

Research Data Management in Canada
Independent Study LIS 9410
Winter 2016

Prepared for:
Paulette Rothbauer

Prepared by:
Meg Miller

Literature Review

This review of the literature is an overview of Research Data Management (RDM) and how it relates to the academic library. It is broken down into five (5) major themes: Defining Research Data Management, The Drivers, The Stakeholders, The Role of the Library, and Implementation.

DEFINING RESEARCH DATA MANAGEMENT

DEFINING DATA

Before we begin to discuss research data management, we should start by defining data itself. Almost all of the articles surveyed noted the difficulty they had in this exercise, as depending who they were talking to and what their background was, data meant something different. Cox, Allchin, Collins, Lewis & Scott (2013, p. 37) note that “Working methods and departmental cultures are so diverse that part of the challenge will be to explain what we even mean by research data.” For the purpose of this review I will use the definition that Doucette & Fyfe (2013) found most appropriate to use in describing data. It is paraphrased from a Canadian Social Science and Humanities Research Council (SSHRC) needs assessment and states that “research data includes digital information that has been structured by methodology for the purpose of producing new knowledge. Content of digital data collections may include text, numbers, images, video or movies, audio, software, algorithms, equations, animations, models, simulations, etc., generated by various means including observation, computation, or experiment” (p. 165).

Data is also governed by the data lifecycle. Again there are many different versions, but they can mostly be broken down into the simple four (4) stage model used by the Canadian Federal Research Data Strategy Working Group in their 2008 Gap Analysis: 1) Data Production, 2) Data Dissemination, 3) Long-term management of data, 4) Discovery and repurposing of data.

RESEARCH DATA MANAGEMENT

Research data management can then be seen as managing the data that is produced by research as it passes through its lifecycle. After conducting a survey of various Australian institutions, Henty (2008) noted that data management also means different things to different people. In completing this review of the literature, I also found that there are many different ideas of what RDM should entail, not only in steps to be taken, but in its conceptualization (service vs. technology). Because of the newness of the field, much of the literature is still at the point of defining what RDM means to their specific institution. This review is meant to be a broad overview of the topic, as opposed to a hashing out of definitions.

Kathleen Shearer was tasked by the Federal Government of Canada to come up with a comprehensive brief on the state of RDM internationally. In her 2015 report, she notes four (4) common elements in policies between institutions. They include language on: data quality and standards, data access and sharing, data retention and preservation and the creation of data management plans. They also include provisions for privacy, traditional/indigenous knowledge, data of sensitive nature, and intellectual property. Some, but not all policies mentioned driving principles, scope of the policy, roles and responsibility and monitoring and enforcement. Different institutions have different ideas about how these policies will be enacted. Nine (9) points are made regarding this in the Tri-Agency's Draft Proposal on Research Data Management (2015). 1) Data management plans need to be created before the project begins describing how data will be collected, formatted, shared and preserved. 2) Constraints and obligations must be met (to granting agencies, research offices etc.). 3) Adherence to relevant standards and best practices of the discipline should be maintained. 4) Data should be collected and stored using formats to ensure preservation and access beyond the span of the project. 5) Data should be accompanied by

appropriate metadata to enable future users to understand the data. 6) Preservation, retention and sharing should be ensured. 7) Data should be shared as early as possible in the research process (when they are considered to be useful and of appropriate quality). 8) All users of research data should acknowledge through citation the sources of data they are using. 9) It is not necessary to preserve and share all project data. Consider reuse value when the data management plan is being constructed.

Shearer (2015) mentions that it is an intensive process with the major stumbling blocks being coming up with qualified personnel, the complex policy environment of research institutions, infrastructure support and clarifying roles and responsibilities of those involved being the major issues. The pay-off can be huge if done correctly Doucette & Fyfe (2013) note that “Proper management [...] of research data is essential for productivity, securing grant funding, enabling collaboration and ensuring the future use of data (p. 196)”.

THE DRIVERS

Institutions adopting these Research Data Management strategies are not doing so altruistically. In 2004 the Organization for Economic Co-operation and Development (OECD) put forth the Declaration on Access to Research Data from Public Funding. Canada was one of the 33 countries that signed (Canada, 2015). In Canada, SSHRC and CIHR have policies related to research data. SSHRC’s policy on data sharing has been in place since 1990, but it is not well known and lacks teeth (Vince Gray, UWO RDM team member, Personal Communication). The 2nd Tri-Council Policy Statement on the Ethical Conduct for Research Involving Humans also includes sections covering research data. We are now in the middle of a push to align our institutions with the policies mandated (or being strongly suggested) by the government in response to that 2004 initiative.

Higman and Pinfield note in their 2015 UK report on drivers of data sharing that institutions are trying to make RDM plans happen, not because they want to or because they see value, but because it is what the funders want, their opinions are reflected in much of the literature coming from the UK, Australia, US and Canada.

Aside from the 'because we were told to' standpoint, Borgman (2010) presents four (4) arguments for research data management and data sharing. 1) To improve data quality. 2) To enable others to ask new questions of extant data. 3) To advance the state of science and 4) To reproduce research (p. 14). These points are especially relevant as Henty (2008) notes that there are more and more large scale interdisciplinary projects with complex data sets coming down the pipes, so managing this data should be a top priority.

One of the biggest things to come out of this research data management push, is that with this heightened awareness being given towards research data is that "data is [now] seen as research output of equal importance to publications" (Henty 2006, p. 2).

