



CHAPTER 3*

Collaborative Research Data Curation Services A View from Canada

*Eugene Barsky, Larry Laliberté, Amber
Leahey, and Leanne Trimble*

In Canada, as in many developed countries, requirements for data management are being established across a wide range of scholarly disciplines. Barriers to data management and sharing are being addressed through the recommendation and use of community standards such as research data management plans (DMPs). Canada's federal granting agencies—known as the “Tri-Agencies,” consisting of the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council of Canada (SSHRC)—recently released a draft statement on digital data management.¹ Through this statement, the Tri-Agencies actively encourage research institutions to provide their researchers with an environment that enables robust stewardship and curation practices and to deliver support for the management and deposit of research data in secure, curated, and accessible repositories.

There are several library-led collaborative initiatives currently underway that aim to develop interoperable and sustainable data curation services in Canada in anticipation of future government requirements for data management. These initiatives, in combination with existing local expertise, are directly contributing to the capacity for research data management in Canadian universities. This chapter

* This work is licensed under a Creative Commons Attribution 4.0 License, CC BY (<https://creativecommons.org/licenses/by/4.0/>).

provides a brief history and overview of initiatives related to the coordination of data curation and preservation services at university libraries in Canada. Case studies from the Ontario Council of University Libraries (OCUL), the University of British Columbia Library (UBC), and the University of Alberta Libraries (UAL) are presented, with a focus on the library as a central facilitator of data curation and preservation. Some considerations about the financial and consortial business models are discussed. Finally, these efforts are placed in the context of Canada's overarching infrastructure initiative, the Canadian Association of Research Libraries (CARL) "Portage" project, which aims to develop a robust, collaborative national infrastructure network for Canadian research data.

Canadian Academic Library Involvement in Research Data Management

Canada, like the United States, lacks a centralized data-archiving service. National data archives, like national libraries, provide government-supported services and expert staff to ensure that information produced within a country is permanently preserved. To date, there have been several attempts to establish a national data archive, but none have been able to secure adequate support or the funding required for its establishment.² Centralization tends to be challenging in a country that has a relatively small and geographically dispersed population characterized by regionalism. Nevertheless, libraries have been strong advocates for improved access to data in Canada. For example, the Canadian Association of Public Data Users (CAPDU, <http://www.capdu.ca/>) is a library-based organization whose members advocate for improved access to data in Canada. The Canadian Association of Research Libraries (CARL, <http://www.carl-abrc.ca/about-carl>) also has advocacy as part of its mandate and is involved in research data management activities. The efforts of Canadian academic librarians have seen success in strengthening the data collections available to researchers for secondary use.

The Data Liberation Initiative (DLI, <http://www.statcan.gc.ca/eng/dli/dli>), a subscription-based service providing access to Statistics Canada data, is an excellent early example of Canadian academic libraries collaborating on data management. The DLI program began in 1996 as a result of consultations between Statistics Canada, the Canadian Association of Research Libraries (CARL), and the Humanities and Social Sciences Federation of Canada.³ The founding of the DLI was a response to both the high costs of Statistics Canada's public microdata files (which, due to budget cuts in the 1980s, were priced on a full cost-recovery

basis and were out of reach of all the most well-funded researchers) and the lack of data infrastructure at Canadian universities to provide access to these data.⁴ The sheer size of the DLI collection, including thousands of data files for hundreds of survey series, and the demands from researchers for this data, has directly contributed to the growth of library data infrastructure to manage and preserve access to this data. When the DLI was formed, there was little expertise in many libraries to support data services; however, because Statistics Canada required a point of contact within the library who would be responsible for distributing data to end users, libraries quickly developed staff expertise through DLI training activities.⁵ In addition, the DLI program prompted consortial initiatives to expand the available technical infrastructure. For example, in Ontario the development of <odesi> (<http://odesi.ca>) provided a centralized storage infrastructure and an innovative Web-based data access platform.

Some disciplines, particularly in the sciences, have developed a culture of data sharing through disciplinary repositories. In Canada, examples of domain repositories include the Polar Data Catalogue (a project of the Canadian Cryospheric Information Network, CCIN), the Canadian Astronomy Data Centre (an initiative of the Canadian Advanced Network for Astronomical Research, CANFAR), and CBRAIN (an initiative of the McGill Centre for Integrative Neuroscience, MCIN). Many disciplines, however, do not have these kinds of coordinated resources to turn to. Therefore a natural role for academic libraries is to develop institution-based data repositories and catalogues for disseminating and archiving data, particularly data sets that fall within the “long tail” of research data, meaning the large number of relatively small datasets that are produced in a wide range of disciplines.⁶ Long-tail data sets have a great deal of diversity and can have high curation requirements. Libraries, with their expertise in preservation of research output (e.g., through institutional repositories) as well as their history of engagement in data management and dissemination activities, are well-equipped to take on these challenges, given sufficient resources.

The federal government has been consulting with various research communities, including libraries and archives, about the benefits and challenges of research data management for some time. In 2005, the Canadian government released the report of the *National Consultation on Access to Scientific Research Data* (NCASR), the cumulative work of an expert task force of more than seventy leaders Canada-wide from research, administration, and libraries, among other areas.⁷ The list of recommendations included the development of a national steering body to coordinate data management and project funding across sectors in Canada; however, the approach ultimately failed to gain support politically.⁸ In 2008, a new group was formed, the Research Data Strategy Working Group (RDSWG), that sought ways to move forward on the NCASR recommendations. In 2011, CARL and the RDSWG held a Research Data Summit, which resulted in the formation of Research Data Canada (RDC) in 2012.⁹ RDC has facilitated a range of

committees and technical projects and partnered with other organizations internationally to advance research data infrastructure and expertise.

CARL has been an active participant in many of these important national discussions. In an effort to improve library preparedness for research data support services, it ran an extremely popular research data management course for libraries in early 2013.* Building on the momentum generated by the course, a forum was established for ongoing dialogue around related activities in Canada, known as the Canadian Community of Practice for Research Data Management (RDM) in Libraries (<https://cancoprdrdm.wordpress.com/>). CARL has recognized that one of the ways forward for the library community is to establish more formal relationships with those organizations that provide Canada's research computing infrastructure, such as CANARIE (network infrastructure), Compute Canada (high performance computing), CUCCIO (chief information officers at Canada's universities), and the National Science Library (formerly known as CISTI, and the home of DataCite Canada).

Today academic libraries across Canada are putting plans in place to actively deliver a range of research data management services to their communities.¹⁰ Infrastructure remains a central challenge, but one that is being addressed through collaborations between libraries and with the broader research community, through current CARL initiatives such as Portage. The Portage initiative brings together many stakeholders in a collaborative effort to develop distributed infrastructure, in contrast to earlier unsuccessful attempts to create a single national institution to manage data preservation. This bottom-up approach may be the key to success in the Canadian context.

