The Shared Stewardship of Research Data

Dalhousie University
June 5-6, 2017

Chuck Humphrey, Director of Portage
Context

• Changes in our research environment
  • Emerging Data Policies: Tri-Agency, Open Science, other funders, and publishers
  • Digital Research Infrastructure
    • DRI Summit on June 27, 2017
    • LCDRI report in July 2017
• Institutional responses these changes
  • National and local
1. Institutions: Institutional Strategy

• Each institution administering Tri-Agency funds is required to create an institutional research data management strategy. The strategy will outline how the institution will provide its researchers with an environment that enables and supports world class research data management practices.

• The strategy must be posted and publicly available on the institution’s website, with contact information to direct inquiries about the strategy.

Source: Jeremy Geelen (SSHRC) presented at CARA on May 8, 2017
2. Researchers: Data Management Plans

- Grant recipients are required to create data management plans (DMPs) for research projects supported wholly or in part by Tri-Agency funds. Grant recipients should submit these plans to their institution’s research office as a condition of the release of grant funds.

- For specific funding opportunities, the agencies may require DMPs to be submitted to the appropriate agency at time of application; in these cases, they may be considered in the adjudication process.
3. Researchers: Data Deposit

- For all research data and code that support journal publications, pre-prints and other research outputs that arise from agency-supported research, grant recipients are required to deposit these data and code in an appropriate public repository or other platform that will ensure safe storage, preservation, curation, and (if applicable) access to the data.

Source: Jeremy Geelen (SSHRC) presented at CARA on May 8, 2017
Digital research infrastructure

Adapted from CFI DRI submission
DRI Ecosystem
Core Components

Key Ingredients
- Infrastructure (HW & SW)
- HQP
- Operational policies & processes
- Funding

Key Outcomes
- Research
- Knowledge
- Commercializable products/outcomes

Source: LCDRI Data Management Working Group
Institutional Responses: national and local
Facilitating the development of partnerships and best practices to build a sustainable Canadian RDM ecosystem.

- RDM multi-sector stakeholders’ forum
- International RDM voice for Canada
- Advocacy for RDM stakeholders
- RDM strategic policy setting & best practices identification

Supporting the development and delivery of RDM services and resources to Canadian universities and partners.

- Leadership and coordination of library-based RDM initiatives
- RDM community of applied practice
- National RDM network of applied expertise
- RDM services & infrastructure platforms

Supported by: canarie

Supported by: CARL ABRC
The Canadian Association of Research Libraries (CARL) is the leadership organization of Canada’s 29 largest university libraries and two federal institutions (LAC & NRC Science Library). CARL members support Canada’s research community by:

- Working to improve access to knowledge;
- Improving services to students, faculty and researchers;
- Promoting effective and sustainable scholarly communication;
- Ensuring that Canada’s researchers have world class information management capacities;
- Sharing best practices and experiences,
- Advocating public policies that enable broad access to scholarly information.
The Canadian Association of Research Libraries (CARL) is the leadership organization of Canada’s 29 largest university libraries and two federal institutions (LAC & NRC Science Library). CARL members support Canada’s research community by:

- Working to improve access to knowledge;
- Improving services to students, faculty and researchers;
- Promoting effective and sustainable scholarly communication;
- Ensuring that Canada’s researchers have world class information management capacities;
- Sharing best practices and experiences,
- Advocating public policies that enable broad access to scholarly information.
The Goals of the Portage Network

• Foster a **community of practice** for research data management (RDM)
• Facilitate and provide leadership in the **development of RDM infrastructure**
• Engage and advocate for research data management with **stakeholder communities**
acceleration = \frac{\text{change in velocity}}{\text{time taken}}
Portage Principles

- Research data are a public good
- Intelligent access: openness, with respect for privacy
- Commitment to standards and interoperability
- Open source: Tools will be contributed back to the community
- Collaborative approaches: cost savings and shared expertise
- Inclusiveness: aim to serve all researchers and create a more level playing field
- 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
- Stewardship: a sense of responsibility for managing research data over the long term

December 2013
Networks of Expertise Model

Kirchner, et.al. (2015) examined the use of the centre of excellence model to provide digital information services to multiple institutions. They proposed a variation which they called “networks of expertise.”

This approach applies a community-building strategy and brings together experts with complementary skills for a limited time to address a particular issue. Furthermore, it keeps experts at local institutions by relying on an active network to address issues across a spectrum of institutions.

