CARL & E-Research

Gwendolyn Ebbett
University of Windsor
IATUL Seminar
24 - 26 OCTOBER 2016
University Library of Bozen-Bolzano (South Tyrol, Italy)

CARL & E-RESEARCH

Gwendolyn Ebbett
University of Windsor
Ontario, Canada
CARL members include Canada’s twenty-nine largest university libraries as well as two national libraries. Enhancing research and higher education are at the heart of its mission. CARL develops the capacity to support this mission, promotes effective and sustainable scholarly communication, and public policy that enables broad access to scholarly information.
Open science context

Open science means:

• Sharing and access to all types of research outputs
• Transparency of research findings
• Open peer review & open usage metrics
• Equitable flow of knowledge

Began with open access to publications, but now includes data and other types of research outputs

Parallels to Open Government/Open Data movement
Current policy context in Canada
Canada’s Action Plan on Open Government 2014-2016

“...the Government of Canada will establish a government-wide approach to Open Science to increase access to federally-funded scientific publications and data.”

Tri-Agencies (CIHR, NSERC, SSHRC)
- Consultation in 2013 (Report: Capitalizing on Big Data: Toward a Policy Framework for Advancing Digital Scholarship in Canada)
- Comprehensive Brief on Data Management Policies, April 2015
- Tri-Agency Statement of Principles on Digital Data Management
Recent RDM Principles

Tri-agency Statement of Principles on Digital Data Management

- Canadian Institutes of Health Research
- Natural Sciences and Engineering Research Council of Canada
- Social Sciences and Humanities Research Council of Canada
- January 2016

Research Data Management in Canadian Universities: A statement of principles

- Research Data Canada and Vice-Presidents of Research from the U15
- March 2016
But we need more than just policies
Research Data Management is like a three-legged stool...

- Policies
- Culture
- Infrastructure and services
A lot of this is about cultural change socializing the community
Many researchers would rather share their toothbrush than their data...
Infrastructure and services

Disciplinary repositories

Integrated domain data centres

But these only collect a small portion of research data produced in Canada
The majority of datasets fall into the “long tail”

Canadian academic libraries: strong collaborations
Multiple RDM Stakeholders
Vision

A future in which Canada capitalizes on the trend towards data intensive research and is a world leader in research and innovation

A national, library-based research data management network that coalesces initiatives in research data management to build capacity and to coordinate activities better

https://portagenetwork.ca/
Portage Mandates

- Developing a national research data culture
- Fostering a community of practice for research data
- Building national research data services and infrastructure
From pilot to network

- March 2014 – library-based, CARL-led, one-year project, aiming to leverage expertise and emerging initiatives in research data management. Objectives:
  - data management plan support
  - network of expertise
  - demonstration data preservation activity
  - plan for ongoing RDM network
- May 2015
  - CARL agreed to fund Portage Director to lead two-year transition to operational network
- September 2015 – Chuck Humphrey, Director
Portage Principles
(December 2013)

- Research data are a public good
- Intelligent access: openness, with respect for privacy
- Collaborative approaches: cost savings and shared expertise
- Inclusiveness: aim to serve all researchers and create a more level playing field
- Commitment to standards and interoperability
- International relationships: liaise internationally and ensure Portage keeps pace with international practices
- Respect for differences: flexibility to meet the needs of different regions, institutions, and disciplines
- Open source: Tools will be contributed back to the community
- Stewardship: a sense of responsibility for managing research data over the long term
Bottom-up approach

Mandated institutions working collaboratively

Information professionals

Researchers, societies and their domain projects

Granting council and federal agency digital data management principles

Academic research data management principles
Shared stewardship at institutional level and nationally...

Libraries

Researchers

Grad Studies

Individuals, Groups and Services

Ethics

IT

Research Services
Stakeholder engagement

Build a community of practice for research data management through well established working relationships among stakeholders across research sectors

○ Regional library consortia

○ Individual academic libraries

○ Academic RDM stakeholders, e.g., VPR’s, REBs, Scholarly Associations, Networks of Centres of Excellence

○ Other RDM stakeholders, e.g., Research Data Canada, Leadership Council for Digital Infrastructure
THE PORTAGE NETWORK is dedicated to the shared stewardship of research data in Canada through:

- Developing a national research data culture
- Fostering a community of practice for research data
- Building national research data services and infrastructure

Launched in 2014 by the Canadian Association of Research Libraries, Portage works within the library community to coordinate expertise, services, and technology in research data management, seeking to collaborate with other research data management stakeholders.

Research data culture represents widely shared values and principles for digital data management.