Overview of Case Studies

The authors of this chapter work at institutions across Canada that each has a unique approach to offering research data management services. Canada's small and spatially distributed population makes effective organization on a national level challenging. Canadian academic libraries tend to work together primarily within the context of regional consortia. In this chapter we will use several examples to illustrate the Canadian context. This chapter is not intended as a comprehensive description of all of the important research data management services undertaken at Canadian libraries, yet the case studies presented in this paper will show a good cross section of the kinds of research data management activities underway, ranging from libraries independently providing local services to comprehensive regional and national collaborations.

* The outline from this course is available online as Canadian Association of Research Libraries, "Data Management Workshop," accessed August 3, 2016, <http://www.carl-abrc.ca/strengthening-capacity/workshops-and-training/data-management-workshop/>.

Local Services: University of Alberta Libraries

The University of Alberta Libraries (UAL) has a long history of providing data services. In 1977, the precursor to the Data Library was established by data librarian Chuck Humphrey in University Computing Services (UCS), which ran a facility for data deposit and retrieval. The early Data Library started as a database registry of data sets generated by university researchers. By 1980 the database had grown into a full data library comprising local research data, such as the Edmonton Area Survey, as well as data obtained, through mediated access, from large data archives such as the ICPSR and The Roper Center. In 1992 the Data Library and its staff, a coordinator and a data librarian, became part of the libraries' Humanities and Social Sciences unit. The Data Library staff provided a full complement of research data support services, including data acquisition and cataloging, assistance with data analysis, instruction related to data, and the provision of data archiving services to university researchers. With the formation of the library's Digital Initiatives (DI) unit in 2012, the Data Library and its staff became part of a larger unit with a renewed focus on the development of new RDM services (<http://guides.library.ualberta.ca/data>).

Since 2014, a working group for Research Data Management Services (RDMS) has been coordinating services for the broader University of Alberta Libraries. The RDMS working group consists of ten members from various campus subject libraries, including health, sciences, and the humanities. The mandate of the working group is to develop an effective communication and outreach strategy for liaison librarians around research data management. To facilitate this role, the working group consults with librarians in order to provide them with the resources they need to provide information to their faculty in areas related to research data management. These resources include the collection of RDMS user stories reflecting these services and the development of a librarians' tool kit, which includes links to informational and educational resources and slide templates that can be modified and tailored to various teaching settings and levels.

One of the most prominent promotions of library services and training opportunities for researchers on the University of Alberta campus is the annual Research Data Management Week, which debuted as the Campus Data Summit in 2012. The week, also coordinated by the RDMS working group, is comprised of a mixture of keynotes, presentations, and workshops. The event is well attended, with over 200 attendees in 2015, and continues to thrive. In 2015, Compute Canada (<https://www.computecanada.ca/about/>) became heavily involved by offering a concurrent stream of workshops in order to introduce faculty to

Compute Canada's advanced research computing (ARC) systems, storage, and software, which provide services and infrastructure for Canadian researchers and their collaborators. The week also offers an opportunity to roll out new library services to a wide audience.

In 2014, the University of Alberta Libraries launched a Dataverse instance (<https://dataverse.library.ualberta.ca/dvn/>) to serve as an optional research data repository for the campus. Since the launch there have been many Dataverse workshops and one-off sessions for faculty and students; promotional slides and quick reference material have been added to the liaison librarian tool kit. As of March 2016, the UAL Dataverse contains thirty-four published Dataverses with 234 studies, 2,541 files and 1,986 downloads. There are also 115 unpublished Dataverses (many of which are ongoing projects).

Since 2015, the library has sponsored a data purchase program, noting that while open data is becoming more widely available, there are still many cases where data is available only commercially. Therefore, the libraries piloted a demand-driven data purchase program with the primary goal of purchasing data to better support University of Alberta researchers. Once the data is purchased, it is immediately made available to the researcher, and when the project is completed the data is added to Dataverse, provided the licensing allows for open distribution, for use by other interested campus researchers. If the licensing is restrictive, the files are still added to Dataverse for discoverability; however, access is mediated.

Finally, the Education and Research Archive (ERA), the University of Alberta's institutional repository, was developed and supported by the University of Alberta Libraries. ERA's open-access content includes the intellectual output of the university. In October 2015, all of ERA's content was migrated to a new Hydra-based digital asset management system (DAMS) environment. The new platform, called HydraNorth, is the first phase for consolidating all the diverse digital assets managed by the library. It currently harvests metadata from the Dataverse instance so that data sets can be discovered when users search ERA; then users are linked back to the data files in Dataverse via their persistent DOIs.

The UAL is on the leading edge of research data management services in Canadian academic libraries and serves as an excellent example of what can be achieved at universities with reasonable staffing and infrastructure funding. However, many Canadian universities may not have the resources to undertake these activities alone, and one solution is to seek opportunities to collaborate.

Informal Regional Consortia: University of British Columbia Library

The University of British Columbia (UBC) Library is one of the largest university libraries in Canada and has been conducting ad-hoc research data management activities since the early 1970s. UBC Library's Abacus data repository (<http://dvn.library.ubc.ca/dvn/>) has, over the last fifteen years, moved from tape to custom database to a more complex data management system. In 2008, DSpace (version 1.5) was installed to run Abacus and replaced a home-grown system based on PHP and MySQL. As its input format was metadata-agnostic (using the Dublin Core metadata standard), it was suitable for the migration of UBC's licensed data sets, and the metadata management was the best available at the time of its adoption. Over time, the data needs of faculty and students increased dramatically. Data sets became larger and more complex. For example, geospatial data has gained wide use among research fields not normally associated with the use of data or geospatial imagery. The open-source software DSpace does not provide automatic version control, embedded data integrity checks, or granular access to data and data analysis in a web browser. As a result, the decision was made to upgrade UBC Abacus to a more data-user-friendly system, another open-source data repository solution, Dataverse.

Willing to assist smaller regional schools, in 2008, UBC entered into an arrangement to make the Abacus data repository available to other universities in the province. At the time of writing, four major university research libraries in British Columbia (Simon Fraser University, University of Victoria, University of Northern British Columbia, and University of British Columbia) are using the UBC instance of Dataverse, primarily as a licensed data repository. Using EZ-proxy for access control, data is provided to users from each institution according to their data licenses. Moreover, the UBC Abacus Dataverse has expanded to allow researchers from the universities to submit their open research data. Currently, UBC Abacus has more than 30,000 managed data files, with more than 10TB of managed data. The researcher-submitted data collection is approximately 10 percent of all data files but is steadily growing.

A UBC Library research data team provides basic and advanced Dataverse training to groups, departments, and labs on UBC campus as well as its partners in other university libraries and research institutes. After training, the goal is for these groups to manage their own data within the appropriate Dataverses assigned to them. The UBC team assumes responsibility for the entire Dataverse instance; however, individual researchers, labs, and libraries are trained and assigned to be the data curators for their own data sets.

