDM. Further, some small incremental funding was needed for certain aspects of the design and implementation of DI [Digital Infrastructure]. Although the latter amounts are not likely to be significant, they are not currently "line items" in the budget of any of the key organizations. This gap needs to be addressed."[106]

Infrastructure and services:

The infrastructure for research data management in Canada is piecemeal, with some fields having very good coverage and others very little. Domains that are well covered are generally those that have access to large national repositories and have established traditions of data sharing (e.g. astronomy, ocean science, Statistics Canada Data Centres, polar/arctic research data, genomics). There are also large-scale international data repositories that preserve and provide access to data in their fields (e.g. PubChem, GenBank, Protein Data Bank, Global Biodiversity Information Facility, Inter-university Consortium for Political and Social Research). The Canadian government maintains repositories that house data in many areas deemed of national importance. In terms of multidisciplinary repositories, both Figshare and Dryad, which offer more generic repository services, are available to Canadian researchers to deposit and several institutions are running repository services that collect and provide access to the data produced by researchers on their campus (University of Alberta, Simon Fraser University, Scholars Portal in Ontario, University of British Columbia, University of PEI).

RDM can be very complex and support services may be required at various stages throughout the data lifecycle, from preparation of data management plans, to documentation of data for access and preservation, to the re-use and analysis of datasets. There is some support in the context of domain repositories, as well as a growing number of libraries that are offering support to the researchers at their institution, but there remain numerous gaps in the infrastructure and services required to comprehensively support RDM across all communities.

A recent initiative, now referred to as Portage (formerly Project ARC), is developing a national library-based network for research data management in Canada to address both the infrastructure and services gap.[107] Portage is managed by the Canadian Association of Research Libraries and has two components: a national network of

[106] http://digitalleadership.ca/wp-content/uploads/2014/02/Summary-Report-of-Summit-2014-Final-March-2014.pdf
[107] http://data-carl-abrc.ca/project-arc/

expertise and a national preservation and discovery system. The network is in its early stages, but the aim is to provide support for data management planning across Canada, as well as begin to build the infrastructure that would support widespread data management, sharing and preservation. Portage is working closely with several universities, Research Data Canada and Compute Canada to develop solutions that will improve our ability to store and preserve the range of research data produced in Canada.

Skills and expertise: Research data management requires specialized skills and knowledge, both for the researchers handling the data and for the support services (such as libraries and IT staff). RDM expertise must be embedded throughout the lifecycle of the research data, from their collection, to dissemination, to preservation. In a blog post about RDM, Chuck Humphrey, Data Librarian at the University of Alberta and one of Canada's experts on research data management, states, "[d]ata management activities span the research lifecycle and involve many different skills, drawing upon a variety of expertise. The demands for data management expertise depend on the scale of the research project. A small project may involve only a couple of people, who can manage with a general set of skills. A much larger project may require a team of experts with each team member responsible for a specific specialization."[108]

There are some efforts to improve data management support for researchers. Many Canadian universities are developing services to provide data management support on campus, which range from basic information resources to more comprehensive support, such as provided by the University of Alberta, which offers expert advice, repository services, and data management planning tools.[109] Portage is an initiative being led by the Canadian Association of Research Libraries which aims to launch a distributed national network of expertise that would offer information resources, training and consulting services for Canadian institutions, libraries, and researchers. In addition, CARL has offered a course for librarians on RDM services and is organizing another course about research data management planning in 2015.

Despite these efforts, there is still much work to be done to improve awareness and expertise for RDM in Canada. The 2011 Gap Analysis observed "[r]esearchers rarely have the skills to appropriately manage their data and there are few data professionals to assist them."[110] In a report published in January 2013, the Research Data Canada Education and Training Committee went on to say, "[c]urrently within Canada there are few training opportunities available in the area of RDM. Canada lacks the national-level coordination of a body like Jisc in the UK, and federal granting councils do not yet provide the necessary policy incentives regarding RDM."[111] In their report, the RDC Committee made a number of recommendations about how Canada can improve the current situation, including, "Canada needs a multipronged strategy for the delivery of education in order to build capacity for research data stewardship in Canada. This should include integrating RDM in graduate curricula for future researchers, implementing RDM courses in information schools and other relevant academic programs, and providing a variety of training that will assist current researchers, librarians and other stakeholders to up-skill for RDM."[112]

