



Portage Progress Report

For the period of April 2018 - June 2018

The following report summarizes Portage activities and accomplishments in the second quarter of 2018, corresponding to the goals outlined in the Portage Business Plan: 2017 - 2018, and provides an update on Portage operations and communications undertaken in this period.

Reporting areas include work undertaken towards:

1. Fostering a community of practice for research data management (RDM);
2. Facilitating and providing leadership in the development of RDM infrastructure;
3. Engaging and advocating for RDM with stakeholder communities; and
4. Developing Portage operations and communications.

The mission of Portage is to contribute actively to RDM in Canada through a sustainable network of library-based services and collaborative infrastructure. The following report aims to communicate our progress towards the development of this envisioned national community of practice.

Report prepared by:

Jeffrey Moon (Director, Portage)
Julie Morin (Project Officer, CARL)
Lee Wilson (Service Manager, Portage/ACENET)

July 2018

1. Fostering a community of practice for RDM

A primary objective of Portage is to build a network of expertise for RDM. Critical aspects of this objective are to coordinate and expand on expertise and services within Canadian academic libraries and to build capacity in specific areas of RDM.

Activities and Accomplishments

The following update is centered around the work of the Portage Network of Experts.

Portage Council of Chairs:

- The Portage Council of Chairs met for a two-hour teleconference on April 4th, where they had the opportunity to discuss updates presented by members of the Portage Secretariat, and each of the Expert Group Chairs provided an update on the progress of their respective groups.
- The Council also met for an all-day in-person meeting at the University of British Columbia on June 4th, where members of the Portage Secretariat provided detailed updates on Portage activities, including federated architecture, and each Chair presented on the activities of their Expert/Working Groups. The group also had the opportunity to discuss proposals to the CANARIE RDM funding call, document management for the community of practice, coordination amongst groups, and work plans for 2019 and beyond.

Curation Expert Group (CEG):

- The objective of the CEG is to identify, evaluate and promote best practices, resources and workflows that support data sharing and reusability for researchers within projects, as well as data within repositories.
- CEG is developing a Data Curation Survival Guide, a one-pager which covers data curation essentials for individuals in curatorial roles.
- CEG is collaborating with the Research Intelligence Expert Group (RIEG) to design an information-gathering and outreach strategy/process for identifying data curators, expertise, and activities in Canada.
- CEG is in the process of drafting a data curation white paper on developing a national approach to data curation in Canada.
- CEG is working to establish a Canadian data curation network (following the model of the US Data Curation Network, or DCN) through interest groups and a possible national training event in collaboration with the Training Expert Group (TEG).

Data Discovery Expert Group (DDEG):

- DDEG contributes expertise to the Portage network around the discovery of Canadian research data and the development of a national data discovery service.
- DDEG identified five gaps in their submission to the CANARIE community consultation on the RDM Funding Program.
- DDEG is in the process of drafting a metadata-harvesting policy for FRDR.
- DDEG, in collaboration with UBC, Compute Canada and Scholars Portal, submitted a proposal to the CANARIE RDM Funding Call on federated geospatial data discovery.

- DDEG's FRDR Discovery Service Working Group has initiated four new "task forces," each focusing on a different area of development for a national discovery service:
 - **Research Data Repositories Task Force** aims to identify and make recommendations on priority repositories for metadata harvesting to the FRDR team and Communications and Outreach TF.
 - **Controlled Vocabularies & Subject Headings Task Force** will explore appropriate existing subject headings and controlled vocabularies for use in FRDR.
 - **SHARE Project / Discovery of Related Research Outputs Task Force** will connect with the US-based SHARE project and examine other open linked data projects for research management information.
 - **Communications and Outreach Task Force** will forge connections with repositories identified by the Research Data Repositories TF or other sources, develop materials and strategies for conducting outreach, and track outreach activities.

Data Management Planning Expert Group (DMPEG):

- DMPEG continues their work overseeing the development of the Portage national data management planning platform, the DMP Assistant.
- DMPEG was well-represented at IASSIST/CARTO 2018 in Montreal, with a Poster presentation on Data Management Planning.
- DMPEG was also represented at the CASRAI Conference (Weiwei Shi & Ted Strauss) with a well-received presentation on the DMP Assistant and future developments (see below).
- DMPEG working closely with the University of Alberta on DMP Assistant 2.0 update/migration.
- DMPEG has two new Working groups:
 - Exemplars WG: discipline-specific DMPs and
 - Policy WG: looking at criteria for evaluating DMPs, e.g. for adjudication purposes.