Formal Regional Consortia: The Ontario Council of University Libraries

In Ontario, several universities have a long history of providing data archiving services.* The Carleton University Social Science Data Archive began in 1965 and was housed in the Sociology and Anthropology Department until around 1994, when it moved to the MacOdrum Library and become known as the Data Centre (now Data Services, <https://library.carleton.ca/contact/service-points/data-services>). The University of Western Ontario (now Western University) launched its Data Resources Library in the late 1970s (now known as the Map and Data Centre, <https://www.lib.uwo.ca/madgic/>), which worked with the Social Science Computing Laboratory to disseminate and archive several faculty research projects. The University of Toronto established its Data Library in 1988 (now the Map and Data Library, <http://mdl.library.utoronto.ca/>), with services that included the acquisition and preservation of data sets produced by University of Toronto researchers. By the late 1990s, as was happening across Canada after the initiation of the DLI program, additional universities in Ontario began to develop data expertise and to offer data support services to their communities.¹¹

In Ontario, there are twenty-one universities, which vary widely in size, focus, and available resources. Since the 1960s, the libraries at these twenty-one universities had been collaborating through the Ontario Council of University Libraries (OCUL). In its early years, OCUL was involved in traditional library services such as consortial licensing of journals and facilitating effective resource sharing. In 2002 OCUL formed Scholars Portal (<http://www.scholarsportal.info/>), a shared technology infrastructure that hosts and provides access to OCUL's growing digital collections. As data services came to greater prominence, Ontario libraries saw an opportunity to collaborate under the OCUL umbrella in order to improve services, reduce duplication of effort, and better manage limited resources. Therefore, over the last decade, OCUL has undertaken several successful data infrastructure projects, including the development of <odesi>, a social science data portal, and Scholars GeoPortal (<http://geo.scholarsportal.info>), a geospatial data portal. While each of these does contain some research data, <odesi> and Scholars GeoPortal are intended as curated collections of "published" data sets from authoritative sources such as government statistical agencies and as such are not conducive to the widespread inclusion of member libraries' institutional research data outputs. These systems are also primarily focused on discovery and access rather than long-term preservation.¹²

* Canadian university data services are listed in this chronology (a work in progress) developed by members of the International Association for Social Science Information Services and Technology (IASSIST): "Chronology of Data Libraries and Data Centres," accessed August 3, 2016, https://docs.google.com/spreadsheets/d/1qmC_z50UDHh-3Jwdlu6wGrzt1xjdfumBaad6P6cUpBdY/edit#gid=0.

For this reason, other solutions were needed in Canada to address the growing demand for library research data repositories, and in 2011, Scholars Portal installed an instance of the Dataverse open-source software and offered it to the OCUL community as a pilot program. The pilot was intended to address a community-identified need for an Ontario-based repository service that would allow for easy-to-use, Web-based self-deposit by researchers. Dataverse was chosen for the pilot due to its support for research data, including the Data Documentation Initiative (DDI) metadata built in. Scholars Portal staff developed some documentation and training materials to inform and train staff at OCUL libraries about the benefits of incorporating Dataverse into the suite of services offered for data management and deposit of research data. As a result, the Scholars Portal Dataverse instance has allowed some OCUL libraries to launch research data management services without needing to have the technical infrastructure and staffing to support repositories of their own. Models for the service vary from library to library, ranging from self-serve deposit to library-mediated curation. Some examples of OCUL institutions that have launched research data management services based upon the Dataverse platform are the University of Guelph and Queen's University.¹³ Due to the uptake of Dataverse within OCUL, the successful pilot became a core Scholars Portal service in 2012. Today, support for the use of Dataverse is largely provided by local library staff and is independent of the infrastructure hosted and supported by Scholars Portal.

In Ontario, several libraries have been offering longstanding RDM services, while others have recently embarked upon new RDM initiatives or are still in the planning phases. There is no doubt that this is a strategic area for most academic libraries, but it is unclear how RDM services will be funded at a time when budgets are very tight and researcher demand is in its infancy (with Canadian funder requirements still in flux). A community of librarians interested in research data management has begun to emerge, with the creation of an OCUL-wide Listserv to discuss topics of interest and an RDM theme for the 2015 Scholars Portal Day (<http://www.ocul.on.ca/node/4479>). Continued collaboration through the OCUL consortium will likely be extremely important to the success of emerging RDM services in Ontario libraries.

Data Repository Services in Canadian Libraries

There are many factors that libraries must consider when selecting software to form the basis for a research data repository. A suite of software is needed that can support access and discovery as well as long-term preservation. Access and discovery are facilitated through support for established metadata standards and

harvesting protocols, granular search tools, and data exploration tools. Data preservation involves the ability to manage data identification (through persistent identifiers), integrity, sustainability, and authenticity.

Discovery and Access Platforms

As we saw in the previous section, Dataverse has been the data repository software of choice for all of our example institutions. Dataverse (<http://dataverse.org>), developed by Harvard's Institute for Quantitative Social Science, is open-source software that allows researchers to share, cite, preserve, discover, and analyze research data.¹⁴ Its open-source nature means that an institution or group of institutions can host its own instance of the Dataverse software and offer a customized solution tailored to its own community. This is an important factor in Canada, where many universities prefer to store data on local servers hosted within the country. A local installation also provides the opportunity for local branding and for offering custom training resources to users.

Dataverse is designed as a self-deposit platform, organized into Dataverse networks, where individual researchers, research teams, and institutes can create their own account and deposit their own data into "Dataverses" that are part of a bigger "network." It is also possible for university libraries or other data custodians to curate contributions and manage the data submission process on behalf of researchers. In this sense, Dataverse is very flexible. For example, in the University of Alberta Libraries' Dataverse, the entire network is devoted to research data from one institution, and an individual Dataverse is created for each research project being deposited. In British Columbia, the Abacus Dataverse Network focuses on library-curated Dataverses for each participating institution. In Ontario, the Scholars Portal Dataverse (<https://dataverse.scholarsportal.info/>) is completely open-ended, with some institutions hosting a library-curated Dataverse within the network, in addition to researcher-created Dataverses. Local branding is possible for both the network and each individual Dataverse contained within it.

Dataverse also provides data analysis functionality in the browser; users do not necessarily need to download the data files to interact with them. Tabular data files that are uploaded to the system can be further analyzed in the integrated web-based data analysis and visualization tool. Offering some data visualization and analysis within the Dataverse tool eliminates the need for desktop software to perform similar tasks and adds to the interactiveness of the data, potentially broadening the audience and range of users. Moreover, the Universal Numeric Fingerprint (UNF) feature in Dataverse works to enhance the reproducibility of science. A UNF "is a unique signature of the semantic content of a digital object. It is not simply a checksum of a binary data file. Instead, the UNF algorithm approximates and normalizes the data stored within. A cryptographic hash of that

normalized (or canonicalized) representation is then computed.”¹⁵ This means that same data object stored in, say, SPSS and Stata, will have the same UNF. And if the same analysis was used on the same data set, the UNF should be the same. Moreover, specific analyses done in Dataverse are given a special citation that mentions the analysis performed.

