

Portage: Organizational Framework

April 7, 2015

Prepared by Martha Whitehead and Kathleen Shearer

In March 2014, the Canadian Association of Research Libraries (CARL) launched a project to develop a library-based research data management network in Canada. Project ARC, as it was called, had a one-year mandate to develop several aspects of the network, now named Portage.

The aim of the Portage network is to coordinate and expand existing expertise, services and infrastructure so that all academic researchers in Canada will have access to the support they need for research data management. The Portage network will have two major components:

Network of Expertise: Research Data Management (RDM) requires specialized knowledge and expertise, which many institutions and researchers do not have. The Portage network of expertise will provide access in both English and French to a comprehensive set of resources, tools and experts that will provide access to up-to-date, relevant and trusted information about RDM.

National Preservation and Discovery System: Portage has also been working on a project to connect the various infrastructure and service components needed for a national preservation and discovery network.

Principles

The principles that were identified at the outset of Project ARC will guide the development and ongoing operations of the Portage network:

- Data are a public good
- Intelligent access: openness, with respect for privacy
- Collaborative approaches: cost savings and sharing 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 our work is in keeping with international practices
- Respect for differences: flexibility to meet the needs of different regions, institutions, languages and disciplines
- Open source: Tools will be contributed back to the community whenever possible
- Stewardship: a sense of responsibility for managing research data over the long term.

Assumptions

1. The principles stated above will guide the development and operations of Portage.
2. Portage is one element in a larger digital infrastructure that supports RDM in Canada. It will function within this larger context and collaborate with other stakeholders to develop a sustainable and coherent national RDM environment.
3. Significant in-kind contributions will be provided by university libraries. As an extension of their operations, participating libraries will provide infrastructure and staff support for both the preservation and discovery system and the network of expertise.
4. We will develop partnerships to identify storage capacity for research data and to secure funding for the development of the network.
5. CARL will provide project management support during the first year of operations and staffing levels will increase over a three-year period.
6. CARL and its members will have governance roles in the Portage network.
7. Other institutions and research centres may contribute to Portage through an affiliation of their repositories with Portage's preservation system, by adhering to standards requirements, and by potentially providing expertise.

Operations

(More details about the current activities of Portage are available in the Portage [Workplans](#) that are in development)

Different operational approaches are required for the two components of Portage. However, it should be noted that there are numerous areas of overlap between the services provided. For example, many of the staff contributing to the preservation and discovery system will also serve in the network of expertise.

Network of Expertise

Managing research data can be very complex and support is often required at various stages throughout the data lifecycle—from preparation of data management plans, to documentation of data for access and preservation, to the re-use and analysis of datasets. While some researchers are becoming aware of data management and its numerous benefits, very few have the expertise and knowledge to address the issues on their own.

The Portage Network of Expertise will provide recommended resources, expert advice, and practical help to anyone working in a Canadian university or research institution wanting to store, manage, protect and share digital research data in both English and French. The primary target group will be front-line support staff in the libraries and data centres at Canadian universities, with the aim of building capacity and knowledge in those communities. Front line staff will then provide support to researchers in their communities. In some cases, such as when an institution has no such support, help will be provided directly to a research team or individual researcher.

Operations:

- Provide free access to RDM tools and resources and manage the Portage DMP Builder via the Portage website. These resources will be aimed at providing support for researchers in managing their data across the data lifecycle, as well as developing data management plans.
- Maintain and foster a community of practice in Canada for RDM. In order to build capacity across the country, the network will play a convening role by bringing people together to share best practices both virtually and in-person. This may also involve liaising with other stakeholders, such as research administrators, IT and other support staff in areas where we need to build capacity around RDM.
- Deliver at-cost training and networking events that build capacity in the library community and beyond. The training will involve the implementation of an online training course for graduate students and librarians, such as EDINA's MANTRA, as well as hosting events that adopt the "train the trainer" approach enabling those attending these training events to return to their institution and share their knowledge with others.
- Develop, identify and share best practice for RDM in several key areas: (1) privacy, security, and confidentiality, (2) skills and training, (3) data management plans, (4) metadata and data discovery, (5) data curation, (6) data access and dissemination, and (7) preservation.
- Contribute to the development and adoption of new and innovative standards, tools and practices via local, national and international venues including liaising with disciplinary projects and national/international organizations such as Research Data Canada, as well as the Research Data Alliance and CODATA.
- Develop metadata guidelines and procedures for the production of metadata that both make the data discoverable and independently understandable (an OAIS principle). The latter type of metadata is much more detailed than the former and may include documentation on the workflow of the research process.
- Develop data access guidelines and procedures that facilitate dissemination and reuse. Access involves a bundle of procedures around licensing, permissions, dissemination procedures, data formats, citation attribution, etc.
- Offer consultancy services to those institutions or projects that are seeking more specialized guidance and advice.

