

Recommendations for a National Dataverse Service

Prepared by the Dataverse North Working Group's Business Models Subgroup

Eugene Barsky (University of British Columbia)
Corey Davis (COPPUL)
Alan Darnell (Scholars Portal)
Jason Flynn (Dalhousie University)
Lisa Goddard (University of Victoria)
Meghan Goodchild (Queen's University/Scholars Portal)
Amber Leahey (Scholars Portal)
Erin MacPherson (Dalhousie University)
Pierre Roberge (UQAM)
Brienne Selman (University of Winnipeg)
Lee Wilson (Portage/ACENET)

Background

The Portage Network's Dataverse North Working Group¹ is developing a community of practice for Dataverse in Canada. As part of its work, the group has been looking at opportunities that could be addressed by nationally coordinated strategies, including hosting services provided by regional Dataverse providers. Specifically, the Working Group was tasked to:

- Coordinate and develop a framework for Dataverse hosting and support services for designated libraries or other special interests that do not currently have a place to deposit research data.
- Explore a common business model to level the access to universities across Canada.²

Over the past six months, the Business Models Subgroup has undertaken an evaluation of the current Dataverse landscape, including information gathering and an assessment of hosted Dataverse services and library users across Canada. The Subgroup completed two surveys in the Fall of 2017 of both institutional Dataverse users and Dataverse hosting providers in Canada.

¹ Portage's Dataverse North Working Group - <https://portagenetwork.ca/wp-content/uploads/2017/09/DataverseNorthWG.pdf>

² <https://portagenetwork.ca/wp-content/uploads/2017/09/DataverseNorthWG.pdf>

Using the results from the surveys and a subsequent environmental scan of existing repository service models, the Subgroup developed several potential models for how the Dataverse community could proceed. In evaluating these models, it became clear that a single national Dataverse service hosted by an experienced service provider would work best in the Canadian context, where funding is limited and expertise is widely dispersed. While the Subgroup members recognize that any national effort should adequately recognize and support those institutions and regions that choose to operate their own repository infrastructure, we believe that significant benefits would accrue to the Canadian research community through a unified national service.³

Recommendation

Portage, through its Dataverse North Working Group, should work with key stakeholders⁴ to establish a national Dataverse North service, based on a sustainable business model⁵ and hosted by Scholars Portal at the University of Toronto Libraries, that will enable all Canadian researchers and the entire academic library community to effectively utilize a robust, scalable, affordable, and open research data repository platform that aligns with the suite of national Research Data Management (RDM) Services currently under development by Portage through its Networks of Expertise. This service would engage with other Dataverse providers in Canada to support and strengthen regional and institutional efforts through the Dataverse North Working Group.

Key benefits of a national Dataverse North service

While the centralization of repository infrastructure at a national level does not necessarily make sense in every domain, this group believes that a national service will significantly benefit the Canadian research community in a number of important ways:

- *Realizing true economies of scale.* While recognizing that some institutions or regions will choose to operate their own research data repositories for a variety of sound reasons, it's also important to understand that the greatest threat to digital materials over the long term is often economic.⁶ The economies of scale realized through a

³ This is of particular moment, as the Tri-Agency Digital Data Management Mandate approaches in 2018.

⁴ University of Toronto Libraries, Scholars Portal, OCUL, the other regional academic library consortia (COPPUL, BCI, and CAUL), and existing Canadian Dataverse providers (e.g. University of British Columbia, University of Alberta, University of Manitoba, Dalhousie, University of New Brunswick, and others)

⁵ This will likely be a mix of structural funding provided through Portage from the federal government, from the academic library community through annual service fees, and from project or grant funding in areas such as feature development and other enhancements.

⁶ <http://blog.dshr.org/2013/10/the-major-threat-is-economic.html>

national service could provide the kind of cost optimizations needed to effectively and sustainably manage digital assets over time: “collaboration and federation can help to manage, share, and reduce costs.”⁷ A national service will have a significant impact on the ability of many small and medium-sized institutions to develop institutional RDM repository services.

