Future-Proofing the Canadian DRI/RDM Ecosystem through Openness and Collaboration

White Paper delivered by the Canadian Association of Research Libraries (CARL)
to NDRIO as part of its Needs Assessment Consultation - *December 2020*

Introduction

The future of our academic libraries is completely different than our past. No longer simply a repository of published research and primary sources, today's academic libraries are vibrant, active partners in the full lifecycle of the research enterprise, from creation, to dissemination, to preservation. As such, they have significantly invested in, and contributed expertise in recent years to, Canada's digital research infrastructure (DRI), and have effected a remarkable growth in research data management (RDM, or DM) capacity through their proactive launch and funding of the Portage initiative. The academic library community is committed to continually meeting evolving scholarly needs, and collaborating across regional and national networks to implement appropriately scaled, sustainable, bilingual delivery models for essential research services. We are pleased to offer thoughts about future areas of need within the DRI ecosystem.

Key Areas of Ongoing and Growing Need

Highly Qualified Personnel (HQP)

Highly qualified personnel (HQP) within academic libraries have specialized skills and expertise essential to supporting researchers in the key DRI areas of Research Software (RS), Advanced Research Computing (ARC), and Data Management, but with a primary focus on DM. Librarians and library-based functional specialists bring multidisciplinary expertise in the areas of GIS, data management planning, data curation, data stewardship, digital humanities, preservation, copyright, intellectual property, publishing, ethics, training, standards, information management (IM), and IT. They provide essential support to researchers throughout the entire data lifecycle and are often an integral part of successful research projects. While alternative supports are sometimes available through regional/national, commercial and disciplinary providers, "researchers are most comfortable adopting new tools and approaches when the infrastructure and services are close to them"[1] Academic libraries have invested significantly[2] in providing these local supports. Librarians are

---

[1] Baker, David, Bourne-Tyson, Donna, Gerlitz, Laura, Haigh, Susan, Khair, Shahira, Leggott, Mark, ... Martha Whitehead. (2019, December 13). *Research Data Management in Canada: A Backgrounder*. Zenodo. http://doi.org/10.5281/zenodo.3574685

[2] For example, a recent CARL paper reports "the total annual staff time devoted to open [scholarship] activities [including open research data] in all 28 libraries is 194.08 FTE, representing spending of about $17 million CAD" (Shearer, 2020).

also highly networked on a pan-Canadian basis — as evidenced by their active participation in all Portage Expert Groups.

What is needed going forward
There are signs – for example, the rising profile of "FAIR" and "CARE" data practices — that researchers are increasingly aware of, adopting, and confident in RDM best practices, recognizing them as foundational to successful research projects. To build on this momentum, sustained national investment in researcher training and in bilingual RDM tools, guidance and services that enable best practices is required. Investment in training and professional development for HQP based in libraries will help foster that all-important local layer of data specialist expertise, and will allow academic libraries to continue to provide timely, informed and cost-effective support to researchers, leveraging their knowledge of national platforms and services at the institutional level.

Open Scholarship
The COVID-19 pandemic illustrates that immediate open access is imperative to solve major global challenges. Canada must learn from this experience and prioritize open science and open access to scholarship and associated research data in its national policies, infrastructure, and services.

CARL is committed to achieving a scholarly communication model where open is the default for the vast majority of research outputs, but which allows for alternate approaches that protect the privacy of an individual's personal data and respect the right of Indigenous communities to maintain control of their data and traditional forms of knowledge. Managing sensitive data ethically is vitally important, and some progress has been made in this area through Portage, but the challenge is by no means met.

CARL institutions are committed to the use of infrastructure platforms that support open scholarship and have made a financial investment in the development and maintenance of such platforms to ensure their availability, sustainability, and interoperability over time.[3]

What is needed going forward
NDRIO must ensure that DRI in Canada is fully aligned with the principles of open science. Accordingly, NDRIO must make investments that prioritize and ensure the

---

[3] In its 2020 report *Investments in Open: Canadian Research Libraries' Expenditures on Services, Staff, and Infrastructures in Support of Open Scholarship*, CARL reported that 28 participating university libraries spent a total of $23 million CAD in 2018-2019, with an average spend per institution of $827,086 CAD. This includes important collective investments in Coalition Publica and Portage, as well as individual institutional contributions across a variety of external services and local developments.

sustainability of infrastructure, tools and systems that support open as a default but enable sensitive data to be subject to appropriate restriction.

Interoperable, Scalable Platforms

Libraries have invested substantially in research infrastructure and have built expertise, promoted standards, and developed bilingual tools supporting data deposit, scholarly metadata, permanent identifiers (PIDs), interoperability, and data management planning.

The benefits achieved through scalable infrastructure are tangible in library-supported data repository platforms that support managing, publishing and preserving research data such as [Dataverse Canada](#) and the [Federated Research Data Repository](#). The shared infrastructure, policy, resources and technical support afforded by these regional and national efforts benefit many institutions, improve sustainability at the local level, and strengthen equitable access by researchers across the entire Canadian research ecosystem.