THE STAKEHOLDERS

In terms of who the stakeholders are in this initiative, the Federal Draft Policy (2015) report lists researchers, research communities, research institutions, and research funders as those impacted. Upon reviewing the literature in the context of the academic setting, this can be further broken down into: funders, administration, intermediate-expert researchers, research office, library and students. While it is widely accepted that these are the groups involved, their roles and responsibilities are slightly more amorphous.

FUNDERS

While the Tri-Agency Draft (2015) describes the responsibility of research funders as being those to develop policies and requirements, provide applicants with assistance, promote excellence and develop assessment materials, the literature reflects the funders as drivers of RDM, as a one way unbending, un-assistive force (Higman & Pinfield, 2015).

ADMINISTRATION/INSTITUTION

The responsibility of Administration/the Institution is tasked with supporting the researcher (Tri-Agency Draft, 2015), but there is very little in the literature at all discussing this. Jain's 2010 study looking at Botswanan institutions and adoption rates of RDM strategies was one of the few that brought up administration. He notes that when strategies had institutional support (financial, verbal etc.) there were higher rates of adoption of the strategy being implemented as users had higher levels of trust towards them.

RESEARCHERS

There is significantly more research done on the needs and expectations of the researchers themselves. The federal directive notes that it is up to the researcher to "follow the requirements of applicable institutional policies and professional or disciplinary standards, & acknowledge and cite datasets that contribute to their research (Canada 2015, p. 4). Many articles note what this might entail, but the 'how' they will actually do this in practice is lacking (i.e.: who will reach out to who, who will enforce standards etc.).

RESEARCH OFFICE

The research office was only mentioned specifically in one of the 30 articles reviewed. Henty (2008) noted that it was only after being forced to work together over the institutional repository did the University library and University Research Office's mandates align with that of the wider

institution. If this is the case, then the same sort of strategic partnership could arise out of the RDM process, with both parties being concerned with the stewardship of the scholarly research process.

Why is this not looked at more closely?

LIBRARY

See section following 'Stakeholders' for an in-depth look at Libraries and their relationship with RDM.

STUDENTS

Students are also skipped over in the Tri-Agency Draft. Students (graduate and undergraduate) have less research experience than their supervisors/professors. One of the few articles covering this area by Doucette & Fyfe (2013) surveyed students at all levels from a variety of disciplines. All thought RDM was important, and claimed to understand how to do it. Further questions demonstrated they did not know near as much as they thought. Educating these students in appropriate RDM methodologies before they hit the workforce would be a valuable thing to do.

THE ROLE OF THE LIBRARY

WHERE LIBRARY SEES ITSELF FITTING

The in a recent review of the literature by Pinfield, Cox & Smith (2014) they noted that most bodies within the institution agree with the fact that there is ambiguity in what their roles are and should be. In terms of the library, there are strong feelings about the link between open access and open data/data sharing with the overall feeling being "preservation is a good thing, but we don't know what we're doing yet" (p. 14). One of the librarians interviewed also made the point that informational RDS was much more popular to offer than technical services.

There are many different mindsets about how the library should be best integrated. In a summary piece on realizing the role of the academic library in the research data management process Cox et al. (2013) summarize the points of how the library could be involved in (from a British perspective). They see the library as having expertise in metadata and conservation, copyright and data licensing; as well as expertise in enabling resource discovery. Additionally, libraries are often affiliated with the institutional repository and have expertise in deposit and long-term preservation. Their reasoning is that the library has the potential, and therefore should be part of most aspects of the research data management plan. To facilitate this point, Henty (2008) notes that the library needs to become integrated into the research lifecycle itself, not just become involved at the end. To do this they need to adopt different tools and skills, forge new partnerships and engagement. By acquiring all these new things, then the researchers will be more accepting towards library involvement. Bates (2014, p. 23-24) furthers this mentality, noting that “Libraries should not just be about support, [they] should involve with collaboration. Borgman (2010) on the other hand, agrees on the point of expertise, but is not as anxious force to be involved in everything. She suggests that if the library evolve from “having a focus on reader services, to having a focus on author services” (p. 13), then this integration into the research lifecycle would come naturally. She goes on to note that no matter what the RDM structure of the institution, the library would play a role in facilitating the process of sharing and using the research data created.

PROBLEMS THE LIBRARY RECOGNIZES

While some in the library see themselves as a solid fit to take on the role of research data management shepherd, others are not so sure. Verbaan & Cox (2014) completed an interesting study entitled ‘Collaboration or Competition? Responses to Research Data Management in UK

Higher Education by Librarians, IT Professionals, and Research Administrators'. They uncovered that in many cases the library ended up with a lead role in RDM planning because no-one else wanted to deal with it, and as librarians are constantly trying to define themselves, they scooped it up because they thought it might give them weight. When compared with other groups (IT, the research office) they had a much narrower view of what RDM should entail and often it was short-sighted. Their main focus was on getting people to adopt their service, not the service itself. The authors made the note that the library was the least well equipped of all stakeholder groups to take a leadership role (fewest PhDs, lowest research output, limited technical skills (p. 288)).