- *Commitment to responsible data management.* A national Dataverse instance demonstrates a commitment to data management, including the Tri-Agency mandates. The Tri-Agency Statement of Principles on Digital Data Management outlines the roles and responsibilities of researchers, institutions, research communities and funders. A national Dataverse instance will assist in meeting some of these roles and responsibilities by providing a place to securely deposit, promote and share data.
- *Provides equal access to data services for all institutions.* A national Dataverse will provide access to RDM platforms and services to small and medium sized institutions that may be under resourced in this area.
- *Help researchers and journal publishers navigate a complex environment.* Researchers looking to increase visibility and discoverability of their data, and to fulfill deposit mandates, and journals looking to manage the submission, review, and publication of data associated with published articles, must navigate an increasingly complex environment associated with RDM. A well-publicized national service will help declutter this environment by providing a well-articulated suite of services available to all researchers and university publishers in Canada.
- *Strengthening technical staffing and other support services.* Based on our recent survey of Canadian Dataverse users, local support is limited and most institutions report that librarians and others undertake Dataverse-related activities as a part of broader responsibilities. A national service could provide centralized technical support, and at the same time encourage the collective creation of materials to support, for example, advocacy and awareness, outreach activities, the creation of metadata templates, and discipline specific guidance, with everyone benefiting from working with the same version of software system.
- *Leveraging existing expertise and infrastructure.* Scholars Portal runs one of the largest Dataverses in Canada, in addition to other established technical library services on infrastructure at the University of Toronto Libraries. UTL/Scholars Portal Data Centre is a secure environment that conforms to industry best practices for maintaining data integrity and longevity, and UTL/Scholars Portal staff are able to upgrade software on a regular basis to enable the latest features and fixes, and to address the latest security vulnerabilities.

⁷ OECD (2017), "Business models for sustainable research data repositories", OECD Science, Technology and Industry Policy Papers, No. 47, OECD Publishing, Paris. Available at <http://dx.doi.org/10.1787/302b12bb-en>

- *Pooling resources for new feature development.* Canadian Dataverse users have indicated a number of new features they would like to see, including better visualization tools, data curation support, file organization, media streaming, digital preservation, and support for large files, to name a few. New feature development would more efficiently take place on a single platform where resources are pooled nationally; and, where applicable, the code-base for enhancements and new features could be made openly available via Github (or a similar service) so that other Canadian Dataverse providers might take advantage of these developments.
- *Improving funding opportunities.* A national service with broad participation across the country would also be a more attractive recipient for grant funding and research dollars, such as that made available through CANARIE⁸ and other CFI and even Tri-Agency funding sources. Developments to a national Dataverse North system would also improve outcomes for future integration with other RDM systems and tools in support of RDM workflows.
- *Supporting local and regional efforts.* Some institutions and regions might have compelling reasons to run their own Dataverse or other research data repositories (local expertise and capacity with different platforms, privacy impacts and related local or provincial policies, policies and practices in support of local or regional culture, etc.). A national service would work constructively with these institutions or regions running their own Dataverses and help them build capacity by invigorating the Canadian research data management community of practice for the benefit of all. We are seeking to build and foster relationships across Canada, while working in good-faith to develop a shared national service where it may be beneficial to all.
- *Alignment with national RDM efforts.* A national Dataverse instance will be better aligned with, and poised to take advantage of, other national RDM services currently in development. For example, a single instance of Dataverse can more easily work with Preservation Service Providers to meet the preservation processing specifications for research data outlined by the Preservation Expert Group (PEG),⁹ receive curation staffing support from the regionally distributed, nationally coordinated (and technology agnostic) curation network being developed by the Portage Curation Expert Group (CEG).
- *The ability to influence international efforts.* A national Dataverse instance will be able to influence international efforts that intersect with research data repository developments, such as the work being done under the auspices of the Digital Curation Centre (DCC) and the Research Data Alliance (RDA).

