Research Data Management – Approaches to Capacity Building by Acting Locally While Thinking Nationally

Talia Chung

University of Ottawa, talia.chung@uottawa.ca

Talia Chung, "Research Data Management – Approaches to Capacity Building by Acting Locally While Thinking Nationally."
http://docs.lib.purdue.edu/iatul/2017/challenges/2

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Abstract
The role of university libraries in research data stewardship has been in rapid growth and evolution. Key principles for good research data management standards are emerging and stabilizing internationally providing an opportunity for institutions to encourage and facilitate sound research data management practices among its students and researchers. Libraries must consider how we strengthen our collective ability to anticipate and respond to these needs. Using recent theoretical models of research data management services, this paper looks at an approach to build capacity around research data management services at a local level (the University of Ottawa Library) against the backdrop of a maturing national initiative (the Canadian Association of Research Libraries’ (CARL) Portage Network).

Introduction
In the last decade, academic libraries around the world have been increasingly attentive to the question of research data management (RDM) and the needs and opportunities arising on their own campuses. Technology-enabled research methods are generating vast quantities of digital information over the life of research projects, and challenging institutions traditionally engaged in the long term preservation of knowledge, such as research libraries, to steward new forms of the scholarly record.

Like many institutions, the University of Ottawa (uOttawa) is implementing supports and services to assist researchers in managing their research data based on best practices in literature, counterbalanced by what we are hearing on campus about our own school’s particular context and needs. uOttawa Library’s RDM services seek to contribute to the success of our researchers while ensuring that digital information arising from their research activities are effectively and responsibly managed throughout the data lifecycle.

Using a theoretical framework consisting of OCLC’s Tour of the Research Data Management (RDM) Service Space [Bryant, Lavoie, Malpas & OCLC, 2017], COAR’s Librarians’ Competencies Profile for Research Data Management [Schmidt & Shearer, 2016] and Cox, Kennen, Lyon & Pinfield’s (2016) model of research data service maturity, this article will examine how a single institution, the University of
Ottawa, is progressing in developing its own RDM service offering against the backdrop of a maturing national initiative, Portage, led by the Canadian Association of Research Libraries (CARL).

**Frameworks for RDM services and competencies**

Research data management, defined at its simplest, can be described as “the effective handling of information that is created in the course of research” [JISC, 2016]. Recognizing RDM as a significant development requiring high levels of engagement by academic libraries, research library associations, such as the Association of Research Libraries (ARL), OCLC and the Canadian Association of Research Libraries (CARL) have been actively engaged in these areas, in response to interest from their member institutions.

The benefits of good data management, from the perspective of researchers, administrative and funding agencies, and journal publishers, are well documented in the literature. Important issues such as researcher efficiency, re-use of publicly funded research data to create new investigational possibilities, and the support of published articles by sharing data which underpin conclusions, are driving the development of research data strategies on campus.

These benefits, along with other factors, contribute to the rising interest and need for robust discussions around each institution’s unique profile of priorities, research activities, funding, policies, capabilities and aspirations. More recent models of RDM services place a greater emphasis on consultation services and access to infrastructure, illustrating how our shared understanding of RDM services is evolving and, in the case of some institutions, has begun to shift from being a strategic issue towards more of a procedural issue [Pinfield, Cox & Smith, 2014].

In March 2017, OCLC published the first of a four report series looking at the research data management service space [Bryant et al., 2017]. Reviewing services developed by over a dozen research libraries across three continents, the authors propose a model representing three categories of RDM services commonly found throughout academic libraries. These categories include (i) education service, (ii) expertise service, and (iii) curation service.

In 2016, COAR’s Joint Task Force on Librarians’ Competencies in Support of EResearch and Scholarly Communication published a set of competencies [Schmidt & Shearer, 2016] describing the skills and knowledge needed by academic librarians to support researchers in managing their research data. Schmidt and Shearer define three categories of RDM support that may be offered by libraries, these include (i) providing access to data, (ii) awareness and support for managing data, and (iii) managing a data collection.

