



The New England Collaborative Data Management Curriculum Pilot at the University of Manitoba: A Canadian Experience

Mayu Ishida

University of Manitoba, Winnipeg, Manitoba, Canada

Abstract

Canada's federal funding agencies are following the directions of funding agencies in the United States and United Kingdom, and will soon require a data management plan in grant applications. The University of Manitoba Libraries in Canada has started planning and implementing research data services, and education is seen as a key component. In June 2014, the New England Collaborative Data Management Curriculum (NECDMC) (Lamar Soutter Library, University of Massachusetts Medical School 2014) was piloted and used to provide data management training for a group of subject librarians at the University of Manitoba Libraries, in combination with information about

data-related policies of the Canadian funding agencies and the University of Manitoba. The seven NECDMC modules were delivered in a seminar style, with emphasis on group discussions and Canadian content. The benefits of NECDMC – adaptability and flexible framework – should be weighed against the challenges experienced in the pilot, mainly the significant amount of time needed to create local content and complement the existing curriculum. Overall, the pilot showed that NECDMC is a good, thorough introduction to data management, and that it is possible to adapt NECDMC to the local and Canadian settings in an effective way.

Introduction

Canada's federal funding agencies are following the directions of other funding agencies in the United States and United Kingdom, and are expected to require a data management plan (DMP) in grant applications in the near future. A DMP is a supplementary document that describes how research results -- including datasets -- will be disseminated and shared in the proposed research program. At the University of Manitoba Libraries in Canada, we have started planning and implementing research data services, and education is seen as a key

component. In anticipation of the new requirement, we hope to assist our researchers with data management plans and to prepare our librarians to address data management questions. In June 2014, the New England Collaborative Data Management Curriculum (NECDMC) (Lamar Soutter Library, University of Massachusetts Medical School 2014) was piloted and used to provide data management training for a group of subject librarians at the University of Manitoba Libraries, in combination with information about data-related policies of the Canadian funding agencies and the University of Manitoba. The Lamar Soutter Library of

Correspondence to Mayu Ishida: mayu.ishida@umanitoba.ca

Keywords: data management, instruction, NECDMC, Canada, University of Manitoba

University of Massachusetts Medical School in the United States has been leading the development of NECDMC with an increasing number of partners.

Requirements of Canada's Federal Funding Agencies

Canada has three federal funding agencies: 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). Together, they are referred as the Tri-Agencies, and currently none of them requires a DMP in their grant application processes. However, CIHR and SSHRC do have policies about depositing and archiving data sets in the reporting process. CIHR requires open access to bioinformatics, atomic, and molecular coordinate data from CIHR-funded projects at the publication of research findings (Canadian Institutes of Health Research 2012, 5.1.2). SSHRC requires that datasets from SSHRC-funded projects be preserved and made available to others within two years after the project completion (Social Sciences and Humanities Research Council of Canada 2012).

The Tri-Agencies are currently coordinating their policies about sharing and publishing research results, and will introduce the new Tri-Agency Open Access Policy in fall 2014 (Government of Canada Tri-Agencies 2013). In its draft, the new policy states that all peer-reviewed journal articles resulting from federally funded research must be made freely accessible within 12 months of publication (3.1). Furthermore, the SSHRC Executive Vice President stated in June 2014, "The Tri-Agencies and CFI [Canada Foundation for Innovation], in collaboration with Genome Canada, are proposing to develop a new policy, following developments in other countries, which would require all grant applications to include data management plans" (CIHR, NSERC, and SSHRC 2014, 15). It is anticipated that the Tri-Agencies'

DMP policy will be introduced as early as fall 2015, after its Open Access Policy takes effect.

Institutional Policies at the University of Manitoba

Since the Tri-Agencies plan to require DMPs in the near future, academic institutions in Canada are to prepare their faculty and students for the requirement, and to develop policies for data ownership, storage, and retention. At the University of Manitoba (UM), a working group was recently formed to develop institutional guidelines for data management, and a pilot of a secure file sharing system was conducted to offer UM researchers an alternative to free (but not secure) external services.

