

2017

Research Data Management: A Library Practitioner's Perspective

Siu Hong Yu

Western University, syu333@uwo.ca

Follow this and additional works at: <https://ir.lib.uwo.ca/fimspub>



Part of the [Library and Information Science Commons](#)

Citation of this paper:

Yu, Siu. "Research Data Management: A Library Practitioner's Perspective." *Public Services Quarterly* 13 (2017): 48-54.

This is an Accepted Manuscript of an article published by Taylor & Francis in PUBLIC SERVICES QUARTERLY on January 23, 2017, available online:

<http://www.tandfonline.com/10.1080/15228959.2016.1223475>

FUTURE VOICES IN PUBLIC SERVICES

Nancy H. Dewald, Column Editor

The **Future Voices in Public Services** column is a forum for students in graduate library and information science programs to discuss key issues they see in academic library public services, to envision what they feel librarians in public service have to offer to academia, to tell us of their visions for the profession, or to tell us of research that is going on in library schools. We hope to provide fresh perspectives from those entering our field, in both the United States and other countries.

Siu Hong Yu is an MLIS student at the University of Western Ontario.

Research Data Management: A Library Practitioner's Perspective

SIU HONG YU

University of Western Ontario, London, Ontario, Canada

Often simply known as “the book,” a prompt book is an annotated master copy of a script that contains the information necessary to create a successful theatrical production. It lays out all the dialogues, actor positions, technical light and sound cues, as well as any other relevant details including contact lists for the cast and crew, information

© Siu Hong Yu

Comments and suggestions should be sent to the Column Editor, Nancy H. Dewald, Reference Librarian, Penn State Berks, Tulpehocken Road, P.O. Box 7009, Reading, PA 19610-6009. E-mail: nxd7@psu.edu

Address correspondence to Siu Hong Yu, [address] E-mail: syu333@uwo.ca

about the venue(s), local amenities, and emergency procedures. As post-doctoral fellow Toby Malone (2015) demonstrated at the Research Data Management Conference organized by the University of Waterloo Library, prompt-book materials are invaluable stage-managerial resources and represent the rich dramaturgical heritage of an institution of great theatre. Their digitization is a practical solution to their remote accessibility and provides tremendous opportunities for future research.

Prior to attending the data management conference, my idea of research data was basically limited to scientific data, which are mostly numerical or statistical, and usually include the description or representation of natural phenomena or technological advances. Similar to a lab book that documents the process of a chemist's scientific inquiries, Malone's presentation opened my eyes to the hidden world of stage managers' playbooks, and to what research data management (RDM) could entail in the arts and humanities. What exactly is RDM? What kinds of RDM services do researchers need? How can academic libraries best support RDM? Why should it be bothered? As a student, I certainly do not have all the answers. Hence, for this essay, I talked to three librarians and one data technician from three Canadian universities in the hope to gain a practitioner's perspective on RDM.

"Data is the new gold," states Vice President of the European Commission responsible for the Digital Agenda, Neelie Kroes (2011). With the recent emergence of RDM as a strategic priority for universities (Pryor, 2012; Whyte & Tedds, 2011), extensive literature has covered RDM in the contexts of academic libraries (Cox & Pinfield, 2014; Pryor, Jones & Whyte, 2014; Ray, 2014; Steeleworthy, 2014; Tenopir et al., 2014). In her book chapter titled *Roles and Responsibilities: Libraries, Librarians and Data*, Corrall (2012) makes a compelling case for library commitment to RDM and lays out the strategies for tactical engagement with research data. Central to RDM is the research lifecycle, where data is 1) created or received, 2) appraised and selected, 3) ingested or transferred, 4) preserved, 5) stored, 6) accessed, used and reused, and 7) transformed (Higgins, 2008). In its core, RDM concerns with "the organization of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. It aims to ensure reliable verification of results, and permits new and innovative research built on existing information" (Whyte & Tedds, 2011, p. 1).