Preservation Expert Group (PEG):

- PEG provides Portage with advice on RDM infrastructure developments supporting the long-term stewardship and access of research data and metadata.
- PEG released its White Paper, [Research Data Preservation in Canada](#), on May 15, 2018.
- PEG identified two gaps in their submission to the CANARIE community consultation on the RDM Funding Program (need for inventory of preserved research datasets & need for national file format registry/metadata standards policy). Neither of these gaps were converted into funding applications at this time.
- Three submissions (IDCC, IASSIST, iPres) were accepted based on the work of this group.

Research Intelligence Expert Group (RIEG):

- RIEG provides Portage with intelligence on RDM in Canada and are responsible for managing materials and outputs of the Canadian RDM Survey Consortium.
- RIEG developed an RDM Taxonomy which was used to analyze key RDM documents to identify and compare topics represented in existing knowledge pertaining to RDM

in Canada and internationally. The taxonomy contains 73 topical terms and 8 categories, including intended audience and document type.

- 110 documents have been classified using this taxonomy.

Training Expert Group (TEG):

- The primary objective of TEG is the development and maintenance of a high-level program for supporting RDM training across Canada.
- Jane Fry and Carol Perry presented a paper titled “Sustaining our data through data management - it’s easier than you think!” at IASSIST 2018. Jane Fry also presented a poster titled “Portage - Research Data Management made easy!”
- The RDM-101 and DMP online training module Working Groups completed the content development phase of their work. Scholars Portal has provided ongoing access to a learning management system, where the learning modules are currently being developed.
- In September TEG will be circulating a call for a new Working Group to develop a Repositories 101, which will cover Dataverse and FRDR.
- TEG working with other Expert Groups to develop additional ‘Busy Bee’ one-page quick guides (e.g., Curation Expert Group plans to add to this series).

2. Facilitating and providing leadership in the development of RDM infrastructure

Portage seeks to advance the development of national platforms for planning, curating, preserving and discovering research data. This will require working with infrastructure providers - both in the library community and elsewhere - to develop new tools where gaps exist and to bridge systems where interoperability is needed.

Activities and Accomplishments

DMP Assistant:

- Planning for migration of the Portage DMP Assistant to a new merged Digital Curation Centre (DCC) and California Digital Library (DCL) codebase (known as DMP Roadmap) is underway, thanks to the in-kind support of the University of Alberta, and specifically, the work of Weiwei Shi at the University of Alberta, in consultation with the DMP Expert Group.
- Version 2.0 of the DMP Assistant promises exciting new features, including a foundation for machine-actionable DMPs, ORCID integration, researcher ‘opt-in’ for DMP sharing, and APIs to facilitate sharing DMPs.
- Launch anticipated in Fall 2018.

Dataverse North Working Group (DVN):

- DVN was formed to contribute to a community of practice for libraries using or interested in using the Dataverse repository platform for research data in Canada. Work is being advanced by three sub-groups investigating gaps and opportunities that could be addressed by nationally coordinated strategies and services for Dataverse; including, the Business Models Sub-Group, the Training Sub-Group, and the Metadata Sub-Group.

- DVN Business Models subgroup, in collaboration with the FRDR-SM Working Group, submitted a gap analysis on repository infrastructure to the CANARIE RDM Funding Consultation.
- Work toward a national instance of Dataverse (as proposed by the Business Models Sub-Group) is evolving. OCUL-Scholars Portal is working with BCI in Quebec to improve ‘internationalization’ features of Dataverse with the goal of providing Dataverse services to Quebec institutions. Portage and the Portage Dataverse North Working Group are monitoring these developments and maintaining channels of communication with the stakeholders involved.
- DVN is working with the Training Expert Group (TEG) to develop repository-related training materials. The goal is to develop multi-modal materials with priority on introductory manuals, videos, and workshops aimed at librarians and library staff.
- The DVN Metadata Subgroup is continuing its work on developing a common ‘base metadata template’ for Dataverse using standard vocabularies and identifiers. Additionally, this group is working to facilitate the development and adoption of a common approach to labelling, describing, and organizing data files in Dataverse repositories.