⁸ CANARIE software funding call for 2018 - <https://www.canarie.ca/software/funding/>

⁹ Research Data Preservation in Canada - <https://open.library.ubc.ca/cIRcle/collections/ubccommunityandpartnerspublicati/52387/items/1.0371946>

Guiding principles for establishing and operating a national service

Research data repositories are an increasingly important component of the digital research infrastructure and open science landscape (contributing to its economic and social benefits). Moreover, research policy makers and funders increasingly mandate open data for publicly funded research. A national Dataverse North service will be guided by principles that will ensure ongoing, sustainable, accountable, and responsive operations.

- *Developing a sustainable and equitable business model.* A sustainable business model will take into account how cost drivers (e.g. data volume, frequency of deposit, mix of users, levels of curation) and available revenue sources (e.g. structural funding from the federal government, service fees from libraries, support from participating institution, and grant funding) will adequately scale to meet future demand. Stakeholder identification and engagement will be key to articulating this model and demonstrating value over time. At the same time, in order to improve equity across Canada, we anticipate there being significant in-kind support offered to offset traditional cost models which may be prohibitive to some institutions and regions in Canada. Different cost models are being explored and we are anticipating a mix of funding from institutions, regions, and government.
- *Cultivating a community of practice.* Repository services will not be effectively utilized without encouraging a community of practice to support capacity building across organizations. A national service will support the Dataverse North Working Group in its efforts to to develop a community of practice for libraries using or interested in using the Dataverse repository platform for research data in Canada.
- *Community-led, community-owned.* A national Dataverse North service will be offered by the academic library community¹⁰ in order to provide Canadian researchers with effective mechanisms to share, digitally preserve, and get credit for their data.
- *Collaboration.* A national Dataverse North service will work with other service providers and platforms like the Federated Research Data Repository (FRDR), institutional and regional repositories, and Canadian-based disciplinary repositories, in order to support research data management in Canada generally, and to look for collaborative opportunities around feature development, enhanced functionality, and cost-sharing where possible.

¹⁰ “Canadian university libraries have a long history of the kinds of collaborations required in the multi-stakeholder RDM environment, deep experience in developing programs to advance research, and critical expertise in preservation.” https://portagenetwork.ca/wp-content/uploads/2016/06/IATUL2016_Multi_Stakeholder_Engagement_in_RDM.pdf

- *Transparency.* “Transparency of digital curation costs will help data repositories identify greater efficiencies and pinpoint potential optimisations. Insight into how and why peers target their investments can lead to the better use of resources, help identify weaknesses and drivers in current practices, and inspire innovations.”¹¹
- *Consultation.* A national service will be created in consultation with researchers, librarians, developers, and other stakeholders within the RDM community, as well as with existing Dataverse service providers in Canada, to ensure that the work done at the national level is as supportive as possible of institutional or regional efforts.
- *Digital preservation.* A national service will be developed with long-term preservation in mind, including the development of integrations with preservation management systems like Archivematica, and preservation storage services provided by institutions or regions (e.g. university-based storage services, OCUL’s Ontario Libraries Research Cloud¹², and COPPUL’s WestVault¹³)