Working as an MLIS co-op student at the University of Waterloo Library at the time, I had the opportunity to continue the pertinent conversation from the Research Data Management Conference in a subsequent journal-club meeting. Based on the lively discussion, the meaning of RDM is, as corroborated by Weller and Monroe-Gulick (2014), influenced by the research methodology and academic discipline of the researcher. With discipline-specific data management practices, academic libraries could focus their effort in point-of-need, individual consultations, as well as facilitating mentorship and peer-to-peer learning among researchers as part of the effective RDM support strategy (Carlson et al., 2015).

Recently, the organizers of the Research Data Management Conference published their survey results regarding the RDM practices and needs at the University of

Waterloo's Faculties of Engineering and Science (Szigeti & Keys, 2016). While about 50% of the respondents indicated that there was sufficient documentation and description for another person outside the research group to understand and reuse the research data or to replicate the methodologies that produced the data, close to 30% indicated there was not sufficient documentation or description, and close to 20% simply did not know. Between the raw data, working data and processed data ready for publication, the length of time research data is typically kept by researchers varied considerably. Less than half of the respondents had ad hoc procedures in managing data (procedures include file naming, backup schedule and recording data descriptions). Only about 21% of the respondents had formal procedures on a personal level and up to 13% had no procedure in data management. In terms of data sharing, while the majority of the respondents shared data by personal request (62%), some researchers also shared their data via a discipline-specific repository (22%), or as part of supplementary material to a journal publication (20%). However, up to 16% of the respondents did not share their data. Incomplete data or associated methods, insufficient time, and privacy, legal or security concerns were among the most significant reasons why researchers were not willing to share their research data.

When asked to draft a data management plan as part of a grant application, more than half of the respondents would prefer to have assistance and/or guided documentation. According to the survey, the RDM service garnering the most interest among respondents was an institutional repository for long-term access and preservation of research data, followed by assistance for preparing data management plans to meet funding requirements (Szigeti & Keys, 2016). Interestingly, the service receiving the least interest was assistance with documenting and describing data (i.e., metadata creation), followed by personalized consultation on data management practices for specific research groups or projects. In terms of methods of RDM service delivery, the majority preferred information on a website, while research group consultation, in-person workshop and online tutorials shared similar, albeit modest, interest. Webinars were the least preferred method of service delivery.

From the library's perspective, the primary need in RDM that should be addressed is education. Some faculty members are resistant to the cultural shift brought forward by RDM, while others mistakenly think that RDM means that their data must be publicly accessible by anyone (K. Szigeti, personal communication, June 12, 2016). The truth is that restricted data may be kept in a repository and that the researcher can grant permissions for data access and usage on an as-needed basis. According to Szigeti, most of the RDM questions came from the faculty (personal communication, June 12, 2016). Interestingly, if a graduate student approaches the library for RDM assistance, the student is usually referred to the library by his or her research supervisor. Questions so far are mostly about where to store active data that is being generated via collaborations between researchers at the home institution and elsewhere around the world. With respect to fulfilling journal-publishing requirements, there has been a question specifically about finding a suitable repository to provide long-term public access to a researcher's algorithm that was generated with a proprietary software (K. Szigeti, personal communication, June 12, 2016). Overall, even though there have been very few questions

regarding RDM from the researchers, the library has been proactive in communicating with the office of research and other campus partners to provide assistance in organizing and storing their data.

During a candid conversation with two data librarians working at another academic library, I learned that while a comprehensive RDM support service has not yet been in place there, a subcommittee under research support has been created (personal communication, June 10, 2016). One of the subcommittee's main purposes is to raise awareness and increase the knowledge base about RDM among librarians and library staff. Since no official RDM policy has been established at the institutional level, the library RDM subcommittee is also working with the research office from the bottom up to raise the profile of RDM, and to find out about opportunities for collaboration. With the RDM subject guide published on the library's website, the associate deans are receptive about RDM and the research office is interested to collaborate on RDM initiatives, according to the two data librarians. However, lots of questions remain in terms of budgets, scale and resource allocation, and a great deal of learning still needs to take place. There is a mixture of curiosity and apprehension within the library, where some workforce analysis, organizational restructuring and strategic repositioning are being considered due to the anticipated growing and evolving demand in RDM. While the library can certainly play a crucial role in RDM at the university, it cannot be a library's solo project. The implementation of a trusted data repository, for example, involves technical infrastructure, IT support, database experts and administrative personnel. Ultimately, robust RDM has to be done through partnerships and collaborations across campus and beyond, insisted one of the data librarians.