But libraries also manage institutional repositories (IRs), which typically contain a publicly searchable record of an institution's research output and play an essential role in institutional memory. Traditionally, Canadian IRs have focused primarily on collecting peer-reviewed articles and electronic theses and dissertations. With growth in other types of content repositories within institutions, across regions, and across research communities — notably, data repositories and open educational resource repositories — we anticipate that distinctions between types of repositories will diminish. Ongoing work is required to increase interoperability among repositories of different types to ensure that, from a researcher's perspective, discovery of and access to open scholarship is as seamless, stable and sustained as possible.

Libraries are increasingly considering how best to scale infrastructure at the national level, recognizing that distributed models of support will benefit smaller institutions that may otherwise struggle to maintain their own installations. This approach reduces duplication of effort by removing the need for individual institutions to develop and maintain local infrastructure, workflows, standards, software patches and updates. Shared infrastructure allows HQP to create shared policies, resources and best practices, and to promote knowledge exchange that will enhance the level of support provided to researchers at local institutions.

When information about Canadian scholarly outputs is not siloed, discovery and access are improved both nationally and internationally. Aggregated discovery is already advanced through the [FRDR discovery layer](#) (which harvests metadata from

Canadian repositories) and its integration into global aggregations such as the OpenAIRE gateway. This work must be sustained and its coverage strengthened.

Similarly, work has advanced to implement and leverage persistent identifiers (PIDs). With the academic library community spearheading consortial investment in DataCite Canada and ORCiD-CA, uptake and systems' use of these standards is growing. Efficiencies will be realized: for example, the national data management planning tool, DMP Assistant, will use PIDs as a means to support machine actionability and integration with research information management (RIM) systems.

What is needed going forward
It is important that NDRIO consider the full spectrum of repositories and platforms devoted to research outputs and information, the need for active repository and preservation storage solutions, and the need to support libraries as partners in developing these services at scale to provide sustainability, interoperability, accessibility, and efficiency in the open science landscape in Canada.

NDRIO must continue to foster the implementation of standards that allow for seamless exchange of information across the various tools that collect and preserve researcher information (e.g., IRs, RIMs, the Canadian Common CV or CCV). Better integration between systems and institutions will allow researcher information to be entered once and reused many times, reducing administrative burden for researchers.

The Multiplier Effect of Networked Collaborative Effort
One of the key strengths and value propositions for library involvement in DRI is the strong sense of service and professional practice in this community. The success of Portage in effecting change in the RDM landscape is due largely to a dedicated Network of Experts and broader community of practice, nurtured initially by CARL, and now by NDRIO, through the national Portage Secretariat. The collaborative and cooperative efforts of over 140 experts from over 60 different organizations across Canada speak to the interest of this community in advancing RDM best practices and bringing about underlying culture change. While library HQP form the majority of the Portage Network of Experts, involvement of researchers and other stakeholder groups including IT, ethics, funding agencies, scholarly and professional associations, government, Indigenous organizations, ARC, and colleges has been essential to ensuring a broad and inclusive vision for RDM in Canada.

As noted above, the connections and consortia libraries have forged provincially, regionally, and nationally have contributed to strong, synergistic relationships that have given rise to practical, researcher-facing RDM platforms, tools, services and training. National coordination and facilitation by Portage and Research Data

Canada have made a difference in this space. It is important to note that national training efforts have targeted both researchers and data specialists (using a 'train-the-trainer' approach).

What is needed going forward
NDRIO must address three main 'community building' needs going forward. First, there must be continued national coordination of the well-established Network of Experts to harness the invaluable energy and knowledge that is contributed in-kind by members of this data specialist community. Second, NDRIO should consider establishing an RDM-specific advisory committee composed of stakeholders from across the DRI ecosystem. Third, NDRIO should fund initiatives that promote collaboration and community building. Examples of this include a repository certification pilot and efforts to establish a national network of data curation support in Canada. Continued focus on expanded training, for both researchers and those who support them, is key as well — with strong collaboration between DM, ARC, and RS.

Conclusion

Through Portage and its affiliated platforms, academic libraries make important contributions and investments to DRI. Libraries have invested and will continue to contribute financially to common DRI infrastructure that advances local researcher needs, as demonstrated by coordinated and consortial investment in Portage, ORCiD-CA, DataCite Canada, and Dataverse Canada, as well as substantial investment in salaries for HQP supporting RDM. An appropriate, rationalized cost-sharing approach for the continued development of these and other infrastructures needs to be developed by NDRIO in consultation with universities and their libraries.

We see a future where Canadian academic libraries, and the enduring post-secondary institutions we serve, continue to play a strong, collaborative role in the development of a world-leading DRI infrastructure for Canada. Optimal continued development of the ecosystem needs strategic national investment in training and professional HQP, in interoperable and scalable platforms advancing open science, and in shared standards, tools and practices that allow seamless discovery and exchange of information and data.

Susan Haigh, Executive Director
Canadian Association of Research Libraries