Collaboration and Platform Integration in Support of a Federated Research Data Management Service in Canada

Donna Bourne-Tyson, President, Canadian Association of Research Libraries (CARL) / University Librarian, Dalhousie University
Lee Wilson, Portage Service Manager / AceNet RDM Consultant
Corey Davis, Digital Preservation Network Coordinator, Council of Prairie and Pacific University Libraries (COPPUL)
Portage is a national RDM network launched by the Canadian Association of Research Libraries (CARL) in 2015 that coalesces initiatives to build RDM capacity and infrastructure with libraries.

1. **Networks of Expertise**:  
   - Pan-Canadian RDM expertise  
   - Provide access to resources, tools, and experts in the area of RDM

2. **Infrastructure Platforms**  
   - Working with library consortia, institutions, and other infrastructure partners to assemble essential RDM infrastructure and service components
Portage Network of Expertise

Six Expert Groups, with associated working groups

• Data Management Planning
• Curation
• Data Discovery
• Preservation
• Training
• Research Intelligence

A number of additional working groups, including:

• Dataverse North WG
• FRDR Service Model WG

And two more coming…
Portage Network of Expertise

Two New Working Groups

- **Portage Institutional RDM Strategy Working Group**, aims to develop a *generic template* for institutions to build upon in response to the first pillar of the anticipated Tri-Agency funding policy.

- **Portage Ethical Treatment of Sensitive Data Working Group**, will help Portage frame and develop resources and supports for researchers around this key data management issue.
DMP Assistant and FRDR
The DMP Assistant is an online, bilingual, data management planning tool that will help researchers create their own data management plans.
What is FRDR?

FRDR = Federated Research Data Repository (DFDR en français)

A scalable, federated platform for digital research data management and the discovery of Canadian research data

Discovery  Storage  Preservation  Curation
FRDR’s harvester indexes data repositories across Canada to make research data held in many repositories discoverable from a single platform

- Currently harvesting 31 repositories (roughly 125,000 datasets discoverable)
- Supports OAI-PMH and CKAN endpoints
- Metadata can be crosswalked or adv. search facets provided

Goal: supplement existing repository sites and improve discovery
- Breakdown repository siloing, avoid being “just another repository”
Storage

- A place for Canadian researchers to deposit large datasets
  - Big data transfer using Globus File Transfer
- A place to deposit datasets if researcher does not have a local option
- Designed for scalability
- Storage may be distributed regionally or managed centrally through infrastructure providers (e.g., Compute Canada)
Preservation Processing

- Archivematica integration: Preservation processing for long-term usability of datasets
- Automatic Archivematica processing for datasets up to 300 GB or 25,000 files
- Archival Information Packages (AIPs) transferred to Preservation Service Providers for long-term storage
  - E.g., Ontario Library Research Cloud (OCUL) & COPPUL’s Digital Preservation Network
Dataverse North and preservation
Dataverse North Working Group

- A Portage Working Group with 26 members from across Canada
- A community of practice to coordinate training, support services, outreach strategies, promotions, and infrastructure development
- Explore nationally coordinated strategies and services
Dataverse North goals

- Provide **equitable access** for Canadian researchers
- **Standardize services** across Canada
- **Support users** (librarians and researchers)
- Guide **feature developments and integrations**
Dataverse North strategic priorities

- **Business Models**
  - Develop a framework for hosting and support services
  - Explore a common business model

- **Training**
  - Identify and coordinate training opportunities

- **Metadata**
  - Recommendations for templates, data files, and DOIs
Dataverse North strategic priorities, phase 2

- **Preservation services**
  - Connecting Dataverse infrastructure to Archivematica and preservation storage services

- **Advocacy and promotion**
  - Identified as a major issue in a recent survey
Portage Preservation Expert Group (PEG)

- Portage is dedicated to building a national research data preservation infrastructure through a disaggregated service model

- In this model, preservation functions are broken apart into separate services distributed over various locations or multiple organizations

- Archival/preservation storage (and related services) will be outsourced to other organizations who have the capability and experience
Disaggregated preservation service model

- Portage keeps abreast of the changes in the external environment and implements strategies that can help to mitigate risks.
- Portage will coordinating activities between participating organizations (Preservation Service Providers).
- It will prescribe a set of basic preservation requirements for each PSP.
- Full details will be made available in early 2018 in a Portage whitepaper (see IDCC abstract for basic details).
Bringing it all together
Vision for Canadian RDM Services

We envision seamless, equitable access to RDM platforms and services for Canadian institutions and researchers
Vision for Canadian RDM Services

- National, institutional, or domain-specific repositories
- Feed metadata into the national discovery layer
- Facilitate AIP creation

- Tools for data management planning
- RDM training modules targeted at researchers, librarians, and policy and decision-makers
- Training at the institutional and regional levels

- Adhere to a set of agreed upon standards for data preservation
- Provide long-term archival services
- Feed metadata into the national discovery layer
Vision for Canadian RDM Services

Infrastructure Providers

National Discovery Layer

Planning & Training

Repository Services

Preservation Service Providers

Network of Expertise
Next steps

- Integrating services on a single, end-user facing platform
- Defining business and service models for national RDM services
- Securing sustained funding for RDM in Canada
- Facilitating funders’ RDM policy development and implementation