A community of practice for research data consists of stakeholders working collaboratively to ensure data are accessible to address complex research issues.

Research data management activities span sectors, domains, and jurisdictions,
Portage Network Research Data Management Framework

**RDM Framework**
- **Data Planning and Production**: DMPs & Project-level research data management
- **Data Deposit**: Staging & institutional repositories
- **Data Curation**: Trustworthy data repositories
- **Data Preservation**: Institutional and domain repositories
- **Data Access**: Virtual research environments
- **Reuse & Integration**

**Portage Network of Experts**
- **Data Management Plans**
- **Data Curation**
- **Data Preservation**
- **Data Discovery**

**Training**

**RDM Tools & Platforms**
- **Data Management Plans**
- **File Transfer and Deposit**
- **Curation Toolkit and Preservation Pipeline**
- **Data Discovery and Data Integration Toolkit**

**Portage Tools**
- DMP Assistant
- DVN2Arch Islandora2Arch Globus FTP
- Open Science Framework FPR-PREFER, Archivematica Globus Publication
- Federated Search (UBC)
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.

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

• The University of Alberta Libraries installed an instance of DMP Online during the Spring of 2014 and agreed in December 2014 to host this service as a Portage platform.

• This partnership led to the formation of a DMP Expert Group in February 2015 to prepare a general data stewardship DMP template in the absence of a DMP requirement by the three primary research funding agencies in Canada.
Jeff Moon, Queen’s (Chair)
Eugene Barsky, British Columbia
Jay Brodeur, McMaster
John Brosz, Calgary
Jane Burpee, McGill
Talia Chung, Ottawa
James Doiron, Alberta
Carla Graebner, Simon Fraser
Alex Guindon, Concordia
Amber Leahey, Scholars Portal
Carol Perry, Guelph
Chuck Humphrey, Portage
DMP Assistant Development Process

1. Started in February 2015 by establishing **terms of reference** and setting **guiding principles**:

   - *Promote sound data stewardship practices*;

   - *Get researchers to respond to template questions by describing the strategies they will use to manage their data*;

   - *Ensure that the information provided can be re-used in other contexts*;

   - *Achieve a balance between seeking sufficient information about stewardship, while minimizing the burden on researchers.*
2. Reviewed international data stewardship best practices to identify the template sections and assigned sections to writing teams.

3. Spent two months writing, reviewing, and modifying questions within the sections of the template.

4. Pilot-tested the questions for clarity and understandability over two months with researchers and data specialists.

5. Translated the full template to French over a two-month period.

6. Developed a communication plan, brand, name, and launch during Open Access week in October 2015.
Continuing with...

• Institutional adoption and localization
  - Allow local administrators of DMP Assistant to brand their institution’s space by editing style sheets (CSS) to change the default colour scheme and by uploading institutional logos

• Help desk ticketing testing
  - Use an online ticketing system (OTRS) to track help requests and to build a database of responses. Examining a triage process to distribute help tickets among experts for answers

• Canadian Access Federation user authentication

• Partnering with SSHRC and CIHR in researcher training pilots

• Collaborating in DCC/CDL merged code: unified DMPonline

• International interest (e.g. Germany)
Develop RDM platforms and tools through partnerships and memoranda of understanding

- Current partnerships in place or nearly signed
  - DMP Assistant (UAL-CARL)
  - National Research Data Repository Prototype (CC-CARL)
  - Preservation Pipeline (CC-OCUL/SP-UTL-CARL)
  - Dataverse-to-Archivematica middleware (OCUL/SP-CARL)

- Forthcoming
  - PREFER: Policy Format Register
  - Federated data discovery system
  - Data curators toolkit
National Preservation Pipeline Service with Compute Canada and OCUL/Toronto (in progress)
CARL/Portage and CC: jointly manage the development through shared planning processes

CARL/Portage: research data management expertise and in-kind contributions

CC: project management and software development expertise and necessary computational resources

All derived software products will be released to the wider research community through an open source licence (except software components from other parties that are covered by their own licensing regime)

Non-exclusive relationship
Next Steps

Working with Steering Committee to further develop business model for ongoing operations, and ramping up training and communication.
Data Collection

What **types of data** will you collect, create, link to, acquire and/or record?

What **file formats** will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

What **conventions and procedures** will you use to structure, name and version-control your files to help you and others better understand how your data are organized?

Collecting information and ‘educating’ researchers
Acknowledgements and thanks to

Chuck Humphrey, Portage Director, CARL
Jeff Moon, Chair, Portage DMP Expert Group
Kathleen Shearer, Research Associate, CARL
Martha Whitehead, President, CARL

https://portagenetwork.ca/

Questions?