UM has not established data management guidelines, but it does have policies related to research data. The UM faculty member is granted the copyright of the work she has produced during her regular duties under the collective agreement (University of Manitoba and University of Manitoba Faculty Association 2011, Article 14), although it is not clear whether research data is a type of intellectual property covered under the agreement. On the other hand, the student should discuss with his supervisor(s) and agree upon the authorship of raw data from his thesis-related research before starting a degree program (University of Manitoba Technology Transfer Office 2013, 2.11). If the student creates a database (i.e., data organized in an original format) from raw data, the copyright of the database is granted to the student (2.11).

The UM Office of Research Ethics and Compliance (OREC) suggests research IT practices including those of data backup and security; UM-affiliate researchers are encouraged to back up data frequently in multiple locations, and to avoid sharing data through systems such as Dropbox, Google Drive, and Gmail that are not controlled by the University (Mazak 2013, 2-3). Regarding data

retention, OREC states that “[c]onfidential data should be destroyed or rendered anonymous as soon as it is no longer necessary scientifically to link data with individual participants”, and that “[a]nonymous data may be kept indefinitely” (University of Manitoba Office of Research Ethics and Compliance 2014, 5) while the Health Canada Food and Drugs Regulations sets the retention period of clinical data at 25 years (Health Canada 2004, C.05.012 (4)). Moreover, the administrative and financial records of a research grant should be retained for seven years after the grant completion (University of Manitoba Office of Fair Practices and Legal Affairs 2013, FINANCE-020).

NECDMC Pilot at the University of Manitoba

As the research services librarian, this author facilitated the NECDMC pilot at UM, providing data management training for eight subject librarians from the STEM fields. This group of participants was chosen because the departments in which they serve as liaisons are highly interested in data management. Additionally, since NECDMC was first developed for the STEM fields, the curriculum already included research cases related to the participants’ subject areas. The purpose of the pilot was twofold: 1) To train the group of subject librarians in data management so that they would be equipped with enough knowledge to address a basic level of data management questions from faculty and students, and 2) To examine the NECDMC instructional materials to see how the materials could be adapted to our local and Canadian settings.

In the pilot, we covered the seven NECDMC modules. Module 1 gives an overview of data management issues and introduces a DMP framework, while the remaining modules (Modules 2-7) describe the six major aspects of a DMP:

- Module 1: Research Data Management Overview (major aspects of data management)
- Module 2: Types, Formats, and Stages of Data (what type of data will be created, how the data will be captured and processed)
- Module 3: Metadata (contextual details needed to make data meaningful to others)
- Module 4: Data Storage, Backup & Security (where and on what media the data will be stored, how the data will be backed up, how data security will be managed)
- Module 5: Legal & Ethical Consideration (how privacy and ethical issues in data will be addressed, who owns copyrights and other intellectual property rights regarding data)
- Module 6: Data Sharing & Reuse (what type of restrictions will be placed on data re-use, how access to data will be granted)
- Module 7: Repositories, Archiving & Preservation (what is the long-term plan for preserving and maintaining the data)

For each module, NECDMC provides a lesson plan, lecture slides, learning activities, and research cases on which the learning activities are based. For the pilot, the facilitator modified the existing materials to include local content, and also moved content about data documentation from Module 2 to Module 3 to cover metadata issues in one module. Before Module 1, the facilitator created “Module 0: Introduction” as an icebreaker to explain the pilot purpose and to ask the participants to share their data management experiences and questions.

The facilitator delivered the seven modules in a seminar style, and emphasized discus-

sions based on research cases from NECDMC. Since we aimed to study the NECDMC materials and activities in depth, we scheduled a three-hour session for each module. The participants gave their feedback on each module by completing an evaluation form provided by the Lamar Soutter Library; the feedback will be used internally to offer better training in the future. Each session consisted of:

- Lectures (PowerPoint slides, Q&A): 60-90 min
- Learning activities (group discussions): 45-60 min
- Module evaluation: 5-10 min
- Break: 15-20 min

One to two sessions were delivered on one day, and it took five days over four weeks to complete the pilot.

The major challenge of this pilot was the amount of time required by both the facilitator and the participants. Each module was given three hours, and seven modules were covered. Each participant was asked to commit to 21 hours in class over four weeks to the pilot, whereas the participant was not required to spend time outside class because no homework was given. In addition to the 21 hours in class with the participants, the facilitator spent approximately 30 hours preparing slides, in-class activities, and local content as well as managing the pilot logistics.

Adapting NECDMC to the Local and Canadian Settings

While offering a comprehensive coverage of data management concepts and a variety of research cases from the STEM fields, NECDMC did not include information about data policies and regulations applicable to researchers in Canada. In the UM pilot, the US policies and regulations cited in NEC-

DMC were compared with Canadian counterparts to provide local and relevant information to the pilot participants. In addition to the policies of the National Science Foundation and the National Institute of Health mentioned in NECDMC, the policies of Canada's federal funding agencies were discussed in the pilot (cf. "Requirements of Canada's Federal Funding Agencies"). Regarding health science data, NECDMC refers to the US Health Insurance Portability and Accountability Act of 1996 (HIPPA) that concerns the management of personal health information. As the Canadian counterpart to HIPPA, the Health Canada Guidance for Records Related to Clinical Trials (federal regulation) and the Manitoba Personal Health Information Act (provincial regulation) were cited in the pilot.

In addition to information about the Canadian policies and regulations, UM institutional policies relevant to data management were studied in the pilot. It was challenging to locate these institutional policies because they were under different offices at UM. The facilitator was also a new UM employee and not yet familiar with the functions and policies of these offices. She gained the knowledge of the institutional structure and policies by searching the UM website and speaking with colleagues. She then reviewed and summarized each of the identified policies (cf. "Institutional Policies at the University of Manitoba").

It took approximately 10 hours to research the institutional policies as well as the regulations and funding agency policies. Although it was time-consuming, the facilitator was able to provide content local and relevant to the pilot participants.

Reflections of the Pilot Facilitator

Due to the small number of the pilot participants and the large amount of time assigned to the pilot, the facilitator was able to include two to three active learning strategies in each module such as Think-Pair-Share ses-

sions (Allen 2014) and group discussions based on NECDMC research cases. The participants come from similar backgrounds (subject librarians in the STEM fields), and the research cases selected for the pilot were relevant to their subject areas. Furthermore, they had known each other as colleagues prior to the pilot. These factors probably contributed to the participants' interactions and reflections in class, in addition to the use of active learning strategies.

To make NECDMC relevant to the participants' liaison roles, the facilitator asked how the NECDMC materials could be adapted to their instruction settings. Since library instruction is usually given in one-shot sessions, we discussed how Module 1 (Research Data Management Overview) would be modified into such an instruction session to introduce data management issues. In particular, the short video from Module 1 (*Data Sharing and Management Snafu in 3 Short Acts* 2012) would be used as a fun introduction for students who do not have prior data management knowledge.

While it is beneficial that each NECDMC research case describes the context of data management issues in details, reading and comprehending the details in class took ten to fifteen minutes for certain research cases. The reading could have been assigned as homework, in order to have more time discussing data management solutions in class. Since the research cases did not come with an "answer key", the facilitator tried her best to provide model answers, but would have appreciated such an answer key provided with the curriculum.

Further training in data management should be provided to meet the library staff's needs. Nonetheless, to become more familiar and confident with the data management practices, it is helpful to start applying the practices. For example, we can start simple with conventions for file and directory naming. Since UM librarians are strongly encouraged to undertake research, they have an opportunity to apply the practices in their research and

scholarly communication. If they can build data management skills in their own research, then they will likely gain confidence in assisting faculty and students with data management.

Conclusions

Because of NECDMC's flexible framework, it was relatively easy to add local content and adapt the existing curriculum to the local and Canadian settings. Nonetheless, the benefit of NECDMC – adaptability – should be weighed against the challenges experienced in the pilot. While the facilitator was able to avoid creating instructional materials from scratch and save time by using NECDMC, it still took her a significant amount of time to create local content. Furthermore, it required a serious time commitment from the participants to cover the seven NECDMC modules. As each NECDMC module is self-contained, data management trainers can better accommodate the time constraints and learning needs of learners by teaching select NECDMC modules or offering the concise versions of the modules.

Based on the NECDMC pilot experience, the UM Libraries are planning to extend data management training to librarians in the health sciences, social sciences, and humanities. It is expected to take less time to prepare for the next round of training because the local and Canadian content was created in the first round/pilot. Depending on the time constraints and learning needs of learners, only select NECDMC modules will be taught in the second round. To complement NECDMC with more local content, guest speakers may be invited from the University offices relevant to data management (e.g., grants, ethics, legal councils, copyright, records management, IT). Moreover, this author customized the DMP template from NECDMC with references to the UM policies (Ishida 2014) and used it to illustrate major aspects of data management in her scholarly communication workshop for graduate students. She plans to use the custom

DMP template in her future training. Overall, the pilot of data management training for librarians at UM showed that NECDMC is a good, thorough introduction to data management, and that it was possible to adapt NECDMC to the local and Canadian settings in an effective way.

References

Allen, Cathy. 2014. "Using the Think-Pair-Share Technique." *Read Write Think by the International Reading Association and the National Council of Teachers of English*, <http://bit.ly/1aJ4BNb>.

Canadian Institutes of Health Research. 2012. "CIHR Open Access Policy." March 12. <http://bit.ly/1yWZsKX>.

CIHR, NSERC, and SSHRC. 2014. "Update on Tri-Agency Programs, Initiatives and Policies." CAURA National Conference. <http://bit.ly/1xpJ1cF>.

Hanson, Karen, Alisa Surkis, and Karen Yacobucci. *Data Sharing and Management Snafu in 3 Short Acts*. 2012. <http://bit.ly/1yvt0lo>.

Government of Canada Tri-Agencies. 2013. "Draft Tri-Agency Open Access Policy." October 15. <http://bit.ly/1wtXyXm>.

Health Canada. 2004. "Guidance for Records Related to Clinical Trials (GUIDE-0068)." Guide. October 25. <http://bit.ly/ZWNky8>.

Ishida, Mayu. 2014. "Data Management Plan Template - Subjects at University of Manitoba." Accessed October 11. <http://bit.ly/1F4bfco>.

Lamar Soutter Library, University of Massachusetts Medical School. 2014. "New England Collaborative Data Management Curriculum." Accessed August 18. <http://bit.ly/1uWtNKy>.

Mazak, Trevor. 2013. "University of Manitoba Research Suggested Practices for IT." University of Manitoba Office of Research Ethics and Compliance. <http://bit.ly/1sxzEGK>.

Social Sciences and Humanities Research Council of Canada. 2012. "SSHRC Research Data Archiving Policy." May 11. <http://bit.ly/1Hqb8MI>.

University of Manitoba Office of Fair Practices and Legal Affairs. 2013. "University of Manitoba Common Records Schedule." <http://bit.ly/1zGpkM9>.

University of Manitoba Office of Research Ethics and Compliance. 2014. "University of Manitoba Fort Garry Campus Research Ethics Board Submission Form." Accessed August 19. <http://bit.ly/1vlyTgS>.

University of Manitoba Technology Transfer Office. 2013. "University of Manitoba Intellectual Property Policy." <http://bit.ly/1pvO4zk>.

University of Manitoba, and University of Manitoba Faculty Association. 2011. "University of Manitoba - University of Manitoba Faculty Association 2010-2013 Collective Agreement." <http://bit.ly/1Hqc7Mn>.

Disclosure: The author reports no conflicts of interest.

All content in *Journal of eScience Librarianship*, unless otherwise noted, is licensed under a Creative Commons Attribution 4.0 International License.

<http://creativecommons.org/licenses/by/4.0/>

ISSN 2161-3974