Dataverse is also easy to integrate with other library resources for improved discovery. For instance, since all partners with UBC Abacus Dataverse are using ProQuest’s Summon as a discovery search engine for their libraries, the corresponding Dataverses are exposed via OAI protocol to their Summon engines. Each OAI feed includes all research data for the partner institutions and appropriate licensed data for that institution.* Improved discovery (especially when assigning DOIs for research data sets) means that curated data could be easily accessed and reused by researchers (e.g., in ORCID, Google, Datacite, VIVO, Crossref, and other services), thereby enhancing citations and improving research metrics for individuals and institutions.

Dataverse has proven to be a flexible platform that can support many models for library RDM services. It offers a range of features that may improve data discoverability and access. It also does a good job of managing data files from a preservation perspective, such as managing versions, conducting checksums to maintain data integrity, and supporting persistent identifiers, such as handles and DOIs. Dataverse is capable of normalizing tabular data files into an ASCII text format with a companion DDI metadata record, which is considered a best practice for long-term preservation.¹⁶ However, Dataverse is not a fully featured digital preservation system. It is format-agnostic and will accept deposit of all file types (not just tabular data), but currently it does not support normalization or metadata extraction from nontabular data files. The library community is in need of a robust long-term preservation solution that can manage a larger range of file formats and establish normalization and migration best practices for them. This preservation system would be used in conjunction with the established Dataverse service.

Long-Term Preservation

Digital preservation activities are designed to secure the long-term future of digital information resources. A successful digital preservation strategy must account for and mitigate the impact of various threats to the accessibility and usability of digital materials over time. Common challenges include software, hardware, and media format obsolescence; hardware failure; and natural disasters, among many

* An example of an OAI feed is available as University of British Columbia Libraries, Summon search result for “DBID: BAXLO,” accessed August 3, 2016, <http://ubc.summon.serialssolutions.com/#!/search?ho=t&q=DBID:%20BAXLO&l=en>.

others. Mitigation strategies may include storage refresh, file format normalization (to open formats), software and hardware migration, data replication, and emulation.¹⁷ Preservation metadata about the original data file, its provenance, and the preservation actions taken on the data (such as data validation or normalization to another file format) are required and therefore desired functionality for long-term preservation systems. Ensuring that that preservation activities are documented and well understood is crucial to ensuring long-term viability of data.

One software tool that has emerged in recent years to support digital preservation is Archivematica (<https://www.archivematica.org/en/>). Archivematica is an open-source software package developed by Artefactual Systems. It takes a “micro-services” approach to preservation, offering an integrated suite of free and open-source tools that allow users to process digital objects by applying format-specific preservation policies in order to prepare objects for archiving and dissemination.¹⁸ Archivematica is essentially a pipeline of services that moves digital information packages through a series of file-system directories. Together these steps process digital objects from ingest to dissemination, resulting in the production of an Archival Information Package (AIP), a Dissemination Information Package (DIP), or both. An AIP is a container holding all the information necessary for long-term preservation of the file; it typically includes the original files and existing metadata, any normalized files created by Archivematica processes, and a preservation metadata file generated by Archivematica. This preservation metadata follows the PREMIS preservation metadata standard, encoded in METS (Metadata Encoding and Transmission Standard) format.* In contrast, a DIP is a package delivered to an access platform and contains the data and metadata needed for discovery. Once created, AIPs and DIPs exist independently from Archivematica and are typically stored in a digital asset management system (DAMS) or other secure storage location. Used together, the Archivematica micro-services make it possible to fully implement the Open Archival Information System (OAIS) reference model, a framework for understanding the responsibilities and processes involved in the design of a preservation system.¹⁹

Digital preservation can be applied to all forms of digital information, including research data. Some work has been done to determine optimal file formats for statistical, geospatial, and other research data,²⁰ and Archivematica is equipped to handle relevant normalizations for a wide range of file formats, including images, spreadsheets, documents, and many other files. Archivematica maintains a Format Policy Registry (based on formats documented in the PRO-

* For more information on PREMIS, see Priscilla Caplan, *Understanding PREMIS* (Washington, DC: Library of Congress, 2009), <http://www.loc.gov/standards/premis/understanding-premis.pdf>. The Library of Congress also provides information on using PREMIS with METS: “Using PREMIS with METS,” Library of Congress, October 15, 2010, <http://www.loc.gov/standards/premis/premis-mets.html>.

NOM format registry), which documents the actions the software can apply to specific file formats.* For example, JPGs are identified as “jpeg image format” and are normalized to TIFF. Archivematica will store the original JPG and the derived TIFF in the AIP, referencing the original and converted file names and locations, and will use the PREMIS vocabulary to describe this normalization in the METS file. There are still many specialized file formats for which normalization tools do not exist and that are not yet described in registries like PRONOM. However, as additional information is acquired and new tools developed, Archivematica is well equipped to integrate new policies. This is an area being explored by libraries within Canada (as part of the Portage project) and elsewhere.²¹

Our institutions have varying degrees of engagement with digital preservation using tools like Archivematica. University of British Columbia (UBC) has engaged Archivematica as its digital preservation system since 2014, hosting the software in UBC’s EduCloud cloud-computing service.²² Importantly for British Columbia, this service meets provincial privacy requirements under the Freedom of Information and Protection of Privacy Act. In addition, EduCloud offers the benefits of a virtual server hosting service, such as server consolidation, resource pooling, high service availability, and regular backups. At this time, three (out of four) UBC Library digital repositories are connected to Archivematica for digital preservation: DSpace (UBC cIRcle), CONTENTdm, and AtoM.

OCUL also has significant experience with digital preservation, having received Trustworthy Digital Repository Certification (TRAC) for its electronic journal repository in 2013.²³ Like UBC, OCUL has also been developing a private cloud storage service, known as the Ontario Library Research Cloud (OLRC, <http://www.ocul.on.ca/node/2126>), being rolled out in late 2015. While not actively using Archivematica at this time, OCUL is undertaking several initiatives to add new functionality to Archivematica, in collaboration with Artefactual Systems, in order to assess the opportunity to incorporate it as a service for OCUL libraries. Scholars Portal’s in-house solution for preservation of electronic journal content is not designed for self-serve access by individual OCUL member institutions for the preservation of their own local content (e.g., digitized collections). Scholars Portal staffing is not sufficient to manage local preservation activities on behalf of member institutions, nor is this considered desirable. Instead, Scholars Portal sees the combination of Archivematica and OLRC as a potential self-serve Web-based solution for supporting local preservation requirements. To this end, Scholars Portal is involved with integrating Archivematica with OpenStack Swift storage, the technology upon which OLRC is based. In addition to storage integration, a number of libraries across Canada (including UBC, University

* Archivematica’s Format Policy Registry is described at <https://www.archivematica.org/en/docs/fpr/>, and the PRONOM registry at <https://www.nationalarchives.gov.uk/PRONOM/Default.aspx>.

of Alberta, and OCUL) are currently engaged under the Portage umbrella in a project to integrate Dataverse and Archivematica. When completed, this project will provide new opportunities for integrating good preservation practices into library research data repository workflows.