Build it and they will come
Build a community of practice and they will use it
How to Build a Community of Practice

• Identify the layers in which the community exists;
• Build the community within and across these layers;
• Coordinate activities that are happening parallel to one another (break down silos!);
• Develop partnerships with those who can contribute to the community;
• Nurture a common culture of service and mission;
• Offer expertise where it lacking;
• Establish relationships with other stakeholders and their communities;
• Provide appropriate governance to engage and steer the community; and
• Have a succession plan.
Layers of a Community of Practice

1. Researcher Layer
2. Institutional Layer
3. Network of Expertise Layer
4. Infrastructure Platform Layer
Layers of a Community of Practice

Strengthening relationships

- **Researcher Layer**: Researcher with subject / liaison librarian
- **Institutional Layer**: Subject / liaison librarian with Portage experts
- **Network of Expertise Layer**: Infrastructure providers with Portage experts
- **Infrastructure Platform Layer**: Infrastructure Platforms & Portage experts with librarians & researchers
Distribution of a Community of Practice

Canada’s four regional library consortia
Develop local partners

Libraries

Researchers

Graduate Studies

Individuals, Groups and Services

IT

Ethics

Research Services
Data stewardship

Who is responsible during and after a research project?
Portage RDM framework

Research Data Lifecycle

Portage Network of Expertise

RDM Tools & Platforms

Portage Tools

<table>
<thead>
<tr>
<th>Research Planning</th>
<th>Data Planning and Production</th>
<th>Data Deposit</th>
<th>Data Curation</th>
<th>Data Preservation</th>
<th>Data Access</th>
<th>Reuse &amp; Integration</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMPs &amp; Project-level research data management</td>
<td>Staging &amp; institutional repositories</td>
<td>Trustworthy data repositories</td>
<td>Institutional and domain repositories</td>
<td>Virtual research environments</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Management Plans</th>
<th>Data Curation</th>
<th>Data Preservation</th>
<th>Data Discovery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Intelligence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Management Plans</th>
<th>File Transfer and Deposit</th>
<th>Curation Toolkit and Preservation Pipeline</th>
<th>Data Discovery and Data Integration Toolkit</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMP Assistant</td>
<td>DVN2Arch Islandora2Arch Globus FTP</td>
<td>FPR-PREFER, Federated Research Data Repository, Dataverse North</td>
<td>Federated Search Open Collections UI (UBC), PIMS Jupyter Notebooks</td>
</tr>
</tbody>
</table>
Network of Expertise

- Six expert groups
  - Data Management Plans (DMPEG)
  - Curation (CEG)
  - Preservation (PEG)
  - Discovery (DEG)
  - Training (TEG)
  - Research Intelligence (RIEG)
- Five working groups
  - Data repositories & collection development
  - Metadata for discovery
  - Online training module on an introduction to RDM
  - Online training module on DMPs
  - CIHR Training Modules
- Dataverse North community of practice
Participation in Portage Groups

Number of Portage Volunteers by the Number of Groups in which They Participate

- 62 volunteers participate in one group
- 13 volunteers participate in two or more groups

Number of Portage Volunteers by Four Canadian Regions

- Atlantic: 7 volunteers
- Quebec: 11 volunteers
- Ontario: 32 volunteers
- West: 27 volunteers

Number of Institution Types with Volunteers Participating in Portage

- CARL Universities: 24 institutions
- Non-CARL Universities: 5 institutions
- Organizations: 4 institutions

Counting institutions

Counting people

28
Portage highlights 2015 to date

2015
- CARL Approval
- Director Starts
- DMP Assistant
- CC Partnership
- Website Launch
- CSA-Life Sciences
- SSHRC DMP
- CIHR Training

2016
- New Business Model
- New Governance Model
- OCUL MOU
- CAUL MOU
- LCDRI
- Dataverse North

2017

2018

Volunteers:
- 2015: 15
- 2016: 42
- 2017: 79
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 2015 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.
Portage service model properties

• Responsive: the service must be attentive to the needs of the data management community to ensure it is providing meaningful support;
• Adaptive: the service must be willing to make adjustments to meet changing needs in data management and to do this with the least amount of disruption;
• Integrative: the service must find ways of incorporating new aspects of data management into its operations;
• Collaborative: the service must be open to work with others who have contributions to make to data management;
• Well-governed: the service must be accountable to those it serves, to its partners, and to its funders; and
• Sustainable: the service needs to secure resources that allow its ongoing operation.
Federated service implementation

- is distributed across a defined community,
- consists of distinguishable service components defined within a foundational framework,
- operates in partnership with others to provide these service components,
- is managed through a secretariat that provides direction, coordination, and administration of service delivery,
- is sustained through a business model of in kind and cash contributions, and
- is governed by stakeholder, operational, and funder oversight.
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.

Sign up with DMP Assistant

Sign in and select a template under Organizations. The Portage template is the default.

Answer the questions that are relevant to your work. Guidance and examples are provided.

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 Lessons

Model of network collaboration and trust-building

- A host institution with many in-kind contributors
- Bilingual development – truly national platform
- Drawing from and contributing to international community (UK Digital Curation Centre and California Digital Library)
- Foundation for strong engagement with federal funding councils
  - “Portage has made the data management policy possible.” SSHRC Executive Director for Corporate Strategy and Performance
Data Repository Functional Framework

Microservice Layers

- Data discovery
- Data file transfer: preservation
- Data file sharing: access
- Data access controls
- Data preservation processing
- Data file transfer: deposit

Data Asset Management

Storage
Principles of Open Scholarly Infrastructure

• Governance
• Sustainability
• Insurance

“Above all, [infrastructure] is trusted and relied on by the broad community it serves. Trust must run strongly across each of the following areas: running the infrastructure (governance), funding it (sustainability), and preserving community ownership of it (insurance).”

Thank you

Chuck Humphrey
chuck.humphrey@ualberta.ca
780-492-9216