Both the OCLC and COAR reports describe RDM services being offered at research libraries around the world. There is considerable overlap in the types of services and activities described in both reports, with instruction along with services to build awareness for managing data, most frequently mentioned. According to both studies, many of the roles that the library may increasingly play can be categorized under OCLC’s *Expertise Services*, indicating that there is a distinct need for customized assistance in support of individual research projects. Schmidt and Shearer touch lightly on the role of the library in technical infrastructure support and development, activities largely found under the *Curation Services* category of the OCLC report. Appendix A provides examples of services organized by each OCLC service category.
University of Ottawa

The University of Ottawa (uOttawa), located in Canada's capital city, has an enrolment of approximately 42,000 students of which 6,500 are enrolled in graduate programs. Offering more than 450 programs, from 10 faculties including a medical school and a law school, uOttawa offers bilingual (English and French) programming, which is unique in Canada [Institutional Research and Planning, uOttawa, n.d.].

Ranking among Canada's top 10 research universities, uOttawa is a member of Canada's U15 Group of Canadian Research Universities and, as outlined in its 2011 strategic plan, seeks to rank among the top five research intensive universities in Canada by the year 2020.

Building uOttawa's RDM services

In the mid-1990s, the uOttawa Library established a numeric data service, which was quickly followed by geospatial data services. Supported by two professional librarians and specialized technical support staff, these service areas were critical in establishing library expertise and capacity in support of data access and management on campus, and enabled the library to address the need for data access, data analysis expertise, support for specialized analytical tools and knowledge of data security requirements of researchers. The university also hosts a Statistics Canada (Canada's national statistical agency) research data centre (RDC) which gives access to confidential microdata files to researchers with approved projects.

Since 2012, the development of an institutional policy for research data has progressed slowly. Currently uOttawa has not put into place a specific policy addressing research data stewardship although a draft policy has been developed jointly by the University’s research office and the Library. The draft policy outlines principles of accountability, roles and responsibilities of researchers and administrators, and includes some recommended practices, such as retention schedules. The policy also situates itself within the existing framework of campus policies which guide research activities, such as those which address the responsible conduct of research and intellectual property rights. It also signals to campus stakeholders the university’s commitment to sound practices in the management of research data. The next step in advancing the development of the policy is to carry out consultations with campus stakeholders, such as faculty members, student groups, research institutes, central information technology services, and others, to ensure that the policy addresses the full range of perspectives of those who would be impacted by its implementation.

Despite the lack of an institutional policy and an absence of policy requirements from major Canadian funding agencies, uOttawa’s Library has moved forward on building knowledge and capacity among librarians for RDM, a necessary precursor to establishing a full range of RDM services. The Library participated in various institutes including the ARL e-science institute (2011), CARL RDM institute (2012), and the DLF E-research network (2016).

In 2011, in response to the University's new strategic plan, the Library began to allocate an increasing amount of staff to provide research support services. Adopting the approach that research support, including assisting with research data, should align with disciplinary practices, the Library created research liaison librarian positions, each assigned to one or more faculties on campus, including: (i) health sciences and medicine, (ii) social sciences, (iii) science and engineering, and (iv) humanities. The intent was to mainstream these responsibilities within a liaison librarian model and within disciplinary
constraints while providing greater library capacity to support research activities on campus. In 2015, the Library created a position to lead the development and coordination of RDM services. The E-research librarian is tasked to develop and coordinate library activities and services in support of research data management. With considerable human resources aligned to support campus researchers, the Library moved ahead in further developing its RDM services strategy.