Service Model:

- The network of expertise will be managed by a central coordinator.
- RDM experts from institutions across Canada will provide expertise, training, and consulting services.
- Institutions will commit to releasing their participating staff member for a 20% commitment and will be reimbursed for that time at a standard rate from central Portage funding.
- It is likely that the network of expertise will launch with a relatively small group of experts (8-10) and this group will grow over time.

Preservation and Discovery System

The network will provide ingest and preservation services requiring a robust technical infrastructure consisting of software that supports the entire research data lifecycle (ingest, preservation, discovery, access, repurposing), data replication services, and networked data storage. This infrastructure will be highly distributed with local, regional and central nodes and will also be based on standards that ensure interoperability across nodes and data types.

In addition to ingest and preservation, a complementary set of services will support the discovery of data contained in data repositories across Canada. For this, metadata from repositories will be aggregated into an open registry through which discovery tools will be built to enable searching across data collections and repositories. Datasets from distributed repositories will be interoperable because they will map to common metadata standards allowing an appropriate level of integration.

Close collaboration with other partners and stakeholders is essential for the development and ongoing maintenance of this infrastructure. Both cash and in-kind resources will be needed to support one-time development and initial implementation, as well as ongoing operations.

Operations

- Provide ingest and repository services for datasets. The network will ingest research datasets and the appropriate metadata to ensure that datasets are interoperable, preserved, and discoverable. In addition, it will support ingest processes from external data repositories into the preservation environment.
- Provide preservation services that maintain the data over the long-term, ensuring that they remain understandable and reusable into the future. Services will include: the addition of descriptive and administrative metadata; standardizing data formats so they can be managed as a “collection,”; migrating data to new formats when their current formats become obsolete. Data replication in at least three separate locations across the network is also an important principle of the Portage model of preservation.
- Develop, implement and maintain an aggregated discovery tool that provides access to the Canadian research data in collaboration with other partners.
- Develop metadata guidelines and procedures. Research data management must produce metadata that both make the data discoverable and make the data independently understandable (an OAIS principle). The latter type of metadata is much more detailed than the former and may include workflow documentation.

Service Model

- One-time development, upgrading, and/or expansion of the technical infrastructure will be funded through a variety of financial mechanisms in collaboration with the various constellations of participants in the network. Portage will seek to obtain support from sources such as government and granting agencies and also rely on participant fees for specific initiatives.
- The ongoing technical infrastructure will be maintained in a distributed but collective manner that appropriately reflects the local, regional and national configuration of the preservation network. This is seen as a way of achieving long-term sustainability. Local infrastructure and services will primarily be the responsibility of the local institution. Resource and cost sharing mechanisms, that also include major stakeholders and

partners, will be developed to support multi-institutional, regional and/or national activities and various service nodes.

- Institutions lacking local resources and mechanisms/repositories for ingesting data into the network will have access to such services via other participants acting as host sites for others. Appropriate financial mechanisms will be developed as the needs and requirements of these institutions become clearer.

Staffing

It is anticipated that staffing will be a mix of part-time contributions from staff employed at institutions (with relevant expertise), secondments, consultants, and permanent employees. In addition, given the current spread of existing expertise in Canada, staff will be housed at different locations across the country, and have a variety of reporting relationships. The network's staffing will expand over three years to eventually include:

1. Network of Expertise (3 FTE)
 - a. Coordinator (1 FTE): maintains web-based resources, DMP Online, coordinates pool of experts, training and consultations.
 - b. Experts based at institutions contributing a certain portion of their time to the network of expertise (10 experts x 20% FTE = 2 FTE equivalent)
2. Preservation and Discovery Network (3 FTE)
 - a. Technical Committee (see description below)
 - b. Technical Coordinator (1 FTE)
 - i. Developers/IT specialists (.5 FTE equivalent)
 - ii. Data modeling/management specialist (1 FTE equivalent)
 - iii. In kind contribution from staff/experts at participating institutions
 - c. Funding for capacity development. .5 FTE equivalent funds set aside to support interns and sabbatical time to further develop a skilled workforce in the area of RDM. (.5 FTE)

3. Portage Secretariat (1 FTE)

Portage Director (1 FTE) + Support staff (in-kind, from organizational home)

(Note: In order to enable a smooth transition from Project ARC to the ongoing operations of the Portage network, CARL has budgeted to fund the Portage Director as a one-year project officer position in 2015.)

Budget

It is anticipated that the budget for Portage will increase each year for the first 3 year period, culminating in an annual budget of \$1 million in year 3. Please note, that this budget **does not** reflect the contributions of institutions, Compute Canada, and potential institutional in-kind support.

Budget organized according to line item expense	
Staff (6 FTE)*	\$600,000
Tools, technologies and infrastructure	\$300,000
Travel, Meetings, Workshops, Training, Translation, etc.	\$100,000
Total	\$1,000,000

*Note: CARL has budgeted to fund the Portage Director as a one-year project officer position in 2015. This position is not included in the budget figures above

Funding Model

We are proposing a membership and partnership model that will scale up to an annual budget of \$1 million by the 3rd year of operations. The model will utilize a flexible cost structure that builds on regional/consortial fees, as well as funding and in-kind support from partner organization. This model will ensure the long-term sustainability of the network, encourage membership involvement, and allow for major participants to have appropriate oversight of the network, while still enabling the contribution of external partners and funders.

The membership fees will reflect the size of participating institutions, specific requirements as well as the contributions of each member organization. Institutions (or their consortia) will fund the ongoing maintenance of the network, while external funding will be directed towards the development of technologies, infrastructure and services.

Annual fees for individual institutions will range from \$5,000 to \$50,000 and consortial fee structures will be developed based on size and contributions.

Governance

The network will gradually develop over the next several years and the governance structures and funding models will evolve as the network matures. At this time, the proposed governance structure is as follows:

Steering Committee

A Portage Steering Committee will be appointed by the CARL Board to represent the stakeholders involved in the Portage Network. It is envisaged that this committee will have majority representation from the major institutions and organizations contributing to the network, as well as of the broader stakeholder community (e.g. non-CARL members, researchers, research administrators, and funding agencies).

Characteristics include:

- Representation from the stakeholder community, participating members, and partner institutions. We will look to strike an appropriate balance between Portage's shareholders (i.e. Portage partners and members) and stakeholders (other organizations that have an interest in Portage's work and success). This Committee consisting of 10 to 12 members
- Accountability for network governance and budget
- Establishment of operating principles, policies and procedures
- Identify priorities for investment and areas of development

Technical Committee

The Technical Committee will advise the Steering Committee on matters relating to Portage technical requirements, standards, and network capacity. The Committee will consist of individuals from participating institutions and will nominate a chair that reports to the Steering Committee.

Members and Partners

Organizational participation will be determined based on their level and type of contribution to the project. The costs for membership will vary depending on the size of the organization and levels of in-kind contribution. There will be two levels of membership and one level of partnership.

Contributing Members are institutions or consortia that provide, set up, and maintain ongoing infrastructure that contribute to the core operations of Portage.

- Contributing members must express a long-term commitment to the Network, with the understanding that a minimum commitment is three years.
- Contributing members will sign a memorandum of agreement that includes a statement of their contribution to the operation of the Network.

Supporting Members are institutions or consortia that adhere to one of the following criteria:

- Provide an expert based at the institution, contributing a certain portion of their time to the network of expertise, and are reimbursed for that time.
- Provide in-kind support through staff participation in other aspects of the network of expertise.
- Provide support in the form of software development or value-added services for a fixed period of time.
- Members will join Portage via a written agreement defining the nature of their relationship with the network.

Partners are organizations that provide significant funding or in-kind contribution to the network. These contributions include:

- Funds towards the development or maintenance of infrastructure, technologies and services for the network.
- In-kind storage and infrastructure support for the network.
- Office space and administrative support for the network.

Transition

Project ARC was a one-year project that aimed to lay the foundation for a library-based research data management network in Canada. As Project ARC transitions into the Portage network, there is already significant momentum in several areas. The Project ARC Working Group will be disbanded, but task-specific groups such as the Portage DMP Experts Group and the preservation and discovery working group will continue to make progress on various parts of network. The intention is to continue with those directions, while establishing the governance structure and funding streams. In order to support these ongoing activities and to expand Portage services to full operations over a three year period, CARL will hire a full-time director for Portage.

Research Data Canada (funded by CANARIE) has already invested in several aspects of the Portage network including providing funding for the Project Coordinator to work on compiling resources for the Network of Expertise. RDC has also paid for some of the development work to integrate Islandora and Archivematica platforms at SFU; and it is also contributing funds towards project management for the installation of the Globus software, which will support data management and replication at Compute Canada. It is anticipated that Research Data Canada and CANARIE will remain important partners in Portage and possibly enable funding of various activities.

During the three year transition period, CARL will continue discussions with Compute Canada, Research Data Canada and CANARIE, and at the appropriate time, seek to develop a memorandum of agreement to ensure that the interests of all parties are understood and achievable.