Federated Research Data Repository (FRDR):

- FRDR is a joint project between CARL/Portage and Compute Canada to create a national discovery layer for Canadian research data and a scalable repository platform.
- FRDR is currently operational in a “limited production” capacity, which means that:
 - Anyone may use the service to search for and download data from the FRDR repository or from a growing catalogue of Canadian data repositories;
 - FRDR’s repository services are being offered to a growing list of Canadian researchers. Research projects vary widely in both scope and discipline, and include:
 - The Global Institute for Water Security (University of Saskatchewan)
 - The Mountain Legacy Project (University of Victoria)
 - Stone Tools, Diet and Sociality at the Dawn of Humanity (University of Calgary)
 - Diversity Cape Breton (Cape Breton University)
 - The Water Institute (University of Waterloo)
 - Ocean Networks Canada-DFO (University of Victoria)
 - Deep learning and the Schrödinger equation (National Research Council of Canada)
 - As well as individual researchers from several universities
- A priority for FRDR in 2019 will be moving out of limited production and into a full service launch.
- Discussions are ongoing with Innovation, Science, and Economic Development Canada to discuss the release of priority funding to support FRDR moving into full production, as well as other service-related initiatives.
- FRDR continues to work closely with the Portage EGs and WGs, most notably PEG, CEG, DDEG, TEG, and FRDR-SM WG in the development of a suite of national data services.
 - DDEG has recently instituted a FRDR Discovery Service Working Group focused on developing aspects of the national discovery layer, including

communications and outreach, controlled vocabularies, open linked data initiatives (e.g., SHARE), and identifying priority Canadian repositories for harvesting.

- As noted above, members of CEG will be working with FRDR in a curation pilot as a means of laying the groundwork for a distributed, Canadian data curation network.

FRDR Service Model Working Group (FRDR-SM):

- FRDR-SM WG was formed to draw on community expertise in the development of service and business models, policies, and user experience and operational workflows for a federated research data repository platform.
- The **Policy Subgroup** has undertaken the development of several policies in support of operating a national-scale data repository service. These efforts are nearly complete, with the majority of the policies now drafted and ready for review by the FRDR Steering Committee in the Fall.
- The **Workflows, User Experience, and Training Subgroup** continues to provide suggestions to improve the FRDR platform user interface design. The group has now finished an extensive overhaul of all user-facing documentation associated with the service. The new documentation will be published on the site shortly.
- The **Business Models Subgroup** is expected to begin activities this coming Fall.

3. Engaging and advocating for RDM with stakeholder communities

Portage seeks to advance the development of national platforms for planning, curating, preserving and discovering research data. This will require working with infrastructure providers - both in the library community and elsewhere - to develop new tools where gaps exist and to bridge systems where interoperability is needed.

Activities and Accomplishments

Collaboration with Regional Library Councils:

- Regularly held group teleconferences between the Portage Director and representatives from all four regional library councils.

Outreach to Non-CARL Libraries:

- On May 15th, Jeff Moon presented at a ½ -day Workshop targeting non-CARL libraries in Canada, sponsored by Wilfrid Laurier Libraries.

Portage is negotiating an agreement with Clarivate Analytics to make Canadian research data discoverable in their Data Citation Index (DCI):

- If an agreement is reached, FRDR and its partner repositories will benefit from the provision of DCI's Article-Match-Retrieval system to provide data-use and citation metrics, a free DCI account for each of the 'data providers' feeding metadata into the FRDR discovery layer, and discoverability in an international metadata aggregator.

Portage Working Group on Responsible RDM Practices for Sensitive Data:

- In April, a dedicated [webpage](#) on the Portage site was created and the group set the following priority objectives: 1) Deposit-friendly text for ethics applications & informed consent & data access agreements, 2) Training materials, 3) Environmental scan of best practices for Indigenous data.
- In May, three sub-groups were established to work on these objectives, and they have been meeting regularly.

Tri-Agency RDM Policy Consultation:

- The Tri-Agency opened its public consultation on their RDM Policy on June 5th. Portage is working on a response to this consultation which they will share in draft form by mid-July. To that end, Portage scheduled three ‘Town Hall’ meetings in early July to give members of the network of experts an opportunity to provide feedback and ask questions. Portage encourages all stakeholders -- individuals, institutions, organizations, and associations, to submit feedback to the Tri-Agency.