[108] http://preservingresearchdataincanada.net/2012/12/17/research-data-management-infrastructure-iii/
[109] http://www.library.ualberta.ca/researchdata/
[110] http://www.carl-abrc.ca/uploads/pdfs/data_brochure-e.pdf
[111] Summit Report
[112] Ibid

Roles and responsibilities: Data management is rarely the sole responsibility of the principle investigator. A number of different stakeholders are involved in the research process and have a role to play in ensuring good practices. Other stakeholders include institutional leaders; co-investigators and graduate students; external contractors involved in data collection; research administrators; institutional IT services; and institutional or external data repositories.

Table 6 is adapted from a list of responsibilities published in 2007 by UKOLN[113]. It outlines the responsibilities of the key players involved with research data management and identifies key obstacles for taking on responsibilities.

Table 7: Roles and responsibilities in RDM

Role	Responsibilities	Obstacles
Researcher	Manage data for the life of the project Meet standards for good practice Comply with funder and institutional data policies Work up data for use by others	Low awareness of appropriate standards in some fields Low knowledge about good data management practices in some fields Lack of funding for support by data scientists Lack of time to properly document data
Universities	Adopt data management policies Raise awareness of funder requirements Ensure standards of good practice are met Provide training and advice to researchers Manage a repository service for long term access and preservation of data	Only gradual uptake of RDM policies and services at institutions (including data repositories) Training opportunities for RDM support is not widespread
Data centre/ repository	Manage data for the long-term Meet standards for good practice Provide training for deposit Promote the repository service Protect rights of data contributors Provide tools for re-use of data	Lack of sustainable funding for data centres in many fields
Funder	Adopt data management policies Monitor and enforce data policies Fund data management activities as part of the project Resource post-project long-term data	Unwillingness to divert funds from research to data management

[113] http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf

	management Support workforce capacity development of data curators	

Currently, the roles and responsibilities outlined in the table are aspirational. There is no common understanding across stakeholders about where the responsibilities lie for the various aspects of RDM. Both researchers and institutions are apprehensive about taking on greater responsibility for managing research data. Researchers are worried about the time and resources required for preparing data, as well as a pervasive lack of expertise within the research community. Institutions are concerned about how they will fund data management support services and repositories.

In terms of coordination across stakeholder communities, Research Data Canada has been active in bringing together different stakeholders to move towards a common understanding of roles and to develop collaborative solutions.

8.2 Readiness for Policy Implementation

In the previous section of the report, we discuss the current state of RDM across a number of indicators. That analysis informs the assessment of national readiness for RDM policy implementation in Canada. There are a number of ways in which readiness can be considered. For a detailed assessment, one could use the Community Capability Model Framework. This framework identifies eight capabilities by which capacity for RDM can be measured: Openness, Legal, Ethical and Commercial Considerations, Collaboration, Economic and Business, Skills and Training, Common Practices, Research Culture and Technical Infrastructure.[114] However, this kind of assessment involves a very detailed, intensive process that goes beyond the scope of the work for this brief. The following table provides a preliminary assessment of the community's readiness according to key policy elements. It should be noted that given the large variations in disciplinary practices and infrastructure, any assessment will have limitations and offers a somewhat general view about readiness to adhere to policy requirements.