The other major point running through the literature was skill level. Cox et al. (2013, p. 37) discuss that the point of offering a RDM service is to make life easier for researchers and demonstrate how good practice can save them time and effort. They note that RDM is very specialist, maybe too specialist to be added to the existing workload of the librarian. "One major challenge is to find a balance between what support [the library] would ideally like to offer, and what is feasible within the staffing and resource levels." Nielsen & Hjørland (2014) second this train of thought. Suggesting that maybe the library should not be the ones making and planning metadata structures, and that it should be left to subject specialists. They see value in bringing in information specialists who are also domain specialists to deal with the creating a solid base for RDM efforts.

Thompson (2015) viewed the biggest challenge for RDM was having adequate support personnel so the service would be adopted and taken seriously. She contrasted the types of staff and their skillsets that were hired by the Research Data team from the National Centre for Atmospheric Research with those who were in charge at local institutions. The main difference was in the technical aspect of those hired, with NCAR hiring in 4 major roles: "Data Managers (curation

of data), Data Engineers (build tools and applications), Data Scientists (build models and simulations), and Managers, (provide oversight and direction.)” (p. 4). Thompson notes that data management position would be best suited for a librarian, and it would be very helpful if they were to have subject specific knowledge.

The other side of this skill-level piece is that those involved in institutional RDM do not feel they have the support of their institution. In the North American institutions that Tenopir, Sandusky, Allard & Birch (2014) interviewed, they noted that their institutions supported them with training initially but ended there. RDM is a rapidly evolving field, continuing education for those engaged in it is necessary for them to keep up.

IMPLEMENTATION

When it comes to the implementation of a research data management strategy, there are no long-term studies because the area is so new. Additionally, as Higman & Pinfield (2015) note, the push is external and lacks administrative support and enforcement, so uptake has been slow. Many institutions (Cox et al. 2013, Nielsen & Hjørland 2014, Verbaan & Cox 2014, Borgman 2010) make a point of noting that for RDM strategies to be successful, they cannot be developed in isolation (from other departments, institutions and or disciplines), which although recognized in theory is not taking place in practice.

Shen and Varvel (2013) use the phrase ‘if they build it, they will come’ to depict what does not happen in terms of RDM strategy building. Downloading a software package or platform is not enough. Human, financial and technological considerations all come in to play. Employees may need to be reskilled so they are taken seriously by faculty (Cox et al. 2013), in the case of Purdue

their entire library system was reorganized to be able to better serve the new “knowledge economy” (Carlson & Garritano 2010, p. 2). Anurada (2005) mentions that to get people to begin archiving their data, involved building a collection of retrospective works as well so that researchers could actively see the benefit of engaging in the process.

Ma, Wang, Zhu & Tang (2009) discuss their institutions attempt at RDM that was very much cobbled together and created in isolation from the researchers. Their attempts at data exchange were not successful at all because they lacked a cohesive framework. Items were not retrievable and were duplicated, submitting the data was seen as a burden that researchers just did not have time for. This idea of RDM adoption as being burdensome came up time and again in the literature (Cox et al. 2013, Pinfield, Cox & Smith 2014, Higman & Pinfield 2015), part of dealing with this will be advocating for the service, but the other large part will be listening to the concerns of the researchers and making the process as straightforward and simple as possible. Childs, McLeod, Lomas & Cook (2014) point out that this is the ideal place for technology to come in, and build custom solutions for researchers to automate certain tasks, track citations and data usage or programmatically embed metadata content, these are just thoughts, and they have yet to be put into action. The main problem is, that RDM is a nascent area, and solutions that have been implemented thus far, have not been around for long enough to have had a chance to fully mature.

NOTEWORTHY FRAMEWORKS

That being said, two (2) noteworthy frameworks for dealing with RDM were uncovered over the course of the literature review, the first being that of the ‘Wicked Problem’ and the second being that of the repository.

WICKED PROBLEM

This method of framing Research Data Management seems to be specific to the UK. Cox Pinfield & Smith (2014) borrowed from the world of systems design and used Rittel and Weber's 1973 idea of the wicked problem to define RDM. A wicked problem "is a unique complex problem that is viewed differently by different stakeholders" (p. 3). Additionally, because of this discord, these types of problems cannot be solved, only resolved. As the concept has been around for many years, there is a whole body of work surrounding it including the best ways to deal with these sorts of problems. Of note: words such as empathy, integrative thinking, optimism, experimentalism and collaboration, which is not usually the mindset that was involved in RDM planning. Once the study authors got the stakeholders of their RDM strategy to see it as a 'wicked problem' they could propose different solutions in line with the literature pushing the ideas of experimentalism and collaboration, as well as having it out on the table that compromises would need to be made in order for a functional solution to be possible. The UK RDM community is small, but arguably the farthest advanced in their implementation of an RDM strategy.

LESSONS FROM THE REPOSITORY

This framework is not properly named, but many of the articles discussed how it was valuable to apply what they learned from implementing an Institutional Repository (which also came as an administrative push to open access) to implementing their RDM strategies.

In 2002 the concept of the institutional repository took off, starting at MIT with the creation of DSpace (Lynch 2003, p. 327). There still is confusion as to whether the repository is considered a technology or a service with those who are directly involved with it comfortable with their role, but often with the rest of campus being unsure of what the repository is aside from an institutional database with articles. Jones (2008 p. 158) and Pinfield, Cox & Smith (2014) also note that the

repository should not be a static thing, but a service that evolves with technology and research outputs. Those developing RDM strategies should keep this in mind, that their strategy will never be 'complete', but be always evolving to best meet the needs of their users.