At a third Canadian academic library, questions regarding RDM have come from both faculty and graduate students more or less in equal numbers, observed one data technician (personal communication, June 21, 2016). Requests on RDM support have ranged from researchers at the beginning of their research in need of data management plan assistance, to retiring faculty members hoping to archive their data, and everything in between. Currently, this library addresses RDM requests from researchers on a case-by-case basis. It has been successfully archiving data from the social sciences and is starting to develop services to accept more data from the natural sciences. According to the data technician, storage requests have ranged from less than one gigabyte to terabytes of data. One of the biggest challenges for the library to develop RDM services lies in figuring out what the researchers need and meeting their broad spectrum of needs with sufficient support in terms of staff time, server space and other data-managerial resources. Similarly, a strong sense of collaborative spirit exists and the library is working with the university's research services to prepare for the upcoming demand in RDM.

Despite their differences in institutional context, the general impression I got from my conversations with the librarians and data technician agrees with the literature finding that RDM services are currently not frequently utilized in libraries and that many services are in the planning stages (Tenopir et al., 2014). Nationally, Canada is lagging behind other countries in terms of RDM policies, services and infrastructure (Guindon, 2014). A

common challenge underpinning all three academic libraries above is the uncertainties brought about by the fact that a national data management policy is currently under consultation and has not yet been established. Within Canada, research and innovation are promoted and supported by three federal granting agencies, namely, 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). As progressive as it is, the *Tri-Agency Statement of Principles on Digital Data Management* (Government of Canada, 2016) is not binding. Several research funders in Canada have adopted more explicit RDM policies while others have not (Shearer, 2015).

The need for effective and affordable RDM services will only grow over the next few years. Proper research data management entails how you are going to preserve your data, where you are going to store the data, how you are going to document the data and who is responsible for maintaining the data files. It certainly involves planning and documentation as well as the working knowledge with database, metadata and archiving. Writing a data management plan is a good start, but sticking to the plan and actually implementing it, however, can prove to be challenging. Researchers are focused and driven to advance knowledge in their respective disciplines. Because of the lack of enforcement in Canada, data management tends to be low on priority or set aside, acknowledged one data librarian (personal communication, June 10, 2016). A significant cultural shift, where the scholarly communication impact of data management and sharing is better merited for tenure and promotion purposes, is called for. Libraries may not be able to drive that cultural change, but they can certainly facilitate some of the underlying RDM process.

Institutional support and direction are very important in moving forward. In the short term, Cox and Pinfield (2014) suggest that the priority for most academic libraries should be to develop an RDM policy. They should then focus on RDM advising and training (especially among early-career researchers and graduate students), followed by the involvement in an institutional repository. On a personal level, librarians will need to address our skills gaps, workload issues due to shrinking resources, and lack of confidence (Cox & Pinfield, 2014). However, in recognizing our existing networks within an institution, our expertise in organizing and managing information, as well as our complementary roles in promoting open access and information literacy, we are well positioned to support RDM (Cox, Verbaan & Sen, 2012). As Ray (2014) points out, the greatest contribution of libraries in RDM is our emphasis on services, “providing the basis not only for future access to digital assets, but also for assistance to data creators in managing their own active data” (p. 6). “Memory records the world as so. Imagination transposes it into the key of as if, transforming experience into speculation. That is why to lose one’s memory means losing the future. Because imagination is memory in the future tense,” argues Rumsey (2016, p. 127). By analogy, all these research data are like the memory of our inquiries. Without proper data management, it is really our future that is at stake.