[114] http://ozk.unizd.hr/proceedings/index.php/lida/article/viewFile/121/123

Table 8: Readiness assessment scores

Policy element	Readiness assessment
Data quality and standards	Standards for the collection of research data vary significantly across the disciplines. Some fields already have long established standards while other fields are still developing best practices. International surveys have found that many researchers do not feel they have the expertise and knowledge to appropriately manage their data. Therefore, support and training for RDM in the research community will be required. Some Canadian universities already provide services through their libraries to provide guidance for researchers in managing their data. For example, a review undertaken by the Research Data Canada Standards and Interoperability Committee found that at least 21 universities are providing some resources about data management planning.[115] Portage is developing a model for a network of expertise that will build on and expand the scope of existing university services and offer information resources as well as in-depth support for researchers across the country from experts in the library community. In some cases, applying appropriate standards and quality control may require extra funding for data management activities. These activities are not always eligible expenses in the grant. In general, researchers can be expected to identify and use standards and best practices for managing research data in their field if support is for this is available.
Data access and sharing	There are significant gaps in the repository infrastructure in Canada. These gaps are being addressed through Portage and other repository initiatives, but it will take time and funding to build comprehensive repository network across the country. Another challenge is overcoming the reluctance of some researchers to share their data. There are a number of issues that contribute to this situation, including lack of policies and incentives, the time and effort it takes to prepare the data, and the fear of being scooped. The principles of data sharing are becoming more widely accepted due to global recognition of the value of access and re-use. These principles need to be further embedded into research culture, through awareness raising, adoption of policies and incentive. Researchers can minimally adhere to data access and sharing requirements by retaining their data, describing them appropriately and sharing them with others when requested. However, for widespread data sharing to occur, researchers need to be able to deposit their data into a repository where it will be maintained, curated

[115] Private communication with Research Data Canada (October 10, 2014)

	and can be discovered by others.
Data retention and preservation	In terms of full-scale preservation, these services are not yet widely available in Canada. There are some well-established government and domain repositories with preservation capacity. Scholars Portal in Ontario, the University of Alberta, and the National Research Council also maintain trusted digital data repositories. As with access and sharing, researchers can provisionally adhere to data retention and preservation requirements by ensuring their data are stored and backed up appropriately.
Data management plans	Data management plans oblige researchers to describe how they will manage their data during the course of the research project and outline their plans for sharing and preserving data once the project is completed. The main challenge of requiring data management plans is acceptance by the research community. Few researchers have an understanding of what a good data management plan entails. Researchers will need support for filling out data management plans. The University of Alberta currently provides access to an automated tool, called DMP Builder, which assists researchers in developing data management plans. This tool is available to everyone and there will soon be a Portage version to offer support in both French and English. In the context of this tool, guidance is also being developed to help researchers understand and respond to requirements. More detailed expertise will also be available at some individual institutions and through the Portage network of expertise. National capacity in this area will only develop over time, but raising awareness of the benefits of DMPs will be important to ensure their acceptance in the research community. Canadian researchers can be expected to develop DMPs and their implementation will help build an understanding of data management planning in the research community.

9. Conclusion

The global trend towards research data management and sharing is being driven by a number of things: to ensure verification of research results; to improve the quality and efficiency of research; and to promote the re-use of research for new discoveries and innovation.

This review found that, although there are many gaps and barriers, the environment for policy adoption for RDM in Canada is improving and Canada has made significant progress since the OECD declaration in 2004. There have been both bottom-up and top-down efforts to advance RDM infrastructure and expertise in Canada. Further targeted government investment and incentives could accelerate these advances.

It is clear that policies cannot be adopted in isolation. Good research data management practices depend on multiple factors being in place – including incentives, skills and expertise, services, infrastructure, funding, and policies. In addition, comprehensive adoption of good data management practices involves significant cultural change across numerous stakeholder groups. This will take time, and will likely progress through steady, incremental steps across multiple factors and domains. Parallel efforts must also be made to increase awareness and acceptance of policy objectives within the research community.

Despite the challenges, it is clear that policies are an extremely powerful lever to push the community forward. They provide a framework that helps to guide best practices and without them it is unlikely that there will be widespread adoption of RDM in Canada. Countries that have chosen to move ahead with policy implementation have found that although full compliance cannot be expected immediately, policies can greatly assist in raising awareness of RDM. As noted in a 2013 TC3+ consultation document, “Canada now stands in direct competition with a host of other countries, including the United States, European Union countries, Australia and other technologically advanced countries, in the race to develop an effective strategy for harnessing the digital wave.”[116] RDM policies are an important component of any such strategy.

[116] http://www.sshrc-crsh.gc.ca/about-au_sujet/publications/digital_scholarship_consultation_e.pdf