Foster & Gibbons (2005) discuss the importance of branding and marketing of their service. Lots of jargon and buzzwords were used in the roll-out of the IR at their institution, and it was not widely adopted. When interviewing faculty about their sentiments on their service they realized that their users had no idea what their service actually did, and had their own ideas as to what their needs were for a solution to facilitate open access to scholarly publications (p. 9). In many cases, the suggestions faculty had, were already implemented, just not described clearly. Once IR staff implemented these suggestions and made their wording clearer in outreach materials adoption of the repository dramatically increased. Being clear about how the service applies to the user. In the case of Bankier & Percali (2008) changing how they advertised the IR from a digital repository to a digital showcase improved faculty perceptions on the usefulness of the service.

Another major theme that can be taken from the implementation of the institutional repository is how important it is to integrate itself into the scholarly communication cycle. Henty (2008) notes that being a part of the cycle brings the repository legitimacy from the rest of the institution. It is not longer an afterthought, but a part of the scholarly communication process. In the case of institutions in Australia, this integration brought the library together with the research office and forced them to better align with institutional mandates, bringing a significantly greater level of support from the institution to both the research office and the library.

If institutions keep in mind these lessons learned from implementing their repository services while coming up with their RDM strategies, it will make things easier for them in the long-

run.

PROBLEMS WITH ADOPTION

Currently research data deposit and the creation of RDM plans is not very high. Pinfield, Cox & Smith (2014) and Tenopir et al. (2014) note that this can partially be attributed to the fact that it is very much an area of development, where everyone is feeling their way around, there is a lack of direction at the institutional level, and these policies are not being enforced.

On the other side of the coin, researchers often think that their processes and data are perfect, so this idea of a research data management plan cannot apply to them specifically (Pinfield, Cox & Smith, 2014). They (the researchers) also see their data as being unique and distinct from those being created by other departments and don't want to see their data get lumped into a general data bank so will use many different excuses (using out of date technology or file-formats) to stop from being required to deposit data (Jantz & Wilson 2008).

Researchers also have concern about the data sharing aspect of this whole push. This is especially prevalent among those doing 'small science'. Cragin et al. (2010, p. 4034) summarize multiple instances of secondary users of the data misusing it; either misinterpreting it, cherry-picking to make a point, or not properly attributing it to its original creator.

In all additional policy, enforcement and institutional support and outreach are needed before research data management strategies will be fully adopted.

WHERE IS CANADA

Shearer (2015) notes that Canada is behind the UK, Australia, the Netherlands and US in the adoption of a cohesive RDM strategy. Literature notes the University of Alberta and UBC are at the forefront of implementation in the academic institutions.

The Canadian Association of Research Libraries (CARL) is to roll out it's Portage Initiative, which is a national bilingual data management planning system. It involves a network of expertise for research data management (which will involve customizable downloadable templates for institutions to aid in their creating of data management plans) as well as a national research data preservation and discovery system.

Many institutions have their own unique solutions to managing their research data (Dataverse instances, integrating into the repository) so it will be interesting to see how a national initiative will fit into this.

To take a closer look at how Canadian academic institutions are supporting Research Data Management a survey of institutional materials was undertaken.

THE SURVEY

The sixty-four (64) institutions surveyed were all publicly funded English/Bilingual Canadian universities. Initially the idea was to write a script to automate the hyperlink scraping process from these websites, but after some thought, it was determined this was not the best way to move forward since context was very important (who was linking to what, why and how). The link collecting process was done manually.

In addition to seeing what content was being used, I was also interested in seeing who was 'in charge' of research data management at institutions. I started by navigating to the main page of each institution and searching 'research data management' and 'RDM' in the search box that was available. Of the 64 universities surveyed, 36 (56%) of them did not return any content. The remaining 28 universities provided me with 916 links to analyse.

THE SURVEY PROCESS

Before I started, I opted to set up an Access database so that I could control vocabulary on input, making clean-up easier in the long run. For each link I collected:

- Institution (name of university)
- Department (who within the university was creating content)
- Location (who was being linked to, external or internal as well)
- What (what was the link: policy document, checklist, libguide, email address...)
- Purpose (controlled vocab created from the categories in the Canadian DMP assistant, Portage)

On average each institution's RDM strategy linked to 30 places. But there were some outliers, with one institution linking out to 164 different things.

THE SURVEY RESULTS

Verbaan & Cox (2014, p. 218) noted that it is often the library that takes control of the research data management strategy, and this survey agreed with this, with the library being in control with all instances but 5 in two of these cases, IT is taking a lead, in 2 other cases, it is a centre of advanced computing, and in one case the research office. In all of these five (5) cases, the library also had created content to support RDM at their respective institutions.

The next level of delineation was to see what was being linked to: something outside of the institution or something internal. These internal links were tagged with the 'self' tag. The following

field was simply described as 'What'. Here I described the link in less than 5 words. If it was a page of a libguide, a homepage of a website, policy, procedure, statement, checklist, template or plan.

THE SURVEY RESULTS: INTERNAL LINKS

Of the internal links (created by the library to library resources), they were either links to the institutional repository or links to a LibGuide created by the institution. LibGuides were usually divided into several sections reflecting various actions a stakeholder might be involved with and be looking for support on.

A common scheme is:

- Home/overview
- Advice/FAQs/Plan
- Organizing/Documenting/Metadata
- Sharing/Repositories
- Data Citing
- Contact/Help/e-mail link

What I found interesting here is that there did not seem to be a hierarchy of sections (planning before, documenting before sharing) and the user could dive in where-ever.