Acknowledgement

The author would like to sincerely thank Francine Berish, Alexandra Cooper, Vincent Gray, Elizabeth Hill and Kathy Szigeti for their time, insights and data sharing during the research and writing of this article.

References

- Carlson, J., Nelson, M. S., Johnston, L. R., & Koshoffer, A. (2015). Developing data literacy programs: Working with faculty, graduate students and undergraduates. *Bulletin of the Association for Information Science and Technology*, 41(6), 14-17. doi: 10.1002/bult.2015.1720410608
- Corrall, S. (2012). Roles and responsibilities: Libraries, librarians and data. In G. Pryor (Ed.), *Managing research data* (pp. 105-133). London, England: Facet Publishing.
- Cox, A. M., Verbaan, E., & Sen, B. (2012). Upskilling liaison librarians for research data management. *Ariadne*, 70. Retrieved from <http://www.ariadne.ac.uk/issue70/cox-et-al>
- Cox, A. M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, 46(4), 299-316. doi: 10.1177/0961000613492542
- Government of Canada. (2016). *Tri-agency statement of principles on digital data management*. Retrieved from <http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1>
- Guindon, A. (2014). Research data management at Concordia University: A survey of current practices. *Felicitier*, 60(2), 15-17. Retrieved from http://cla.ca/wp-content/uploads/60_2.pdf
- Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation*, 3(1), 134-140. doi:10.2218/ijdc.v3i1.48
- Kroes, N. (2011, December). *Data is the new gold. Opening remarks given at the Press Conference on Open Data Strategy*, Brussels, Belgium. Retrieved from http://europa.eu/rapid/press-release_SPEECH-11-872_en.htm
- Malone, T. (2015, October). *Managing dramaturgical data: The Stratford Festival prompt-books*. Paper presented at Research Data Management Conference 2015, University of Waterloo, Ontario. Abstract retrieved from <http://subjectguides.uwaterloo.ca/uwrmdm2015>
- Pryor, G. (2012). Why manage research data? In G. Pryor (Ed.), *Managing research data* (pp. 1-16). London, England: Facet Publishing.
- Pryor, G., Jones, S., & Whyte, A. (Eds.). (2014). *Delivering research data management services: Fundamentals of good practice*. London, England: Facet Publishing.

- Ray, J. M. (Ed.). (2014). *Research data management: Practical strategies for information professionals*. West Lafayette, IN: Purdue University Press.
- Rumsey, A. S. (2016). *When we are no more: How digital memory is shaping our future*. New York, NY: Bloomsbury Press.
- Shearer, K. (2015). *Comprehensive brief on research data management policies*. Retrieved from <http://www.science.gc.ca/default.asp?lang=En&n=1E116DB8-1>
- Steeleworthy, M. (2014). Research data management and the Canadian academic library: An organizational consideration of data management and data stewardship. *Partnership: The Canadian Journal of Library and Information Practice and Research*, 9(1), 1-11. doi: 10.21083/partnership.v9i1.2990
- Szigeti, K., & Keys, S. (2016). *Research data management faculty survey report, University of Waterloo* [Data file and analysis]. Retrieved from Waterloo Research Data Management Dataverse at <http://hdl.handle.net/10864/11519>
- Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2014). Research data management services in academic libraries and perceptions of librarians. *Library & Information Science Research*, 36(2), 84-90. doi: 10.1016/j.lisr.2013.11.003
- Weller, T., & Monroe-Gulick, A. (2014). Understanding methodological and disciplinary differences in the data practices of academic researchers. *Library Hi Tech*, 32(3), 467-482. doi: 10.1108/LHT-02-2014-0021
- Whyte, A., & Tedds, J. (2011). *Making the case for research data management*. DCC Briefing Papers. Edinburgh: Digital Curation Centre. Retrieved from <http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm>