Industry Canada's Consultation on Developing a Digital Research Infrastructure Strategy

Response submitted by the Portage Network of the Canadian Association of Research Libraries (CARL).

September 11, 2015

Canadian Association of Research Libraries

Association des bibliothèques de recherche du Canada

Who we are

The Canadian Association of Research Libraries (CARL) represents the country's twenty-nine largest university libraries. Enhancing research and higher education are at the heart of our mission. CARL promotes effective and sustainable scholarly communication, and public policy that enables broad access to scholarly information.

Introduction

The Portage Network of the Canadian Association of Research Libraries (CARL) welcomes this opportunity to respond to Industry Canada's Consultation on Developing a Digital Research Infrastructure Strategy. This Network supports a national research data management (RDM) service to assist researchers and other RDM stakeholders through (1) a library-based network of expertise on RDM and (2) national platforms for planning, preserving, and discovering research data.

CARL and Portage strongly support Canada's Science, Technology and Innovation Strategy 2014, which 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"[1]. We concur with the vision that open science, including improved research data management, will maximize the public's investment in research and lead to greater utility and impact of research.

The Canadian research library community has a long history of collaborating on cross-sectoral infrastructure, providing others with ready access to information resources. During the print era, libraries and the federal government worked cooperatively to support local access to federal publications through the Depository Services Program. More recently, the Canadian Research Knowledge Network has provided affordable access to digital full-text, searchable databases for post-secondary institutions through a national licence program with commercial publishers. Furthermore, regional library organizations provide mechanisms for sharing resources among academic libraries. This same cooperative spirit of working across sectors, with other stakeholders, and among themselves is now being applied to research data management.

---

[1] Industry Canada. (2014). *Seizing Canada's Moment: Moving Forward in Science, Technology and Innovation 2014* (Catalogue No. Iu37-4/1-2014E-PDF). Retrieved from Industry Canada: https://www.ic.gc.ca/eic/site/icgc.nsf/eng/h_07472.html

Direction and Strengths

CARL agrees with the Consultation's Background statement that "high-performance computing, storage, high-speed networks, and other tools and resources, including software, standards and data management services" are all vital elements needed "to perform data-intensive and computationally-intensive research and data management." We further recognize that these key elements of Digital Research Infrastructure must operate dependably and, as much as possible, transparently together. A national strategy for Digital Research Infrastructure is important for ensuring that Canada has a coordinated approach in this regard to remain a world leader in research and innovation.

We see the transformation, strengthening, and support of digital research infrastructure in Canada to be highly dependent on rewarding and incentivizing collaborative initiatives among the diverse stakeholders in research data management. While the task of RDM is too great for any single stakeholder, Canada can achieve world-class digital research infrastructure through organizations working collaboratively on agreed upon goals. This ecosystem is not a case of winner takes all but rather is one of high interdependency among stakeholders where collaboration is needed for all to thrive. This will require stakeholders sharing skill sets to fulfill the capacity requirements of building and maintaining this infrastructure. In addition, the sustainability of this infrastructure will require funding that is directed at horizontal investments rather than today's funding formulas, which tend to favour vertical, piecemeal, short-term projects.

The effectiveness of digital research infrastructure will be largely determined by a comprehensive set of data policies and by the clarity of the mandates stipulated around the implementation of this infrastructure. Policy guidance and role clarity are necessary in an ecosystem of diverse stakeholders, providing clear boundaries in which to operate and helping to build working relationships among its stakeholders.

The overall success of this approach will be measured by the degree to which DRI access has been simplified for use by researchers and is equitably available throughout the research community.

A Change in Cultures

Canada is moving from a research data culture of neglect to one of long-term stewardship. It is not the case that every aspect of research data management was overlooked in the past. Instead, the management of research data has not been consistently applied across the full lifecycle of research. New expectations around data sharing, data interoperability, data preservation, and data reuse are placing greater

demands on stages in the research lifecycle that previously were largely ignored. As noted in the Comprehensive Brief on Data Management Policies, "Canada currently has only a few select mechanisms to support data management beyond the lifespan of the project"[2]. To be successful, Canada's Digital Research Infrastructure Strategy must address all stages of the research lifecycle, including data management planning, sharing and long term preservation of research outputs.

Preservation services and infrastructure are particularly crucial to ensure that research data are available for sharing and reuse both in the present and the future. Digital preservation is a complex process involving planning, resource allocation, specialized skills, and the application of community-recognized best practices. Furthermore, this must all take place in the context of a stable and persistent institutional environment. Today, Canada simply does not have an appropriate number of publicly curated data repositories to meet the demands for preserving research data. Greater public investment in long-term preservation services and infrastructure will be necessary to support the government's goals of open science.

Some have argued that research data management is an issue that should be addressed only by individual domains or disciplines of study. While domains have significant roles to play in defining and promoting data standards and practices among their own research communities, there is also infrastructure that is common across domains that needs to be co-developed and shared. Without the sharing of infrastructure, the most likely outcome will be an uneven playing field across domains, creating areas of research with and without adequate research digital infrastructure. Where this maldistribution will be particularly evident will be in the early and latter stages of the research lifecycle, i.e., when plans for data management are being made and when infrastructure is required to share, preserve, and reuse research data. Domain engagement in DRI is necessary but not sufficient to build national infrastructure. This fact too needs to be part of the changing culture of research in Canada, especially if we are to reach greater heights of innovation and knowledge generation.

The Portage Network is looking to provide services in these early and latter stages of the research lifecycle. Its Data Management Planning service will help researchers of all domains prepare plans before starting a project and its preservation platform will help researchers share and archive their data.

A strategic, collaborative approach to developing Canada's research infrastructure will ensure long-term sustainability and accessibility of the system, and thus support the broader aim of the Government of Canada "to maximize access to federally funded

---

[2] Shearer, K. (2015, April 7). *Comprehensive Brief on Research Data Management Policies*. Retrieved from Science.gc.ca: http://www.science.gc.ca/default.asp?lang=En&n=1E116DB8-1

scientific research to encourage greater collaboration and engagement with the scientific community, the private sector, and the public"[3]. Internationally, Australia, United Kingdom, Netherlands, and the European Union are investing in such horizontal infrastructures to support leading-edge research in those regions.[4]

The private sector has an important role to contribute in this environment, also. In particular, they have a role in adding value to open data and in building products around open access to data. The most harmful action that could happen would be to allow any sector to privatize access to data, which would effectively deprive access for large segments of the research community and thereby stifle further innovation and limit downstream impacts of research.

Meeting All Levels of DRI Use

The vision of Portage is "to lay the foundation for a library-based research data management network that will improve our national capacity for the management, preservation, and reuse of research data." This will require Portage to work at three specific levels: the project or program level, the institutional level, and the network level. A national DRI Strategy must recognize each of these three levels in identifying the development requirements for research infrastructure. The provision and maintenance of DRI happens mostly at the institutional and network level, where the latter may be organized regionally, nationally, or internationally. The actual use of DRI, however, typically occurs at the project or program level, which may be distributed locally, regionally, nationally, or internationally. This complexity raises cross-cutting jurisdictional issues that are political, financial, and legal. Consequently, a national DRI Strategy needs to include mechanisms for dealing with jurisdictional challenges in sharing research infrastructure.