With the goal of bringing together expertise and resources in the Library to support RDM, the Library created a working group to define and develop a strategy to support RDM on campus. Its membership was drawn from units across the library with an emphasis on those with responsibilities to support RDM and its necessary infrastructure. Bringing together data services librarians, who have experience in managing licensed data, systems librarians, who are building infrastructure in support of digital discovery and preservation, liaison librarians, who are developing library services which contribute to the university’s research enterprise, and the E-research librarian, who is responsible for developing and coordinating library research data support services, the working group enabled the Library to better coordinate initiatives in support of RDM but also provided an important opportunity for shared learning and knowledge transfer within the group. Recognizing that robust research data support services require a certain level of knowledge and expertise to exist within a library, shared-learning to increase RDM support capacity was a key goal of the working group. Together, the group accomplished a number of goals including, (i) actively advocating on campus to raise awareness of RDM issues and the role of the library in research support, (ii) creating the University’s RDM website and developing learning materials for researchers, (iii) offering workshops on data management planning and good data management practices, and (iv) offering consultancy services in data management best practices and guidance on data repositories.

Benchmarking uOttawa’s RDM service development

Returning to Appendix A, a list of services currently offered, at any level, by uOttawa library is mapped against the OCLC RDM service categories.

In addition to the OCLC RDM service space framework [Bryant et al., 2017] and COAR’s competencies for librarians [Schmidt & Shearer, 2016], Cox, Kennen, Lyon & Pinfield (2017) published the results of a 2014 survey of RDM services offered by 170 academic libraries from around the world. Using a list of 21 RDM service activities derived from the literature, the authors were seeking to answer the question, “what services are libraries currently offering and what services do they wish to offer in the future?” The services most widely offered by institutions, including uOttawa, include,

- Maintaining a web resource/guide of local advice and useful resources for RDM (88%),
- Offer a research data management advisory service to researchers (85%),
- Offer data citation advisory services (79%),
- Offer advice on copyright and/or intellectual and/or licensing property rights relating to data and data management (79%)

In 2013, the Association of Research Libraries (ARL) surveyed academic research libraries in Canada and the U.S. to ask whether they offer RDM services, defined as “providing information, consulting, training or active involvement in: data management planning, data management guidance during research (e.g., advice on data storage or file security), research documentation and metadata, research data sharing and curation (selection, preservation, archiving, citation) of completed projects and published data” [Fearon, Gunia, Pralle, Lake, Sallans & Association of Research Libraries, 2013, p. 12]. Of the 73 responding institutions, 74% were providing RDM services as defined.
Interestingly, skills gaps identified in the surveys include knowledge of data curation, legal and policy issue advisory skills, data description and documentation skills, and a good understanding of the research lifecycle in specific subject domains.

Using the results of the 2013 ARL survey and Cox’s findings, uOttawa’s service offerings strongly resemble what is being offered in many institutions, but a weakness in this approach is in the belief that RDM services develop in a uniform and linear fashion thereby enabling a “maturity model” for services. This approach does not reflect a current, and evolving, understanding of what is needed to support the full research data lifecycle and excludes important activities such as managing access restrictions and privacy assurances for sensitive data.

In reality, many factors such as policy environment, resourcing, individual school priorities, and disciplinary practices, influence both the trajectory of service development as well as the pace or rapidity of development. For uOttawa, these factors have had an influence on the development of a research data stewardship policy but have not prevented the Library from proceeding to put into place services which would support such a policy. As in the case of many institutions, capacity and expertise development, particularly in the areas of data curation, preparation and preservation, continue to be a challenge for uOttawa; fortunately these are areas of active development by CARL’s Portage initiative.

The Portage Network, an initiative of the Canadian Association of Research Libraries (CARL)

The Canadian Association of Research Libraries is a membership organization representing the 29 major research libraries in Canada and two federal institutions. In 2013-2014, following a summit of stakeholders involved in the development of digital research infrastructure in Canada, CARL undertook to implement a national approach to RDM service, capacity, advocacy and infrastructure development. This work resulted in the creation in 2015 of CARL’s Portage Network, an initiative to develop a library based network of expertise in support of RDM for the primary benefit of CARL research libraries and their communities.

Portage is dedicated to a shared stewardship of research data in Canada through,

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

[Portage Network, n.d.]

Portage activities are organized around two themes, both of which highly favour a shared and distributed model of collaboration.