Operational Costs of Data Repository Services

The costs of operating of a data repository can vary widely depending on the level of services provided, but in all cases there will be technology (hardware, software, and storage) and staffing costs. The use of open-source software like Dataverse eliminates the cost of software licensing fees; however, it can become necessary to invest software development resources in order to implement desired features in the software, as we will describe with an example in the Future Directions section.

The University of Alberta has taken on the operational costs of running Dataverse locally. The service is directly supported by four staff members, in addition to their other duties, who not only manage the technical infrastructure but also provide data curation services to researchers, including one-to-one consultation sessions on metadata creation, file permissions, the value of data sharing, and the importance of data attribution. Most of the technical implementation work was up front to get the service out the door, and episodic during software updates. Once the service was up and running, any operational costs related to its promotion (presentations/workshops) have been spread out to all librarians with portfolios relating to RDMS.

When university libraries work together in consortia (as it is frequently done in collections management), it is possible to share costs and reduce duplication of effort. British Columbia's Abacus Dataverse Network is an example of a collaborative service that is still in its early days. Since the collaborative work led by the University of British Columbia Library does not function as a formal consortium, it has been challenging to formalize a cost-sharing model; such models are not common in the province of British Columbia's academic libraries to date. However, it is not sustainable for UBC Library to continue paying for both the technical and human side of the operation, which in 2015 ran around \$250,000 CAD.

In Ontario, where there is a long-standing history of cost sharing through formal consortia, the growing pains are fewer. OCUL has an established model where new services are proposed to the governing group composed of the library directors for each member university. If the proposal is feasible and fits within OCUL's strategic directions, then OCUL will typically seek grant funding to

cover any one-time project costs such as development of a new software platform. When the service nears its launch date, the OCUL directors review a sustainability plan and make a decision as to whether to include this new service in the suite of “core services.” Once a service is considered a core service, it is integrated into the OCUL costing model, which calculates the contribution each member institution makes towards the OCUL annual budget.

In the case of the Scholars Portal Dataverse service, the model has been somewhat less formal. Because there was no new software to develop, grant funding was not sought. Also, the service was initially launched as a pilot with Scholars Portal assuming the up-front hardware costs, which were minimal at that time as the service was being used primarily for testing. To date, Scholars Portal staff have taken on a primarily technical support role for its Dataverse instance; users in need of more in-depth support for their data management activities are referred to designated staff at their home institution’s library. This differentiation of roles allows for technology-related costs to be centralized and shared among the OCUL consortium members, while research support costs are incurred by individual libraries as local expertise is needed. Today the OCUL Dataverse service is no longer considered a pilot, but the overall use of the service is still in its early growth phase. A sustainability plan is needed to establish requirements for data storage, staffing and resources for curation support services, and ongoing development projects, such as new features to meet local institutional or disciplinary needs. Additionally, OCUL has yet to finalize a costing model for long-term preservation of research data from member institutions.

National Collaboration: Portage

In 2015, the Canadian Association of Research Libraries (CARL) launched the Portage network, an initiative to develop a library-based research data management network in Canada (<https://portagenetwork.ca>). The aim of Portage 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 goals of Portage are two-fold:

1. To develop and support national infrastructure platforms for planning, preserving, and discovering research data.
2. To provide services to researchers and related stakeholders through a national library-based network of expertise on research data management (RDM).²⁴

Canada’s challenges in organizing nationally to support research data management and preservation has changed significantly in recent years. There is much greater awareness among funding agencies, campus research offices, and research-

ers themselves of the importance of data sharing and preservation. In addition, individual libraries have made inroads in supporting research data management locally and have positioned themselves as important partners in this area. The timing seemed right for something like Portage to bring about something that the library community has long desired: a national data archive.

Goal 1: Portage National Data Preservation Infrastructure

The Portage initiative has participated in a series of pilot projects involving partners from within and beyond the library community through RDC's Federated Pilot initiative (<http://www.rdc-drc.ca/activities/federated-pilot/>). In particular, three projects have been central, and all of them have involved collaboration between Portage and Compute Canada. The goal has been to test a number of possible software stacks for ingesting data from a range of research data repositories (both institutional and disciplinary) into a distributed national preservation infrastructure. One project under the this umbrella,* currently underway and described here, aims to integrate Dataverse and Archivematica, with the involvement of participants from across Canada, including OCUL's Scholars Portal, the University of British Columbia, the University of Alberta, Simon Fraser University, Artefactual Systems, and Dataverse.

The Dataverse-Archivematica integration has involved the development of customized open-source middleware that pulls published data sets from Dataverse instances using API calls and processes them for ingestion into Archivematica.²⁵ This involves the creation of a Submission Information Package (SIP), which combines a METS file describing the contents of the transfer, with the associated data files and metadata.²⁶ The middleware then initiates the ingest of the SIP into Archivematica. Processing the ingested content through the Archivematica pipeline is configured by the user on a case-by-case basis and therefore not part of the middleware. This middleware is under development for v4.x of Dataverse and is intended to be straightforward to update as Dataverse evolves

* Another project under the Federated Pilot umbrella, spearheaded by Simon Fraser University and Compute Canada, integrated Islandora and Archivematica (Melissa Anez, "Archidora," DuraSpace wiki, last modified by Tim Hutchinson October 2, 2015, <https://wiki.duraspace.org/display/ISLANDORA715/Archidora>). A third explored integrating Archivematica with Globus Data Publication, a new tool that is already in use by Compute Canada. Some background information about all of these projects is available in a presentation given at the 2015 CNI Meeting (Martha Whitehead, Brian Owen, Dugan O'Neil, Leanne Trimble, and Geoff Harder, "Collaborating to Develop and Test Research Data Preservation Workflows" [slides from presentation, CNI Spring 2015 Membership Meeting, Seattle, WA, April 13–14, 2015], https://www.cni.org/wp-content/uploads/2015/05/CNI_Collaborating_Whitehead.pdf).

(updates will not require any changes to the Archivematica software, only the middleware).

