Opportunities for Academic Libraries to Shape National Approaches to Research Data Management: A Canadian Perspective

Susan Haigh, Executive Director, CARL; Lee Wilson, Service Manager, Portage; Jay Brodeur, Associate Director, Digital Scholarship Services, McMaster University
Overview

• Opportunities to shape RDM at various levels:
  – National policy and funding level
  – Systems and services level
  – Expert capacity and community level - the need for curation support
A national initiative of the CARL to **build research data management (RDM) capacity in Canada** through a **network of experts** in a growing community of practice.

### Network of Experts

- Pan-Canadian RDM expertise with over 100 in-kind experts
- Contribute to development of RDM resources, tools, training, and policy

### Infrastructure Platforms

- Work with library consortia, institutions, and other infrastructure partners to develop & deploy essential RDM infrastructure and service components
Portage Expert & Working Groups

Expert Groups

- Data Management Planning
- Curation
- Data Discovery
- Preservation
- Training
- Research Intelligence
- Data Repositories

Working Groups

- Dataverse North WG
- Responsible Research Data Management Practices for Sensitive Data
- FRDR Service Model WG
- FRDR Discovery Service WG
- Institutional RDM Strategies

130+ Experts
60+ Organizations
National policy and funding
New Federal Funders’ Policy on RDM Expected in 2020

Draft Tri-Agency RDM Policy:

- Institutional Strategy
- DMPs
- Deposit
Portage responses to enable policy implementation...

Draft
Tri-Agency
RDM Policy

Institutional Strategy
DMPs
Deposit

RDM Strategy Template
DMP Assistant
National, online, bilingual, Data Management Planning Tool
National, Multi-disciplinary Repository Options
Harnessing Big Data

“The Government proposes to provide $572.5 million over five years, with $52 million per year ongoing, to implement a Digital Research Infrastructure Strategy that will deliver more open and equitable access to advanced computing and big data resources to researchers across Canada.”
Moving toward a new DRI organization

Government of Canada

$375 million

CANARIE

Digital Network for Research and Education

$137 million

New DRI Organization

$2 million

Research Software

CANARIE CFI

Compute Canada

$50 million

Advanced Research Computing
Systems and services
Key transition priorities for Portage

A. National Data Stewardship Support
Providing coordinated support at the national level

B. RDM Platform Support and Development
Develop and operationalize national, collaborative RDM platforms and services
A. National Data Stewardship Support

All 6 positions filled. Starting now through September 2020

National Portage Secretariat:
- Training Coordinator
- Curation Coordinator
- Discovery & Metadata Coordinator
- Preservation Coordinator
- Project Manager
- Communications Officer
  + translation, website, travel, training & outreach

$2 Million (1-year transition)
DMP Assistant is a bilingual tool for preparing data management plans (DMPs). The tool follows best practices in data stewardship and walks researchers step-by-step through key questions about data management.

- **Step 1**: Sign up with DMP Assistant
- **Step 2**: Sign in and select a template under Organizations. The Portage template is the default.
- **Step 3**: Answer the questions that are relevant to your work. Guidance and examples are provided.
- **Step 4**: Revisit the tool throughout your research to review or revise your answers.

**Sign in**

If you have an existing account with DMP Assistant or previous version of DMP Builder.

**Sign up**

New to DMP Assistant? Sign up today.

Please note that we are currently working on single sign-in authentication. For now, please create a new DMP Assistant account. You will have the option to link your DMP Assistant account to your campus ID when that feature becomes available.
DMP Assistant -- Transition to DMP 2.0

Portage DMP Assistant 2.0

Based on single, open-source platform: DMP Roadmap

Joint initiative:
• Digital Curation Centre
• University of California Curation Centre

New features:
• Provides foundation for machine actionable DMPs
• Supports DMPs as “living documents”
• Supports use of ORCiD IDs
• Includes APIs for creating and sharing DMPs
• Supports researchers ‘opting in’ to share their DMP
Developing National Repository Options
What is Dataverse?

- Open-source repository platform developed by the Institute of Quantitative Social Science (IQSS) at Harvard University
- 48 installations around the world
- Active development community
- Developed with quantitative social science data in mind, but not limited to any domain

From: *National Repository Updates*, Canadian Data Curation Forum, Oct 18, 2019, Brad Covey, Meghan Goodchild, Kaitlin Newson, Reyna Jenkyns, Lee Wilson
Scholars Portal Dataverse

Scholars Portal Dataverse by the numbers:

# Institutions: 48

# Dataverses: 450

# Datasets: 2,307

# Files: 29,862

# Downloads: 178,642

Nov 12, 2019

From: National Repository Updates, Canadian Data Curation Forum, Oct 18, 2019, Brad Covey, Meghan Goodchild, Kaitlin Newson, Reyna Jenkyns, Lee Wilson
Building a national service

National multi-stakeholder engagement

1. Developing partnerships and agreements with other regional library consortia
   a. Agreement with BCI to extend service to 19 university libraries in Quebec
   b. Working on agreements with COPPUL and CAUL

2. Engaging with Portage’s Dataverse North Working Group
   a. Policy subgroup, e.g., Curation

From: National Repository Updates, Canadian Data Curation Forum, Oct 18, 2019, Brad Covey, Meghan Goodchild, Kaitlin Newson, Reyna Jenkyns, Lee Wilson
Building a national service