1. **Network of Expertise**: bringing together the expertise found across Canada’s research library community, Portage seeks to develop and sustain a network of expertise which develops and advises Canadian researchers and those supporting researchers with reliable and up to date information about RDM.

2. **Infrastructure Platforms**: Portage is working to connect various infrastructure and service components needed for data management by engaging key stakeholders from Canada’s research computing environment as well as regional library consortia, many of which have extensive experience in administering shared infrastructure systems for digital content. The infrastructure components reflect the full lifecycle of data, including systems for data curation,
data discovery and data preservation.

Portage draws expertise from the ranks of CARL member institutions, such as uOttawa, to carry out the work of its expert groups, each tasked with understanding, building and fostering a particular aspect of RDM support.

Six expert groups have been established, each addressing a facet of research data management, services and infrastructure, which together will form essential components of a national network. While responsible for advancing tangible outcomes for Portage, each expert group is equally tasked with understanding the “state of the art” in their respective areas.

The Data Management Plan Expert Group, the first expert group created and responsible for the development of Portage’s national DMP platform, DMP Assistant (in French, Assistant PGD - plan de gestion des données), its source code based on the DCC’s DMP Online. Based on international best practices, the group has created a general data stewardship template which has been implemented in the DMP Assistant along with guidance and instructions on data management planning.

The DMP Assistant has already been used by Canadian granting agencies to train funded researchers on good data management practices, by individual researchers to prepare DMPs, by librarians providing instruction to research teams and by administrators seeking to better understand data management when formulating institutional policies.

The Preservation Expert Group is working to identify research data management infrastructure and related best practices, with a focus on long term preservation of research data and its accompanying metadata. It is currently working with Compute Canada to explore workflow and infrastructure needed to support the preservation process.

Researchers will benefit greatly from the work of this expert group, as preservation infrastructure and expertise continues to be widely regarded as a critical gap in Canada’s digital infrastructure.

The Data Discovery Expert Group strives to support data creators and curators in planning, producing, and managing descriptive metadata to more effectively enable the discovery and reuse of research data across a wide range of disciplines. The group has published a white paper on research data discovery and the scholarly ecosystem [Barsky, Brosz, & Leahey, 2016] and a series of recommendations, most notably to develop a single point of access, rich in functionality, for data discovery. The group has also been actively collaborating with Compute Canada in the development of the Federated Research Data Repository (FRDR), a scalable national data discovery engine, repository and discovery platform for research data management [Compute Canada, n.d.].

Many individual data repositories currently exist in universities across Canada and the work of the Data Discovery Group will contribute to finding better ways to highlight and expose the data found in local data collections, rendering institutional data collections more visible, discoverable and more likely to be reused.

The Curation Expert Group is working to identify, evaluate, and promote best practices in preparing data and metadata for research, dissemination, and preservation in Canada. Evaluating workflows, standards and tools, the group’s goal is to ensure quality data and metadata are generated for research-in-progress as well as for long term preservation, with an emphasis on machine actionable practices. The Curation Expert Group is currently drafting a white paper to identify challenges and to make recommendations for
data curation practices and infrastructure.

As in the case of digital preservation, curation is another area of expertise and capacity long identified as a gap in Canada’s infrastructure for digital research data management [Compute Canada, 2016]. With the ultimate goal of rendering the data reusable, the work of the Curation group will benefit both researchers and data curators by identifying and supporting tools to enable more efficient workflows for in-research data and, eventually, more efficient submission to research data repositories.

The **Research Intelligence Expert Group** is responsible for gathering information from both Canada and internationally to provide an evidence-based approach to the development of RDM best practices in Canada. Its activities include coordinating the work of the Canadian RDM Survey Consortium, a project to survey researchers from 14 Canadian universities, representing all regions of the country, in order to gain a better understanding of researcher data management practices. The survey will provide results by broad disciplinary groups (social sciences & humanities, science & engineering, medicine & health sciences) at both an institutional level and at an aggregated national level.