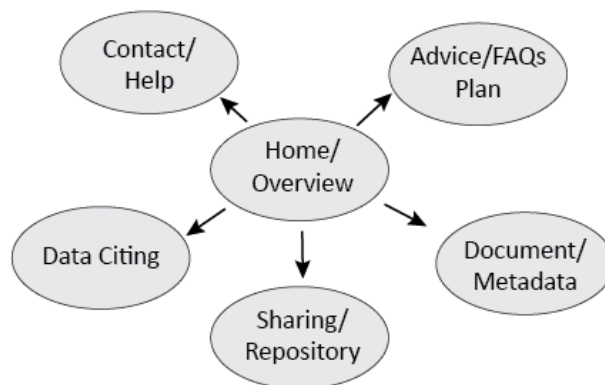


FIGURE 1: WEBSITE LAYOUT

Only one institution had a page dedicated to talking about all the stakeholders in the process and how the library supported them in the RDM process, I expected to see more of this for some reason. This institution and two (2) others were the only ones that dedicated separate pages to the 'why' or the impetus of all of this RDM push. Other institutions either did not mention it or crammed it all into a hyperlink rich paragraph full of jargon.

THE SURVEY RESULTS: EXTERNAL LINKS

For the external links, I created a vocabulary based on the six (6) main sections from Portage's Data Management Plan Assistant. I also added a category for 'policy' so I could see whose policies were being linked to, and one called 'guide' to cover general RDM instruction that did not fit into one single DMP category. I called these links 'Learning Materials', and they were categorized by purpose as: plan, collection (structure, name control etc), documentation/ metadata, storage/backup, preservation and sharing/reuse.

POLICY

Of the external links, 117 of them linked to various policy and procedure documents such as ethics policies by individual university research offices, SSHRC's data archiving policy, the tri-agency's statement of principles. Numerous links out to SSHRC, NSERC, CIHR and Tri-agency documents were expected. What was unexpected was determining that approximately 18% of the policy documents linked to were American in origin, and 2% were from the UK. As previously stated, these were usually found in a small paragraph on why a person needs to adopt RDM strategies.

LEARNING MATERIALS

The remaining external links, I categorized as 'Learning Materials'. Content that was linked to, to augment the materials provided by the institution. After 100 or so I noticed that while some

links would link to exact content (a checklist for instance), others would just link to the homepage of the site that the checklist could be found in. I found this interesting so revisited the links I had already done, and tagged them if they linked to the general homepage instead of actual content.

The breakdown of these links can be seen in the following chart. The first column contains the percentage of links from each category when the links that were just to a homepage were included, the second column are those same categories with those general 'homepage' links are removed and only ones linking to actual content are used. When removing these general 'homepage' links, there was a significant change in one category: documenting, with these links making up an increase of 7% of this category (see Figure 2). All other changes were not statistically significant.

	n=713 (includes general links)	n=593 (does not include general links)
<i>Collecting</i>	2%	2%
<i>Documenting</i>	19%	12%
<i>Guide</i>	12%	13%
<i>Plan</i>	13%	16%
<i>Policy</i>	18%	19%
<i>Preserving</i>	2%	3%
<i>Sharing</i>	30%	30%
<i>Storage</i>	4%	5%

FIGURE 2: COMPARISON OF PERCENTAGE OF LINKS FALLING INTO EACH OF THE 'PURPOSE' CATEGORIES.

Cox et al. (2013) posited that the areas of RDM the library was best suited to support were:

documenting, preserving and sharing. The survey of Canadian links uncovered that Sharing was the

most popular area of support, with Documenting coming second, followed closely by Policy.

Preserving was tied with Collecting as the lowest supported category.

LOCATION

The external links (minus policy) were then analyzed for impact by totaling the number of times institutions in a location were linked to (see figure 3).