Local connections and touchpoints

1. Scholars Portal Dataverse community
   a. Support email listserv
   b. Monthly meetings
   c. Testing and feedback

2. Training and outreach to help build local capacity
   a. Introductory YouTube videos
   b. User Guide with updated FAQ
   c. New Scholars Portal Dataverse Blog
   d. Dataverse North Training Group Materials

From: National Repository Updates, Canadian Data Curation Forum, Oct 18, 2019, Brad Covey, Meghan Goodchild, Kaitlin Newson, Reyna Jenkyns, Lee Wilson
Transitional funding will support storage expansion to further subsidize costs for institutions
A scalable, federated platform for digital research data management and the discovery of Canadian research data

- Partnership between Compute Canada Federation (CCF) and CARL Portage
- A full-featured repository for research data publication
  - A place for Canadian researchers to deposit large datasets
  - Designed for scalability (TB+ scale), leveraging Globus File Transfer
  - Storage and compute geographically distributed at multiple hosting sites
- A national discovery service for Canadian research datasets
National Discovery Layer

Metadata harvested from >40 repositories indexing over 100,000 datasets

Metadata harvested into FRDR

Repository Storage

Regional/Institutional Repositories

Discipline-specific Repositories

Research data & metadata Deposit & Download

Active Storage

Researchers

BIG data

Federated Research Data Repository (FRDR)

Regional/Institutional instances of Dataverse

Portage

Portage: Dataverse North

COPPUL
Council of Prairie & Pacific University Libraries

OCUL
Ontario Council of University Libraries

BCI

CAUL
Council of Australian University Librarians

CBUA
National Discovery Layer

- Improve discovery of Canadian research (meta)data
- Break down repository silos
- Drive traffic to existing repository sites
- Create interoperability between Canadian and international platforms

https://www.frdr.ca/
• Limited Production deposits as of Nov 2019:
  • 38 partnering research groups
  • 100 datasets published (8.4 TB)

• Full-production launch planned for Fall 2020

https://www.frdr.ca/
FRDR Transitioning to Full Production

- Funding will support:
  - Storage hardware and System Administrators at two national CCF hosting sites
  - Developer and Technical Operations Manager

- At launch, all researchers (PIs) and librarians, or their designates, from Canadian postsecondary institutions can deposit

- Users will have access to **curated repository storage** without cost up to a certain threshold, with mechanisms in place to obtain additional storage as needed
FRDR Transitioning to Full Production

- Work with Network of Experts (TEG/FRDR UX WG) to develop user guides, training materials, and webinars
- Interim governance model for FRDR
- Work with host site partners to bring new storage online and define service level expectations
Expert capacity & community
Portage Curation Expert Group (CEG)

Formed in 2016

10 members, from across Canada
• Librarians + library staff
• CARL Portage staff
• Faculty members & postdoctoral fellows

Goal
• Identify, evaluate, and promote best practices in data curation
• Inform national, regional, and institutional data curation activities and capacity building
Tapping (out) the CEG

Data curation activities: Who, where, what?
- Geographically, disciplinary
- Activities, skills, and knowledge

What are the barriers and needs?
- Within research groups, organizations, gov’t agencies, data repositories
- During active research, archiving, preservation

How might a national data curation approach meet these needs?
- High-level coordination, governance, funding
- Services, expertise, training
- Communities of practice
- Standards, expectations, workflows
The Canadian Data Curation Forum

3-day event | Funded by SSHRC, Portage, McMaster University Library

Training event
- 6 workshops | 13 facilitators | 57 attendees
- **Topics**: data curation, responsible & reproducible research

Community-Building Forum
- 61 participants | institutions + research orgs + funders + support orgs
- **Objective**: Articulate a vision for a national approach to data curation

data-curation.github.io
How could / should data curation be supported at a national level?

What are the existing needs and challenges?

What resources & services could meet these requirements?

What are the strengths, weaknesses, and constraints?

How might these be provided and coordinated (i.e. as a model)?
# CDCF Outcomes: Current Barriers / Challenges

## Capacity
- People, expertise, time
- Service creation / scaling
- Funding

## Guidance & Training
- Training opportunities
- Communities of practice
- Resources & tools

## Coordination
- Roles & responsibilities
- Models for cooperation
- Standards & principles
- Common vocabularies
- Shared workflows

## Advocacy / Culture
- Communicating value
- Incentivizing practices
- Demonstrating impact

## Governance

## Communication

## Equity & Inclusion
- Multilingual support
- Indigenous data
- Non-academe data
CDCF Outcomes: Modelling a national approach

6 working groups | 1.5 days
• **Task**: Design & refine a model for a national data curation approach

Common themes & elements
National activities should *empower* institutions, orgs, researchers

• **Coordinate** between national, regional, local services
  – Centralized & decentralized
  – Formal & informal

• **Provide more training** opportunities and resources

• **Build networks** to foster (broad) communities of practice

• **Advocate and promote** support from governments and funders
Towards a national approach to data curation

CDCF Report

M1

M2

M3

Recommendations Document

- CEG & Portage
- Institutions / Orgs
- NDRIO
Acknowledgments

**CARL Portage** is supported by directed funding from Innovation, Science & Economic Development Canada flowing through CANARIE

The **Canadian Data Curation Forum** was supported by funding from the Social Sciences and Humanities Research Council of Canada, McMaster University Library, and CARL Portage