The work of the Research Intelligence group serves to inform the work of individual universities, libraries and institutional planning offices by providing evidence with which to engage key campus stakeholders. Results of the survey, at the institutional level, will help participating universities to develop more customized approaches to service. The aggregated results of the national survey will also benefit non-participating universities by providing a benchmark against which universities can consider their own progress.

The **Training Expert Group** is responsible for identifying gaps in data management skills and also carries out specific projects, such as developing online training modules for researchers and library staff to learn about RDM best practices. In January 2017, the Training Expert Group published a white paper on the RDM training landscape in Canada (Fry, Doiron, Létourneau, Perrier, Perry, & Watkins, 2017), with recommendations such as the need to systematically identify knowledge and skills gaps, to better understand the various audiences for RDM training in order to customize training materials, developing a delivery framework for training, and exploring various approaches to increase capacity for skills development such as train-the-trainer, MOOCs, etc.

The group is currently developing two online training modules, (i) RDM 101, and (ii) Introduction to data management planning, both of which are intended to be very basic, targeting an audience with little or no knowledge of RDM. As the group further develops its repertoire of training resources and its understanding of the knowledge gap which exists in Canada, its work will benefit researchers and library staff at individual universities in Canada seeking to increase expertise in data management.

**uOttawa service gaps and Portage activities**

In appendix A, key Portage activities are mapped against the conceptual framework of OCLC’s RDM service space and uOttawa’s current RDM service offerings. The table indicates that the greatest opportunities for uOttawa capacity development are grouped in OCLC’s expertise and curation categories, which largely mirror the work being done by Portage’s curation, preservation and data discovery expert groups and includes services such as metadata creation, support for in-research data flows and long term data preservation. The gap in data curation and preservation capacity has also been reported in literature on RDM capacity development. Portage’s initiatives are clearly complementary to many of uOttawa Library’s activities and, as seen in Appendix A, can contribute to uOttawa’s ability to
increase capacity in both expertise and curation categories of services.

Conclusion - Local and national, a symbiotic relationship

The distributed and representative membership of the Portage expert groups affords a closeness or a quality of “being in touch” with the specific needs of Canadian universities. This mutually beneficial relationship provides the foundation for a rich, two-way exchange between Portage and member university libraries.

Engagement in Portage, and adoption of tools developed by Portage, will vary from institution to institution, depending on the state of service capacity development, level of resources available, as well as organizational and cultural elements present at an institution. For example, a small institution with little to no data services capacity could greatly benefit from the materials and tools developed by Portage, but a larger school with a greater ability to develop its own infrastructure may choose to invest in its own system rather than to engage in the development of shared infrastructure which may or may not respond to all of its institutional needs.

At uOttawa, the work of Portage has contributed to capacity development of library staff, access to infrastructure, in particular DMP Assistant, and has facilitated on-campus engagement with campus stakeholders such as the Office of Research.

In turn, uOttawa library staff have contributed to numerous Portage projects including participating in the Canadian RDM Survey Consortium, contributing to the development of learning modules for medical researchers funded by CIHR, and supporting the development of multilingual interfaces for tools and resources developed by Portage.

Moving forward, to maintain the momentum afforded by this collaborative model, Portage will be dependent on the quality of engagement, input and in-kind contributions, such as staff time or services, from individual institutions. Institutions engaged in supporting a shared, national vision of digital research infrastructure, will need to articulate a clear value proposition to its campus partners and leadership to ensure support for an initiative such as Portage which is a model of collective and collaborative action at scale.

Areas for further examination

This paper explores many current issues affecting research data support but, as RDM services evolve, many other questions can be anticipated. Below are a few items worthy of further examination.

Existing service frameworks demonstrate that more research is needed to better understand the range of RDM services required by researchers of all disciplines and, increasingly, by multidisciplinary or international research collaborations. A more granular understanding of services will lead to a broader discussion of competencies needed to effectively support RDM. COAR’s Librarians’ Competencies Profile for Research Data Management begins to describe necessary competencies for RDM support but our understanding of competencies must evolve along with evolving notions of RDM service models.