The overall goal of all of these related projects has been to generate a proof of concept that, through open standards and software, it is possible to ingest research data from a range of data repositories, perform preservation actions on the incoming data, and store the data in a distributed network that can accommodate a range of data types and storage locations. These initial pilot projects have shown promise, though scalability remains a concern. Portage is now working with Compute Canada on a set of requirements for a production platform, which would also integrate access and discovery as well as preservation. The focus for the next two years is on digital preservation and enhanced data discovery mechanisms, with an emphasis on building and improving open-source tools to enable curation and preservation of research data in Canada.

Goal 2: Portage Network of Expertise

The Portage network of expertise is still in its infancy, but its operational goals and service model have been laid out in the network's organizational framework.²⁷ It is anticipated that the network will bring together expertise in key areas such as metadata, curation, access and dissemination, preservation, data management planning, security and confidentiality, and others. The first expert group formed was the Data Management Plan (DMP) Experts Group, tasked with developing the general data stewardship template to be included in a new Portage online tool, known as DMP Assistant (<https://assistant.portagenetwork.ca/>), for creating data management plans.

DMP Assistant is based upon the open-source DMPonline software created by the Digital Curation Centre in the United Kingdom (<https://dmponline.dcc.ac.uk/>) and is hosted at the University of Alberta. This tool is customized to meet Canadian needs with a bilingual interface and a standard DMP template developed in anticipation of the introduction of required data management plans by Canadian research councils. As funding agencies determine their requirements and research communities in Canada articulate the data planning needs that best fit their disciplinary profiles, templates will be incorporated within DMP Assistant to accommodate each new requirement.

In addition to developing the tool, the DMP Experts Group conducted usability tests with researchers and other stakeholders. As a result, the tool not only incorporates best practices in data stewardship, it also provides an easy-to-follow workflow that walks researchers through key questions about data management. Such plans typically identify how researchers will address data security, metadata production, file formats, file handling conventions, data sharing practices, data dissemination methods, and arrangements for long-term preservation.

Future Directions

While the United States has seen data management planning requirements since 2011, which have been a strong driver for research data management activities,²⁸ Canadian efforts have been more anticipatory rather than reactive. For this reason it has been challenging at times to move forward with infrastructure development. Regardless, significant strides have been made and collaborations have been key to success in Canada to date. Many Canadian institutions are involved in RDM infrastructure projects at the local, provincial, or national level. There is a sense of momentum in this area, which must continue to build. But there is much more still to be done.

For example, in order for RDM infrastructure to meet the needs of all Canadian researchers, our user interfaces must be bilingual, since both English and French are official languages in Canada. The Portage DMP tool is an excellent example of new infrastructure being designed with this in mind. However, our data repository tools must follow. A project is underway to accomplish this for the Harvard-based open-source Dataverse software, where Scholars Portal staff are code contributors and are working on internationalizing the code (a project of interest to a number of other countries around the world as well). For example, the Université de Montréal in Québec has undertaken translation of the user interface text from English into French. Once this work is complete, this code may become part of the public Dataverse codebase and available to Dataverse instances around the world.

We anticipate that many projects of this nature will be undertaken under the umbrella of the Portage network. Together, it is hoped, these will come together to form the needed infrastructure for managing and preserving research data on a national level.

Conclusions

It is an exciting time in Canada for research data management. Libraries are seeing new opportunities to engage with their communities and with one another. Along with these new opportunities inevitably come challenges, such as costly digital infrastructure that must be managed on an ongoing basis. A number of approaches to research data management infrastructure have been explored in Canada to date, but no one approach holds all the answers. The Portage project has great potential to meet some significant unmet needs but will need sustainable funding in order to be successful.

The development of open-source tools, infrastructure, and support services for research data management is crucial if Canadian scholars are to successfully integrate these new activities into their workflows. While formal funder require-

ments for data management planning or data sharing are not yet established in Canada, consultations are underway and requirements are expected. Academic libraries have a history of supporting data access, dissemination, and preservation as well as an established mandate to participate in the preservation of the research outputs of their community (e.g., in institutional repositories).^{*} Libraries can provide leadership around the adoption of best practices and open standards and partner with a range of stakeholders in the development of infrastructure and tools. In Canada, the library community has been extremely active in encouraging research data sharing, going back as far as the 1960s, and is well positioned to play a leadership role going forward.

Notes

1. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada, "Draft Tri-Agency Statement of Principles on Digital Data Management," July 9, 2015, last modified July 20, 2015, <http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1>.
2. Chuck Humphrey, "Canada's Long Tale of Data," *Preserving Research Data in Canada* (blog), December 5, 2012, <http://preservingresearchdataincanada.net/2012/12/05/hello-world/>.
3. Chuck Humphrey and Elizabeth Hamilton, "Is It Working? Assessing the Value of the Canadian Data Liberation Initiative," *Bottom Line* 17, no. 4 (2004): 138, doi:10.1108/08880450410567428.
4. Ernie Boyko and Wendy Watkins, *The Canadian Data Liberation Initiative: An Idea Worth Considering?* IHSN Working Paper No 006, International Household Survey Network, November 2011, 2, <http://www.ihsn.org/home/sites/default/files/resources/IHSN-WP006.pdf>.
5. Chuck Humphrey, "Collaborative Training in Statistical and Data Library Services," *Resource Sharing Information Networks* 18, no. 1–2 (2005): 167–81, doi:10.1300/J121v18n01_13.
6. P. Bryan Heidorn, "Shedding Light on the Dark Data in the Long Tail of Science," *Library Trends* 57, no. 2 (2008): 280–99, http://muse.jhu.edu/journals/library_trends/v057/57.2.heidorn.html.
7. David F. Strong and Peter B. Leach, *National Consultation on Access to Scientific Research Data: Final Report* (Canada: Task Force for the National Consultation on Access to Scientific Research Data, 2005), https://www.cni.org/wp-content/uploads/2013/03/CNI_National_Newton.pdf.
8. Humphrey, "Canada's Long Tale of Data."
9. Chuck Humphrey, "Community Actions to Preserve Research Data in Canada," *Preserving Research Data in Canada* (blog), December 11, 2012, <https://preservingresearch-dataincanada.net/2012/12/11/community-actions-to-preserve-research-data-in-canada/>.

^{*} See, for instance, "UBC Library Strategic Plan 2010–2015," accessed May 20, 2016, https://about.library.ubc.ca/files/2012/09/StrategicPlan_2010.pdf. Many similar examples exist.