The current state of research data curation at the institutional level is varied both within and across institutions. This creates a challenge both for a national DRI Strategy and for Portage, for whom one of its guiding principles is inclusiveness: an aim to serve all researchers and to create a more level playing field. This means working closely at the institutional level to determine how best to address each institution's current capacity for research data curation and to find affordable solutions for today while also building greater capacity for the future. Institutions can begin by establishing a baseline commitment to research data management through policies that define local mandates on data stewardship. This will help determine what is needed from national research

---

[3] Open Government. (2014). *Canada's Action Plan on Open Government 2014-16* (Catalogue No. BT22-130/2014E-PDF). Retrieved from Open Government Portal: http://open.canada.ca/en/content/canadas-action-plan-open-government-2014-16

[4] Shearer, K. (2015, April 7). *Op. cit.*

infrastructure services within and between institutions. Knowing with whom to collaborate at the institutional level is vital to move constructively forward. This should also motivate national infrastructure providers similarly to coordinate their support of institutions and researchers better.

Moving Forward

Canada has important components upon which to build a national DRI Strategy. CANARIE and Compute Canada are vital to the provision of basic aspects of this infrastructure. CARL's strong commitment to research data management is a significant contribution from the research library community. Research Data Canada has an important role to play as a convener of stakeholders, helping to identify common ground for cooperation. A growing number of data repositories across Canada and the major data producers across sectors are also important players in the research infrastructure ecosystem. CASRAI, a standards setting organization, is working to advance the interoperability of research administrative information and to make the business operations of research more efficient. All of these key stakeholders have a contributing role to play in ultimately shaping Canada's DRI Strategy.

Over the past year, CARL, Compute Canada, CANARIE and Research Data Canada have conducted pilot projects using existing technology to explore operational models for preserving, discovering, and accessing research data at the national level across research domains. These experiences have resulted in new, positive working relationships among organizations and have provided important lessons about assembling production scale platforms and services. For example, CARL and Compute Canada are currently preparing a plan based on their pilot project outcomes for a platform service to produce archival-quality research data objects that will be securely stored, discoverable, and usable well into the future.

Forging a strategy that capitalizes on the willingness to work together is the next step toward a DRI Strategy.

In Summary

1. Through the formation of the Portage Network, **CARL is actively engaged as a collaborator in Canada's research data management ecosystem**.
2. The transformation, strengthening, and support of digital research infrastructure in Canada is highly dependent on **rewarding and incentivizing collaborative initiatives** among the diverse stakeholders in research data management.

3. The sustainability of this infrastructure will require **funding directed at horizontal investments** rather than today's funding formulas, which tend to favour vertical, piecemeal, short-term projects.

4. The effectiveness of digital research infrastructure will be largely determined by a **comprehensive set of data policies** and by the **clarity of the mandates** stipulated around the implementation of this infrastructure.

5. Canada's Digital Research Infrastructure Strategy must address **all stages of the research lifecycle, including data management planning, sharing and long term preservation of research outputs**.

6. Digital preservation services and infrastructure are vital to the ongoing access and reuse of research data. **Greater public investment in long-term preservation services and infrastructure will be necessary to support the government's goals of open science.**

7. While domains have significant roles to play in defining and promoting data standards and practices among their own research communities, **infrastructure common across domains needs to be co-developed and shared**.

8. The private sector has an important role to contribute in this environment, also. In particular, they have a role in adding value to open data and in building products around open access to data. **The most harmful action that could happen would be to allow any sector to privatize access to data**, which would effectively deprive access to data by large segments of the research community and stifle further innovation and downstream impacts of research.

9. A national DRI Strategy must **recognize three levels of use – the project or program level, the institutional level, and the network level –** in identifying the development requirements in research infrastructure.

10. Forging a strategy that **capitalizes on the willingness of key stakeholders to work together** is the next step toward a DRI Strategy.

Canadian Association of Research Libraries

President

Martha Whitehead

Vice Provost and University Librarian,

Queen's University

Director, Portage

Chuck Humphrey

chuck.humphrey@ualberta.ca

CARL Executive Director

Susan Haigh

susan.haigh@carl-abrc.ca

Tel. 613.482.9344   203-309 Cooper St.

info@carl-abrc.ca   Ottawa, ON

www.carl-abrc.ca   K2P 0G5

CARL ABRC