Meetings with Compute Canada:

- CARL/Portage and Compute Canada are working together to develop a common and collaborative vision of advanced research computing (ARC) and research data management (RDM) within the realm of Digital Research Infrastructure (DRI) to share with Innovation, Science and Economic Development (ISED) Canada. Executive Directors of both CARL and Compute Canada endorse this approach which reflects our shared desire to see ARC and RDM represented positively and collaboratively to ISED in the advancement of DRI in Canada.

Other Engagements:

- Jeff Moon (Director, Portage) presented at the Research Data Canada sponsored/organized meeting for Vice-Principals of Research in Toronto on April 13.
- Susan Haigh, Jeff Moon, Donna Bourne-Tyson (CARL President), and Martha Whitehead (LCDRI member & UL) met with ISED on April 18.
- Shahira Khair (Project Office, CARL-Portage) and Eugene Barsky (Chair of the Data Discovery Expert Group) presented at BCNET, in Vancouver, April 23-26.
- Jeff Moon and Jeremy Geelen (Policy Analyst, SSHRC) presented at the CAREB Conference in Montreal, April 26-28.
- Jeff Moon presented at the CARL Annual General Meeting in Regina, May 1.
- Jeff Moon presented at Queen’s University Data Day, May 2.
- Jeff Moon presented to non-CARL Libraries during a ½ -day Workshop at Wilfrid Laurier University in Waterloo.
- Lee Wilson (Portage Service Manager) delivered a DMP Workshop at Cape Breton University, Cape Breton, NS, May 1 & 2.
- Jeff Moon delivered a webinar for Datacure, “Portage for Research Data in Canada, Eh?”, on May 4.
- Jeff Moon delivered the webinar, “Granted! Planning for Stellar Research: Data, Documentation & Dissemination”, for the Saskatchewan Centre for Patient-Oriented Research.

- Lee Wilson delivered a DMP Workshop to librarians and associated staff at Saint Mary's University in Halifax, NS, May 15.
- Portage Director met with representatives of the FRQ (Fonds de Recherche du Québec) in Montreal in May. The FRQ is similar to the Tri-Agency, with three discipline-based divisions: science/engineering, health, and social sciences and humanities.
- Portage Director gave a presentation on Portage to representatives of BCI in May.
- Lee Wilson presented on "Current Trends in Research Data Management in Canada" at Saint Mary's University in Halifax, NS, May 25.
- Portage Director and Service Manager presented "CARL Portage: The RDM Journey Continues" at the 2018 IASSIST-CARTO Conference in Montreal in May.
- Lee Wilson attended the CANARIE Research Software Conference in Ottawa in May.
- Weiwei Shi (University of Alberta, and systems support for DMP Assistant) and Ted Strauss (McGill University) gave a presentation on the DMP Assistant and migration to DMP Roadmap (merged DCC and CDL codebase) at CASRAI 2018.
- Jay Brodeur, Chair of the Curation Expert Group, joined the Data Curation Network Advisory Committee, and attended the Canadian Data Science Workshop in Toronto.
- Lee Wilson presented as a part of an RDM panel at the Atlantic Provinces Library Association (APLA) conference in Fredericton, NB in June.
- Jeff Moon and Lee Wilson (Service Manager, Portage/ACENET) attended the CANHEIT-TECC conference at Simon Fraser University in Burnaby, BC, June 18-21. Jeff gave a Portage presentation and served on a panel looking at RDM in Canada. Lee co-presented on the FRDR Project.

Portage-related Submissions to the CANARIE RDM Funding Program:

- Portage Expert and Working Groups identified a number of 'gaps' in the RDM ecosystem and flagged these during the community consultation in May, 2018. Submissions included:
 - Dataverse North & FRDR:
Gap: Not all researchers have access to robust Canadian repository services
 - Curation Expert Group:
Gap 1: There is a need to develop a national approach to data curation in Canada
Gap 2: There are challenges associated with deposit and appropriate sharing of sensitive data
 - Preservation Expert Group:
Gap 1: There is no inventory of preserved research datasets across Canada
Gap 2: There is a need for a national registry of file format and metadata standard Policies.
 - Discovery Expert Group:
Gap: There are shortcomings in the discovery layer of Canadian RDM ecosystem.