10. Michael Steeleworthy, "Research Data Management and the Canadian Academic Library: An Organizational Consideration of Data Management and Data Stewardship," *Partnership* 9, no. 1 (2014): 1–11, <https://journal.lib.uoguelph.ca/index.php/perj/article/view/2990/3278>.
11. Humphrey, "Collaborative Training in Statistical and Data Library Services."
12. Erin Forward, Amber Leahey, and Leanne Trimble, "Shared Geospatial Metadata Repository for Ontario University Libraries: Collaborative Approaches," *New Review of Academic Librarianship* 21, no. 2 (2015): 170–84, doi:10.1080/13614533.2015.1022662.
13. Wayne Johnston, "Digital Preservation Initiatives in Ontario: Trusted Digital Repositories and Research Data Repositories," *Partnership* 7, no. 2 (2012): 1–8, <http://criticalvoices.lib.uoguelph.ca/index.php/perj/article/view/2014/2637>; Jeff Moon, "Developing a Research Data Management Service: A Case Study," *Partnership* 9, no. 1 (2014): 1–14, <http://criticalvoices.lib.uoguelph.ca/index.php/perj/article/view/2988/3266>.
14. Mercè Crosas, "The Dataverse Network: An Open-Source Application for Sharing, Discovering and Preserving Data," *D-Lib Magazine* 17, no. 1/2 (2011), doi:10.1045/january2011-crosas.
15. Dataverse, "Universal Numerical Fingerprint (UNF)," accessed March 29, 2016. <http://guides.dataverse.org/en/4.2.4/developers/unf/index.html>.
16. Claire Austin, Susan Brown, Chuck Humphrey, Amber Leahey, and Peter Webster, *Guidelines for the Deposit and Preservation of Research Data in Canada* (Ottawa: Research Data Canada, 2015), <http://www.rdc-drc.ca/wp-content/uploads/Guidelines-for-Deposit-of-Research-Data-in-Canada-2015.pdf>.
17. Brian Lavoie and Lorcan Dempsey, "Thirteen Ways of Looking at ... Digital Preservation," *D-Lib Magazine* 10, no. 7/8 (July/August 2004), <http://mirror.dlib.org/dlib/july04/lavoie/07lavoie.html>.
18. Peter Van Garderen, "Archivematica: Using Micro-Services and Open-Source Software to Deliver a Comprehensive Digital Curation Solution," *Proceedings of the 7th International Conference on Preservation of Digital Objects (iPRES2010)*, 145–49, <https://ipres-conference.org/ipres10/papers/vanGarderen28.pdf>.
19. Ibid.
20. Guy McGarva, Steve Morris, and Greg Janée, *Technology Watch Report: Preserving Geospatial Data*, DPC Technology Watch Series Report 09-01 (Digital Preservation Coalition, May 2009), http://www.dpconline.org/component/docman/doc_download/363-preserving-geospatial-data-by-guy-mcgarva-steve-morris-and-gred-greg-janee; Inter-university Consortium for Political and Social Research, *Principles and Good Practices for Preserving Data*, IHSN Working Paper No 003, International Household Survey Network, December 2009, <http://www.ihsn.org/home/sites/default/files/resources/IHSN-WP003.pdf>; Library of Congress, "Sustainability of Digital Formats: Planning for Library of Congress Collections," accessed November 17, 2015, <http://www.digitalpreservation.gov/formats/fdd/descriptions.shtml>.
21. Jenny Mitcham, Chris Awre, Julie Allinson, Richard Green, and Simon Wilson, "Filling the Digital Preservation Gap: A Jisc Research Data Spring Project: Phase One Report—July 2015," accessed November 12, 2015, http://figshare.com/articles/Filling_the_Digital_Preservation_Gap_A_Jisc_Research_Data_Spring_project_Phase_One_report_July_2015/1481170.
22. Bronwen Sprout and Mark Jordan, "Archivematica As a Service: COPPUL's Shared Digital Preservation Platform/Le service Archivematica: La plateforme partagée de

- conservation de documents numériques du COPPUL,” *Canadian Journal of Information and Library Science* 39, no 2 (2015): 235–44, <https://open.library.ubc.ca/cIRcle/collections/ubclibraryandarchives/494/items/1.0132717>.
23. Center for Research Libraries, *Report on Scholars Portal Audit* (Chicago: Center for Research Libraries, February 2013), <https://www.crl.edu/reports/scholars-portal-audit-report-2013>.
 24. Portage Network, “Governance Structure,” *Portage website*, accessed July 21, 2016, <https://portagenetwork.ca/about/governance-structure/>.
 25. Artefactual Systems, “Dataverse,” Archivemata Wiki, accessed November 12, 2015, <https://wiki.archivemata.org/Dataverse>.
 26. Ibid.
 27. Whitehead and Shearer, *Portage*, 2–3.
 28. Katherine G. Akers, Fe C. Sferdea, Natsuko H. Nicholls, and Jennifer A. Green, “Building Support for Research Data Management: Biographies of Eight Research Universities,” *International Journal of Digital Curation* 9, no. 2 (2014): 171–91, doi:10.2218/ijdc.v9i2.327.

Bibliography

- Akers, Katherine G., Fe C. Sferdea, Natsuko H. Nicholls, and Jennifer A. Green. “Building Support for Research Data Management: Biographies of Eight Research Universities.” *International Journal of Digital Curation* 9, no. 2 (2014): 171–91. doi:10.2218/ijdc.v9i2.327.
- Anez, Melissa. “Archidora.” *DuraSpace wiki*. Last modified by Tim Hutchinson October 2, 2015. <https://wiki.duraspace.org/display/ISLANDORA715/Archidora>.
- Archivemata. “Format Policy Registry.” Accessed August 2, 2016. <https://www.archivemata.org/en/docs/fpr/>.
- Artefactual Systems. “Dataverse.” Archivemata Wiki. Accessed November 12, 2015. <https://wiki.archivemata.org/Dataverse>.
- Austin, Claire, Susan Brown, Chuck Humphrey, Amber Leahey, and Peter Webster. *Guidelines for the Deposit and Preservation of Research Data in Canada*. Ottawa: Research Data Canada, 2015. <http://www.rdc-drc.ca/wp-content/uploads/Guidelines-for-Deposit-of-Research-Data-in-Canada-2015.pdf>.
- Boyko, Ernie and Wendy Watkins. *The Canadian Data Liberation Initiative: An Idea Worth Considering?* IHSN Working Paper No 006. International Household Survey Network, November 2011. <http://www.ihsn.org/home/sites/default/files/resources/IHSN-WP006.pdf>.
- Canadian Association of Research Libraries. “Data Management Workshop.” Accessed August 3, 2016. <http://www.carl-abrc.ca/strengthening-capacity/workshops-and-training/data-management-workshop/>.
- Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada. “Draft Tri-Agency Statement of Principles on Digital Data Management.” July 9, 2015. Last modified July 20, 2015. <http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1>.