Ultimately, the most important indicator of success is an assessment of whether the services meet the needs, within the limits of available resources, of a university’s research community. Because resources are finite, in the case of a local - national service model as outlined in this paper, it is necessary to
consider an approach for cost modeling of RDM related activities and services provided at the local, provincial, regional or national levels. This is of particular importance in the case of shared services and infrastructure and is highly dependent on the model of collaboration.

Finally, when end of life for services and infrastructure occurs, numerous questions must be answered such as, what happens to data and metadata deposited? what is the sustainability plan for infrastructure and its contents? are there pieces that can be ported to other solutions and other institutions?
Acknowledgements

I’m very grateful to the following individuals who have generously given their time to share with me their knowledge and insight on the work of Portage and the state of research data management in Canada.

Portage: Eugene Barsky (University of British Columbia), Jay Brodeur (McMaster University), Dylanne Dearborn (University of Toronto), Jane Fry (Carleton University), Chuck Humphrey (University of Alberta), and Jeff Moon (Queens University).

uOttawa colleagues: Mish Boutet, Melissa Cheung, Jeanette Hatherill, Elizabeth Hayden, Jessica McEwan, Anthony Petryk, and Catelynne Sahadath.

Special thanks to Jessica McEwan, Jay Brodeur and Chuck Humphrey for their patient assistance and honest feedback in peer-reviewing and proofreading this paper.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.
References


<table>
<thead>
<tr>
<th>OCLC’s RDM service space model</th>
<th>uOttawa RDM services</th>
<th>CARL Portage activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Education services</strong></td>
<td>Educating researchers on compliance with funder, journal or institutional policies. Educating researchers on good data management principles including data management planning, data literacy and general skills building workshops. Advocacy for good data management to other key campus stakeholders. Developing general RDM learning resources such as a LibGuides or online tutorials; these may or may not take a discipline specific approach. Promote awareness of reusable data sources, such as data archives or licensed data sources.</td>
<td>Creates tools, such as the DMP templates and supporting instructional materials, used to instruct data management best practices. Identifies data management training needs, among both researchers and library staff, and develops or identifies resources to address the gaps. Assists with advocacy and policy development at an institutional, regional and national level. Bringing together research on RDM issues from Canadian and International sources. Undertakes research to fill knowledge gaps of the state of RDM in Canada. These efforts enable institutions to benchmark and, eventually, assess services and programs.</td>
</tr>
<tr>
<td><strong>Expertise services</strong></td>
<td>Advisory service to researchers for: (i) search and retrieval of external data sources, (ii) data citation, (iii) copyright and/or intellectual and/or licensing property rights relating to data and data management,</td>
<td>Provides tool and templates which can be used by researchers to prepare DMPs for specific projects (DMP Assistant). Identifies and evaluates curation tools to connect in-research data flows of active</td>
</tr>
</tbody>
</table>
These services tend to be individualized and closely match the stage of the research project (just-in-time services) often requiring human-mediated capacities to solve specific data management problems encountered by researchers during the research process.

| **Curation services** | Assigning unique identifiers, such as DOIs. | Portage is developing technical functionality for RDM needed by Canada’s researchers. Developments include:

(i) DMP Assistant, a data management planning tool.

(ii) Curation tools for managing in-research data flows.

(iii) Preservation infrastructure/repository for data deposit with the goal of rendering data reusable in the long term.

(iv) Discovery systems and standards for metadata to optimize discovery and reuse of deposited data across repositories and metadata standards. |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RDM Curation services offer the technical functionality needed to manage datasets throughout the research life cycle, subject to local policy requirements. The scope of curation services will vary from institution to institution.</td>
<td>Providing access to a data repository. Supporting access to data requiring privacy restrictions.</td>
<td>projects to preservation workflows to ensure long term preservation. Supports initiatives to demonstrate the importance of good metadata in enabling data discovery leading to increased dissemination of research results.</td>
</tr>
</tbody>
</table>

(iv) data storage options, (v) data analysis techniques, including data-mining and visualization.