In addition, Portage collaborated on two other gap submissions with CASRAI and Compute Canada:

- CASRAI:
Gap: Need to better integrate RDM with RMI (Research Management Information)
 [e.g. ‘auto-harvest’ information from CVs, the need for portable Data Management Plans (DMPs), interfaces to improve interoperability between RMI and RDM systems, need to make the CASRAI Dictionary machine-readable, interoperable, and bilingual, and need for a federated register of semantic resources to improve system interoperability]
 - Compute Canada:
Gap: Need to address management of ‘active data’ in pre-publication phase of research life cycle
- Portage also provided letters of support for four proposals to the subsequent CANARIE RDM Funding Call:
 - A National Geospatial Discovery Layer proposal that will use GeoBlacklight to improve the discoverability of geospatial metadata harvested by FRDR and create connections between Dataverse and GeoServer that will enhance Dataverse’s ability to visualize geospatial information and output geocoded metadata to be harvested by FRDR. This proposal is led by UBC with support from Compute Canada, Scholar’s Portal, and CARL/Portage.
 - USask and SFU, with support from Compute Canada, WestGrid and CARL/Portage, propose to develop a metadata indexing, aggregation, and enhancement tool for active research data management built to integrate directly (via APIs and custom plugins) with existing tools in this space, specifically HubZero and Open Science Framework, as well as support the packaging of datasets for publication into data repositories.
 - A proposal to develop features in Dataverse to make it a better fit for use as a national repository option. This will include integration with Shibboleth, with the OLRC, and enhancing workflows for uploading and downloading data in Dataverse, especially large datasets (by connecting Dataverse to distributed storage providers e.g. Globus, SWIFT, AWS), to be led by Scholars Portal with support from CARL/Portage.
 - The Érudit project aims to fund the development of a web service allowing to assign unique permanent identifiers (DOIs, ORCID) in a semi-automatized way, in order to ensure standardized tagging of author names, grants agencies and bibliographic references of vast structured textual corpora. It thus aims to accelerate the implementation of good research data management practices, to increase interoperability of funded research data, to enhance its discoverability and to ensure its reusability.

CANARIE RDM Software Funding Call:

- Prior to the CANARIE RDM Funding Call, Portage and Compute Canada collaborated with OCUL Scholars Portal on a submission to the CANARIE Research Software fund in May 2018. This proposal targeted increased support for geospatial data in

Dataverse and a geo-discovery layer for the CARL/Portage-Compute Canada Federated Research Data Repository (FRDR), built on the open-source GeoBlacklight platform.

- Ultimately, this proposal was not funded, but scored very well -- leading to its resubmission under the CANARIE RDM Funding call (*see above*).

4. Developing Portage operations and communications

Portage is a CARL initiative and is currently supported through its membership. The goal is to develop a sustainable network built upon the support of the wider research data management ecosystem.

Activities and Accomplishments

CARL Directors Portage Steering Committee (CDPSC):

- The CDPSC planned for a 'Future Direction of Portage' session at the CARL Spring Meeting, on May 1st. A task group of committee members (Brett Waytuck, Donna Bourne-Tyson, Jeffrey Moon, Martha Whitehead, Beth Namachchivaya, Pascal Calarco, and Susan Haigh) was formed to work on a detailed agenda and an outline of short to medium-term funding options, which were approved at a special meeting of the CDPSC on April 4th. Feedback from the session and the anonymous post-session survey were overwhelmingly supportive, and members were pleased with Portage's accomplishments to date.
- The CDPSC also held a one-hour teleconference on June 13th to discuss Expert Group proposals to the CANARIE RDM funding call and abstracts of external proposals Portage was involved in. Also discussed were the Dataverse North Business Model Report, Tri-Agency Consultation Response, and development of a two-page Briefing Note on Portage. Initially framed as 'key messages for CARL Directors for on-campus discussions', this briefing note was subsequently converted into a more [general document](#). Finally, committee members were debriefed on the 'Future Direction of Portage' post-session poll results and the all-day Council of Chairs meeting.

Portage Advisory Committee (PAC):

- The PAC met on May 22nd and discussed the Portage Winter Progress Report and Director updates, project funding opportunities, updates on the Responsible RDM Practices for Sensitive Data Working Group, FRDR Limited Production, the Portage Session at the CARL Annual General Meeting, and stakeholder engagement activities since the last meeting.

Portage Project Officer:

- Shahira Khair, Project Officer with CARL and Portage since January 2017, left CARL and Portage on May 8, 2018 after accepting a position at the University of Victoria as Data Curation Librarian. A process is underway to hire a new Portage Project Officer. More information will be provided once this process is complete.