- Caplan, Priscilla. *Understanding PREMIS*. Washington, DC: Library of Congress, 2009. <http://www.loc.gov/standards/premis/understanding-premis.pdf>.
- Center for Research Libraries. *Report on Scholars Portal Audit*. Chicago: Center for Research Libraries, February 2013. <https://www.crl.edu/reports/scholars-portal-audit-report-2013>.
- Crosas, Mercè. "The Dataverse Network: An Open-Source Application for Sharing, Discovering and Preserving Data." *D-Lib Magazine* 17, no. 1/2 (2011). doi:10.1045/january2011-crosas.
- Dataverse. "Universal Numerical Fingerprint (UNF)." Accessed March 29, 2016. <http://guides.dataverse.org/en/4.2.4/developers/unf/index.html>.
- Forward, Erin, Amber Leahey, and Leanne Trimble. "Shared Geospatial Metadata Repository for Ontario University Libraries: Collaborative Approaches." *New Review of Academic Librarianship* 21, no. 2 (2015): 170–84. doi:10.1080/13614533.2015.1022662.
- Harder, Geoff, Leanne Trimble, Dugan O'Neil, Brian Owen, and Martha Whitehead. "Collaborating to Develop and Test Research Data Preservation Workflows." Paper presented at the CNI Spring 2015 Meeting, Seattle, WA, April 13–14, 2015. <https://www.cni.org/topics/ci/collaborating-to-develop-and-test-research-data-preservation-workflows>.
- Heidorn, P. Bryan. "Shedding Light on the Dark Data in the Long Tail of Science." *Library Trends* 57, no. 2 (2008): 280–99. http://muse.jhu.edu/journals/library_trends/v057/57.2.heidorn.html.
- Humphrey, Chuck. "Canada's Long Tale of Data." *Preserving Research Data in Canada* (blog). December 5, 2012. <http://preservingresearchdataincanada.net/2012/12/05/hello-world/>.
- . "Collaborative Training in Statistical and Data Library Services." *Resource Sharing Information Networks* 18, no. 1–2 (2005): 167–81. doi:10.1300/J121v18n01_13.
- . "Community Actions to Preserve Research Data in Canada" *Preserving Research Data in Canada* (blog). December 11, 2012. <https://preservingresearchdataincanada.net/2012/12/11/community-actions-to-preserve-research-data-in-canada/>.
- Humphrey, Chuck, and Elizabeth Hamilton. "Is It Working? Assessing the Value of the Canadian Data Liberation Initiative." *Bottom Line* 17, no. 4 (2004): 137–46. doi:10.1108/08880450410567428.
- International Association for Social Science Information Services and Technology (IASSIST). "Chronology of Data Libraries and Data Centres." Accessed August 3, 2016. https://docs.google.com/spreadsheets/d/1qmC_z50UDHh3Jwdlu6wGrtz1xjdfumBaad6P-6cUpBdY/edit#gid=0.
- Inter-university Consortium for Political and Social Research. *Principles and Good Practices for Preserving Data*. IHSN Working Paper No 003. International Household Survey Network, December 2009. <http://www.ihsn.org/home/sites/default/files/resources/IHSN-WP003.pdf>.
- Johnston, Wayne. "Digital Preservation Initiatives in Ontario: Trusted Digital Repositories and Research Data Repositories." *Partnership* 7, no. 2 (2012): 1–8. <http://criticalvoices.lib.uoguelph.ca/index.php/perj/article/view/2014/2637>.
- Lavoie, Brian, and Lorcan Dempsey. "Thirteen Ways of Looking at... Digital Preservation." *D-Lib Magazine* 10, no. 7/8 (July/August 2004). <http://mirror.dlib.org/dlib/july04/lavoie/07lavoie.html>.

- Library of Congress. "Sustainability of Digital Formats: Planning for Library of Congress Collections." Accessed November 17, 2015. <http://www.digitalpreservation.gov/formats/fdd/descriptions.shtml>.
- McGarva, Guy, Steve Morris and Greg Janée. *Technology Watch Report: Preserving Geospatial Data*. DPC Technology Watch Series Report 09-01. Digital Preservation Coalition, May 2009. http://www.dpconline.org/component/docman/doc_download/363-preserving-geospatial-data-by-guy-mcgarva-steve-morris-and-greg-greg-janee.
- Mitcham, Jenny, Chris Awre, Julie Allinson, Richard Green, and Simon Wilson. "Filling the Digital Preservation Gap: A Jisc Research Data Spring Project: Phase One Report—July 2015." Accessed November 12, 2015. http://figshare.com/articles/Filling_the_Digital_Preservation_Gap_A_Jisc_Research_Data_Spring_project_Phase_One_report_July_2015/1481170.
- Moon, Jeff. "Developing a Research Data Management Service: A Case Study." *Partnership* 9, no. 1 (2014): 1–14. <http://criticalvoices.lib.uoguelph.ca/index.php/perj/article/view/2988/3266>.
- Portage Network. "Governance Structure." Accessed July 21, 2016, <https://portagenetwork.ca/about/governance-structure/>.
- PRONOM homepage. Accessed August 3, 2016. <https://www.nationalarchives.gov.uk/PRONOM/Default.aspx>.
- Sprout, Bronwen, and Mark Jordan. "Archivematica As a Service: COPPUL's Shared Digital Preservation Platform/Le service Archivematica: La plateforme partagée de conservation de documents numériques du COPPUL." *Canadian Journal of Information and Library Science* 39, no. 2 (2015): 235–44. <https://open.library.ubc.ca/cIRcle/collections/ubclibraryandarchives/494/items/1.0132717>.
- Steeleworthy, Michael. "Research Data Management and the Canadian Academic Library: An Organizational Consideration of Data Management and Data Stewardship." *Partnership* 9, no. 1 (2014): 1–11. <https://journal.lib.uoguelph.ca/index.php/perj/article/view/2990/3278>.
- Strong, David F., and Peter B. Leach. *National Consultation on Access to Scientific Research Data: Final Report*. Canada: Task Force for the National Consultation on Access to Scientific Research Data, 2005. https://www.cni.org/wp-content/uploads/2013/03/CNI_National_Newton.pdf.
- University of British Columbia Libraries. "Summon search result for DBID: BAX-LO." Accessed August 3, 2016. <http://ubc.summon.serialssolutions.com/#!/search?ho=t&q=DBID:%20BAXLO&l=en>.
- . "UBC Library Strategic Plan 2010–2015." Accessed May 20, 2016. https://about.library.ubc.ca/files/2012/09/StrategicPlan_2010.pdf.
- Van Garderen, Peter. "Archivematica: Using Micro-Services and Open-Source Software to Deliver a Comprehensive Digital Curation Solution." *Proceedings of the 7th International Conference on Preservation of Digital Objects (iPRES 2010)*, 145–49. <https://ipres-conference.org/ipres10/papers/vanGarderen28.pdf>.
- Whitehead, Martha, Brian Owen, Dugan O'Neil, Leanne Trimble, and Geoff Harder. "Collaborating to Develop and Test Research Data Preservation Workflows." CNI Spring 2015 Membership Meeting, Seattle, WA. April 13–14, 2015. https://www.cni.org/wp-content/uploads/2015/05/CNI_Collaborating_Whitehead.pdf.