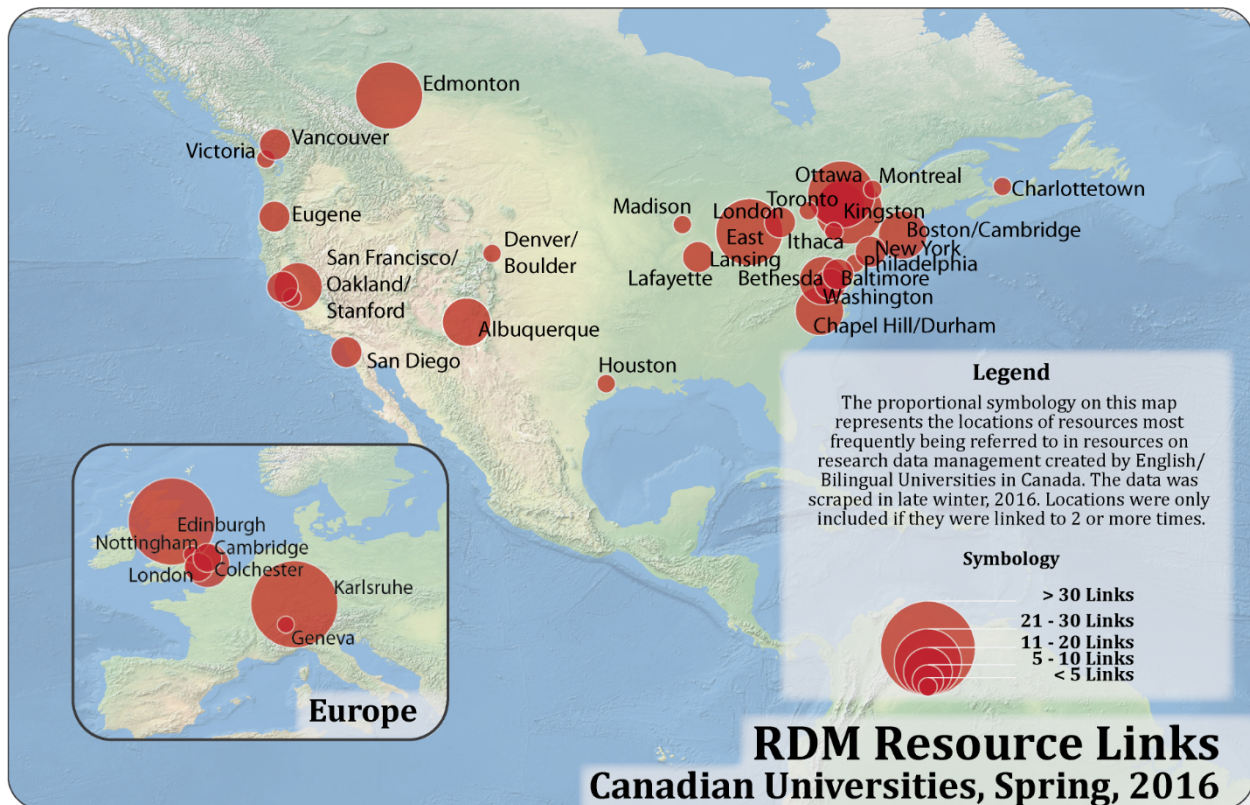


FIGURE 3: LINK IMPACT, BY LOCATION

In Canada, popular locations were Edmonton (University of Alberta, creator of Portage, the Canadian DMP assistant), Toronto (location of Scholars Portal, home of Dataverse), and Kingston (Queen's University, where Jeff Moon who is Chair of CARL working group on RDM is employed). While linking to Portage was a popular option for Canadian institutions, surprisingly not all did,

electing to link instead to American or UK resources, which are not geared toward the Canadian research environment.

Europe contains institutions with the most popular links, with the Digital Curation Centre at the University of Edinburgh, as well as the MANTRA RDM online training course that is offered. The other massive centre is in Germany, at the Karlsruhe Institute of Technology, who are managing the merger of re3data (directory of repositories) and DataBib (data linking service) under the DataCite umbrella.

NOTEWORTHY RESOURCES

A brief list of noteworthy resources has been compiled. Portage and Scholars Portal Dataverse were not included in this list, as all Canadian institutions should be linking to this infrastructure as it has been set up explicitly for Canadian Universities to use.

LIBGUIDES OF NOTE

Michigan State University Library: <https://www.lib.msu.edu/rdmg/>

MSU's LibGuide on Research data management provides a thorough overview of the area. The landing page is broken into five (5) sections, including an overview, an explanation of what the library is able to assist with, a list of training resources (linking to a guide with significantly more detail) as well as a list of upcoming RDM training put on by the library (no Canadian guides did this), another page with all the RDM resources offered by the University annotated (so the user can easily select what they are after, and a contact page. The more detailed guides are broken up into the standard sections of: planning, collection, short-term storage, study and analysis, publishing and reuse, long-term storage and an index of links.

This guide is very clean and approachable, allowing users with a variety of competencies to easily navigate and find a resource that is appropriate for their level. The only way this is able to happen is that the library supports and is actively supported by many different groups and research centres on campus (Statistics department to help with analysis, Computer science for storage, research office for policy), because of this, the library is able to support users of all levels, at all points of the research lifecycle.

MIT Libraries: <http://libraries.mit.edu/data-management/>

MIT's guide is similar to MSU's in that it is designed cleanly and does not come off as overwhelming. There is a newsfeed and calendar integrated into the homepage, so visitors can see exactly what is going on. But while MSU's guide packs a tonne of dense information into an appealing package by burying things many layers deep, MIT takes the opposite approach and lays three (3) main categories out on the main page with several subheadings apiece (Make a Plan, Store your Data, and Share your Data), it is very clean and conversational in its approach, free of jargon (or if there is any, it is defined). They give an explanation of each section, so that someone new to the area can familiarize themselves, but does not duplicate efforts making materials for a variety of competencies.

I prefer this guide to that of MSU because of its clean layout and conversational approach and quality of information, but can see how someone else might prefer that of MSU as it seems more 'science-y', on close inspection the content has been carefully curated instead of added for the sake of adding content.

Queen's University Library: <http://guides.library.queensu.ca/rdm>

Queen's LibGuide is the only Canadian content on the list. Compared to the other two guides it is overwhelming on first navigating to the site, the text is tiny and there is a lot of it. It is divided into nine (9) sections: Overview, Writing a DMP, Funding agency guidelines, Metadata, Data Repositories and archives, Citing data, Best practices, Library as data partner, QUL data archive and Contact.

It was included in this list, because of two of these sections: Library as data partner, no other Canadian guide covered how the library fit into the broader RDM picture of the institution. QUL Research Data Archive is the other section that acts as a list of those studies whose data has been deposited into the, demonstrating to others that people are making use of the service, as well as enabling discovery of the archived data.

REPOSITORY LISTING

DataCite (Re3data.org and DataBib): <https://www.datacite.org/>

DataCite is the result of a merge between the re3data.org directory of repositories and DataBib's data citation service. Here you can get assistance with data citation, find a dataset, get you DOI statistics or find a repository to deposit your research data into. It is free to create an account and brings together members and interest groups from around the world from various disciplines.

DOCUMENTATION/OTHER

DataONE → *Best Practices*: <https://www.dataone.org/all-best-practices>

DataONE (Data Observation Network for Earth) has created a simple list of best practices to consider when engaging in RDM practice. Why it appears on the list, is the way it is organized, with the landing page being a plain language best practices guide, and then at the bottom of each entry are tags for major themes from the entry, allowing the reader to browse and discover details at their own pace. In all other resources listed here, the learning process is very linear (first this, then that), but not so here.

This guide should be augmented with a more detailed resource that covers the nuts and bolts, but for someone who is looking for an unstructured introduction to data management, this would be an ideal resource to start with.

ICPSR → *Data Preparation and Archiving*:

<https://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/>

The Interuniversity Consortium for Political and Social Science Research (ICPSR) has created a guide to the data preparation and archiving process. It is an American publication, but in terms of preparation and archiving of social science data possibly the most complete and detailed guide out there.

UK Data Archive → *Managing and Sharing*: <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf> and <https://www.ukdataservice.ac.uk/manage-data/handbook>

The Data Archive is out of the University of Essex in the UK. This 40 page handbook on Managing and Sharing research data and companion resource do not break much new ground in terms of their main content on managing and sharing research data. It is the 'annotation' and exercises on the content that make this work unique. It touts itself as a practical guide, and it is. It contains a variety of use cases, and my favorite: conversational tips on how to discuss various matters pertaining to RDM. It sounds cheesy, but I think those prompts could prove very useful for people who are new to, or not well versed in the world of RDM. It is also a helpful tool for providing insight into the sentiments various stakeholders may have pertaining to many different issues on the RDM spectrum.

UK Digital Curation Centre → Disciplinary Metadata Creation:

<http://www.dcc.ac.uk/resources/metadata-standards>

This section of the Digital Curation Centre out of the University of Edinburgh was linked to by most institutions as a resource for data documentation, as it is basically a directory to all common metadata standards organized by subject area. Each standard is annotated in plain language so that the user can see its best use. The major shortfall of this resource is that you must browse to the resource, there is no way to search an exact standard (but then I guess you wouldn't be using this resource if you already knew what you were going to use as documentation).

UK Digital Curation Centre → MANTRA: <http://datalib.edina.ac.uk/mantra>

This is another resource from the University of Edinburgh. MANTRA is a free online course for anyone who needs to do anything with research data. Themes include: Research data explained,

Data management plans, Organizing data, File formats and transformation, Documentation, metadata and citation and Storage and security. While it is a UK resource, the vast majority of the information is relevant in the Canadian context (just transpose funding body names).

The material is fairly dense for a beginner, and it might be good to supplement the materials with one of the guides from MIT etc to get a feel for the content before jumping into the course and assessment, which does not hold back with its use of jargon.

References

- Anuradha, K. T. (2005). Design and development of institutional repositories: A case study. *The International Information & Library Review*, 37(3), 169–178. <http://doi.org/10.1016/j.iilr.2005.07.006>
- Bankier, J.-G., & Perciali, I. (2008). The Institutional Repository Rediscovered: What Can a University Do for Open Access Publishing? *Serials Review*, 34(1), 21–26. <http://doi.org/10.1016/j.serrev.2007.12.003>
- Borgman, C. (2010). Research data: who will share what with whom, when, and why. Paper presented at *the Fifth China-North America Library Conference*, 8-12 September, Beijing, China. Retrieved from <http://works.bepress.com/cgi/viewcontent.cgi?article%1237&context%4>.
- Carlson, J., & Garritano, J. (2010). E-science, Cyberinfrastructure and the Changing Face of Scholarship: Organizing for New Models of Research Support at the Purdue University Libraries. *Libraries Research Publications*. Retrieved from http://docs.lib.purdue.edu/lib_research/137
- Canada, G. of. (2015). Draft Tri-Agency Statement of Principles on Digital Data Management - Science.gc.ca. Retrieved from <http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1>
- Childs, S., McLeod, J., Lomas, E., & Cook, G. (2014). Opening research data: issues and opportunities. *Records Management Journal*, 24(2), 142 - 162. Retrieved from <http://dx.doi.org/10.1108/RMJ-01-2014-0005>
- Cox, A., Pinfield, S. & Smith J. (2014), Moving a brick building: UK libraries coping with research data management as a 'wicked' problem. *Journal of Librarianship and Information Science*, p 1-15. DOI: 10.1177/0961000614533717
- Cox, A., Allchin, O., Collins, A., Lewis, J. & Scott C. (2013), Realizing our role in research data management. *CILIP Update*, p 36-38.
- Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 368(1926), 4023-4038.
- Delaney, G. & Bates, J. (2014): Envisioning the Academic Library: A Reflection on Roles, Relevancy and Relationships, *New Review of Academic Librarianship*, DOI: 10.1080/13614533.2014.911194
- Doucette, L. & Fyfe, B. (2013, April 10-13), *Drowning in research data: Addressing data management literacy of graduate students*. Paper presented at the ACRL Conference, Indianapolis, IN.
- Foster, N. F., & Gibbons, S. (2005). *Understanding Faculty to Improve Content Recruitment for Institutional Repositories* (Vol. 11). Retrieved from <http://eric.ed.gov/?id=ED490029>
- Groenewald, R., & Breytenbach, A. (2011). The use of metadata and preservation methods for continuous access to digital data. *The Electronic Library*, 29(2), 236 - 248 Retrieved from <http://dx.doi.org/10.1108/02640471111125195>
- Henty, M. (2008). Dreaming of data: the library's role in supporting e-research and data management. In: *Australian Library and Information Association biennial conference*, Alice Springs, Australia, 2 – 5 September 2008. Retrieved from http://apsr.anu.edu.au/presentations/henty_alia_08.pdf.
- Higman, R., & Pinfield, S. (2015). Research data management and openness: The role of data sharing in developing institutional policies and practices. *Program*, 49(4), 364–381.
- Jantz, R. C., & Wilson, M. C. (2008). Institutional Repositories: Faculty Deposits, Marketing, and the Reform of Scholarly Communication. *The Journal of Academic Librarianship*, 34(3), 186–195. <http://doi.org/10.1016/j.acalib.2008.03.014>
- Jones, R. (2007). Giving birth to next generation repositories. *International Journal of Information Management*, 27(3), 154–158. <http://doi.org/10.1016/j.ijinfomgt.2007.02.004>

- Lercher, A. (2008). A Survey of Attitudes About Digital Repositories Among Faculty at Louisiana State University at Baton Rouge. *The Journal of Academic Librarianship*, 34(5), 408–415. <http://doi.org/10.1016/j.acalib.2008.06.008>
- Lynch, C. A. (2003). Institutional repositories: essential infrastructure for scholarship in the digital age. *portal: Libraries and the Academy*, 3(2), 327-336.
- Lyon, L. (2012). 'The informatics transform: re-engineering libraries of the data decade', *International Journal of Digital Curation*, 7(1), pp. 126-138.
- Ma, J., Wang, Y., Zhu, Z., & Tang, R. (2009). An attempt of data exchange between the institutional repository and the information environment for the management of scientific research—ARP. *Library Collections, Acquisitions, and Technical Services*, 33(1), 1–7. <http://doi.org/10.1016/j.lcats.2009.01.003>
- Maron, N., Yun, J. and S. Pickle. (N.D.). Sustaining our Digital Future: Institutional Strategies for Digital Content. http://www.sr.ithaka.org/sites/default/files/reports/Sustaining_Our_Digital_Future.pdf >
- National Science Board (2005). Long-lived digital data collections: enabling research and education in the 21st Century, National Science Foundation, US. Retrieved from www.nsf.gov/pubs/2005/nsb0540/.
- Nielsen, H.J. & Hjørland, B. (2014). Curating research data: the potential roles of libraries and information professionals, *Journal of Documentation*, 70(2), 221 - 240. (Retrieved from <http://dx.doi.org/10.1108/JD-03-2013-0034>)
- Peters, T. A. (2002). Digital repositories: individual, discipline-based, institutional, consortial, or national? *The Journal of Academic Librarianship*, 28(6), 414–417. [http://doi.org/10.1016/S0099-1333\(02\)00350-6](http://doi.org/10.1016/S0099-1333(02)00350-6)
- Pinfield, S., Cox, A. M., & Smith, J. (2014). Research Data Management and Libraries: Relationships, Activities, Drivers and Influences. *PLoS ONE*, 9(12), e114734. <http://doi.org/10.1371/journal.pone.0114734>
- Priti Jain. (2011). New trends and future applications/directions of institutional repositories in academic institutions. *Library Review*, 60(2), 125–141. <http://doi.org/10.1108/00242531111113078>
- Research Data Strategy Working Group. (2008). Stewardship of Research Data in Canada: A Gap Analysis. *Research Data Canada*. 1-31. <http://data-donnees.gc.ca/eng/index.html>
- Shearer, K. (2010). Research Data: Unseen Opportunities. Retrieved from http://www.carl-abrc.ca/uploads/pdfs/data_brochure-e.pdf
- Shearer, K. (2015). Comprehensive Brief on Research Data Management Policies.pdf. Retrieved from <http://www.science.gc.ca/1E116DB8-E7F3-4B6F-BB44-83342BAAA030/Comprehensive%20Brief%20on%20Research%20Data%20Management%20Policies.pdf>
- Shen, Y., & Varvel, V. E. (2013). Developing data management services at the Johns Hopkins University. *The Journal of Academic Librarianship*, 39(6), 552-557. <http://doi.org/10.1016/j.acalib.2013.06.002>
- Smith, K. (2012). Managing research data. *Library Management*, 33(6/7), 437 - 439. Retrieved from <http://dx.doi.org/10.1108/01435121211266438>.
- Swan, A., & Carr, L. (2008). Institutions, Their Repositories and the Web. *Serials Review*, 34(1), 31–35. <http://doi.org/10.1016/j.serrev.2007.12.006>
- Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2014). Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research*, 36(2), 84-90.
- Thompson, C. (2015, November 6-10), *Building data expertise into research institutions: preliminary results*. Paper presented at ASIST, St. Louis, MO.
- Verbaan, E., & Cox, A. M. (2014). Collaboration or Competition? Responses to Research Data Management in UK Higher Education by Librarians, IT Professionals, and Research Administrators. *In iConference 2014 Proceedings* (p. 281–292). doi:10.9776/14084