Relation to other Portage Services and Platforms

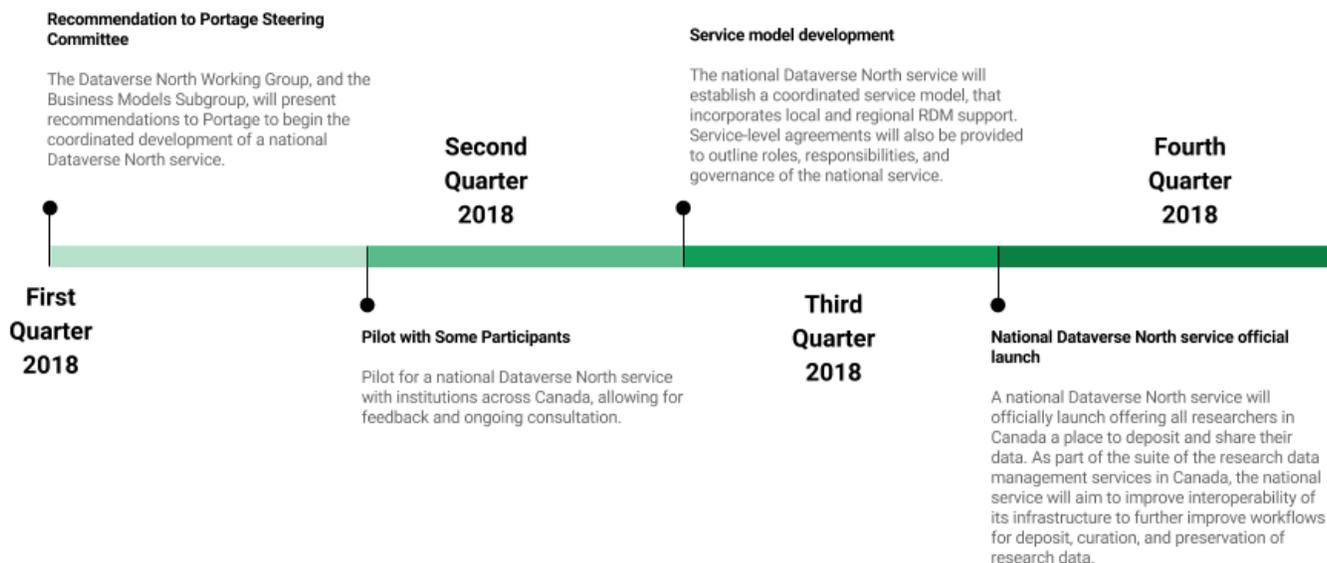
Portage, through its Network of Expertise and in partnership with library consortia, institutions, and other infrastructure partners, is coordinating the development of a suite of national services for RDM that will better enable academic libraries to serve their researcher communities. At this stage, the groundwork is being laid for services related to data management planning (DMPEG), curation (CEG), preservation (PEG), discovery (DDEG), and repositories for the storage, description, and publication of research data (DVN & FRDR), in addition to training modules for a wide range of RDM-related topics (TEG). As a part of this broader landscape, DVN will work in conjunction with existing and planned initiatives to support the vision of providing seamless and equitable access to RDM services for Canadian researchers and institutions.

¹¹ OECD (2017), "Business models for sustainable research data repositories", OECD Science, Technology and Industry Policy Papers, No. 47, OECD Publishing, Paris. Available at <http://dx.doi.org/10.1787/302b12bb-en>

¹² Ontario Libraries Research Cloud - <https://cloud.scholarsportal.info/>

¹³ COPPUL’s WestVault - COPPUL’s [WestVault](#)

Tentative timelines



Summary

Research repositories are an essential part of the infrastructure for open scholarship and open science. Research data repositories provide for the long-term stewardship of research data, thus enabling verification of findings and the re-use of data. They bring considerable economic, scientific, and social benefits. Hence, it is important to ensure the sustainability of research data repositories, especially in Canada, where we do not have central funding for open scholarship in the country. Many research data repositories are largely dependent on public funding. The key policy question to be addressed is how this funding is most effectively provided - by what mechanism and from what source? There are advantages and disadvantages of various business models in different circumstances that can greatly affect data repository operations.

The key recommendation put forward by this paper is the development of a Portage Dataverse North Service, hosted by Scholar's Portal, that will enable all Canadian researchers and the academic library community to effectively utilize a robust, scalable, and affordable research data repository platform.

Related documents

- [Summary of the Dataverse North Business Models Group Dataverse Providers and Institutional Dataverse Users/Clients Survey](#)
- [Dataverse business model evaluation report](#)