



---

## Lessons Learned From a Research Data Management Pilot Course at an Academic Library

Jennifer Muilenburg, Mahria Lebow, Joanne Rich

University of Washington, Seattle, WA, USA

### Abstract

#### *Setting and Objective*

From January-March 2014, three librarians from the University of Washington (UW) taught a course in research data management as a pilot for the New England Collaborative Data Management Curriculum (NECDMC). The goals of the workshops were to a) pilot the NECDMC curriculum to see how effective it was as an out-of-the box solution for teaching research data management (RDM), and b) to gauge interest in an RDM class among certain UW student populations, and c) to teach UW's first RDM workshop offered to non-librarians.

#### *Design and Methods*

The NECDMC consists of 7 modules that can be taught independently or as a series. UW decided to teach all seven modules consecutively, as one-hour long weekly workshops. Each module included a lecture and activity or discussion. We taught at one location on upper campus, and live-streamed

the lecture to another location in the Health Sciences Library. Each module was assessed at the end of the class.

#### *Results*

Interest in a research data management class is high; however, retention for a non-credit, 7-week class is low. Individual assessments show that students thought the content was important and well-delivered.

#### *Conclusions*

Based on registration, graduate students at UW in many disciplines are interested in learning research data management skills. A non-credit, 7-week class had low retention; another type of class structure might increase retention. The NECDMC curriculum is an excellent framework, but modification to individual modules are necessary to provide a thorough and localized curriculum specific to one institution.

---

### Introduction

In fall 2013, librarians at the Lamar Soutter Library at the University of Massachusetts Medical School put out a call for participants to pilot the New England Collaborative Data Management Curriculum (NECDMC) (<http://library.umassmed.edu/necdmc/join>), which was developed by UMass and several other

libraries in the New England Region (the full list is available online, <http://library.umassmed.edu/necdmc/index>). The curriculum was designed "to address universal data management best practices for health sciences, science, and engineering students at the undergraduate and graduate levels" (<http://library.umassmed.edu/necdmc/index>, accessed 4/23/14). The Data Services

---

**Correspondence to** Jennifer Muilenburg: [jmuil@uw.edu](mailto:jmuil@uw.edu)

**Keywords:** research data management curriculum, academic libraries, instructional design

Curriculum and Communications Librarian at the University of Washington (UW) had been in the process of developing research data management (RDM) modules for library staff, and was working on broadening the content from a top-level introduction to a more in-depth approach. When the call for pilot participation was announced, UW data librarians decided to participate as a pilot site and offer the class to students instead of librarians, with the thought that a later iteration would be created for library staff.

The University of Washington is a public research university located in Seattle, WA, with a 2014 enrollment of about 44,000. The UW Libraries employ approximately 115 librarians, 1.5FTE of whom were dedicated to research data management at the time of the pilot. The Data Services Curriculum and Communications Librarian worked with the National Network Libraries of Medicine, Pacific Northwest Region, Regional Technology Coordinator to develop the NECDMC pilot.

## Background

In 2009, UW Libraries Administration appointed a planning group to evaluate the need for data management support on campus. Based on that report, a Data Services Coordinator was hired, a Data Services Team was formed in 2010, and the Libraries Data Services Program was born. The original focus was on outreach and providing support to researchers for new data management planning requirements from funding agencies. Based on the results of a 2012 survey and interviews of UW researchers to identify priorities for data management support, a .5 FTE Data Curriculum and Communications Librarian was hired to provide instruction and marketing surrounding the services provided by the Libraries. In 2014, a proposal for an institutional data repository was approved, and a Repository Librarian and two graduate students were hired, and the Data Services Program officially became the Data Services Unit.

## Content of Course

There are seven modules to the NECDMC:

1. Overview of Research Data Management
2. Types, Format, and Stages of Data
3. Contextual Details Needed to Make Data Meaningful to Others
4. Data Storage, Backup, and Security
5. Legal and Ethical Considerations for Research Data
6. Data Sharing and Reuse Policies
7. Archiving and Preservation

Modules could be taught as stand-alone sessions or used to provide a framework for a series. Each module in the curriculum came with a lesson plan, a PowerPoint presentation, supporting notes used to create the PowerPoint, and one to three content-related activities. Additionally, there were a number of research case studies from a range of disciplines provided in the curriculum that can be drawn on for use in the module-based activities.

## Structure of Course

UW librarians designed a seven-week course around the seven modules. The structure of each weekly session was to begin each class with questions and a lecture (30 minutes); then activities and discussion (30 minutes). Lectures took place in two locations: the primary location on UW's main campus, where lectures were delivered, and streamed via Adobe Connect to a satellite location in the UW Health Sciences Library (HSL). The primary location was in the undergraduate library, with two librarians on hand: one delivering the lecture and activity, the other handling the technology and questions from the satellite location. At the remote site at HSL, there was both a health sciences librarian and a student technology assistant on hand. Lectures were viewed via Adobe Connect, then the livestream was turned off. Activities and discussion were administered independently at both locations. The decision to have two concurrent

**Table 1:** Departments of registrants.

Biology	4
Engineering	6
Epidemiology	3
Information School	7
Librarians	7
Medicine	2
Nursing	7
Other Health Sciences	3
Public Health	5
School of Environmental and Forest Sciences	7
Undergraduates	2
Other	25

sessions was made due to the size of the campus, and a desire to reach students from as many disciplines as possible.

The session began the third week of the winter quarter, with the intent that students would be settled into the new quarter, and with class starting before midterms and completed before finals.

**Recruitment**

The instructing librarians developed email marketing for the course (Appendix A), which was given to liaisons to adapt and distribute to the College of Engineering, the College of the Environment, the Center for Social Science Computation and Research, Health Sciences, and the Information School. Librarians began sending out emails in mid-December, and within a few days registration exceeded the capacity of the designated classroom (30 people). Registration continued to grow, was capped at 78 students and an alternative room was chosen that would accommodate more students. Given the overwhelming response, the decision was made to divide the groups

into two locations, with the larger group attending class in the undergraduate library on campus, and a satellite group viewing a live-streamed lecture at the Health Sciences Library (HSL). In addition to promotional emails, HSL also distributed print and electronic flyers, and presented information to the HSL Graduate & Professional Student Library Advisory Committee.

Registrants were primarily students, but included 2 health sciences faculty and 2 health sciences staff, as well as two additional UW Libraries staff. However, the vast majority of the 78 people who responded to the original email were graduate students; 2 identified as undergraduates, and 7 were employed by the UW Libraries. Not every student had an identified department, but a partial breakdown is shown in Table 1.

**Course Technology**

Live streaming of the PowerPoint slides and lecture audio between the two campuses was accomplished using Adobe Connect. This software also enabled real-time chat

**Table 2:** Content of modules: original, revised, and proposed.

	NECDMC module	UW Pilot Module	Proposed Future Modules
Session 1	Overview of Re- search Data Manage- ment	Overview of Research Data Management	1. Overview of RDM <i>(optional)</i>
Session 2	Types, Formats and Stages of Data	Types, Formats and Stages of Data	2. Data: types, formats, stag- es and metadata
Session 3	Contextual Details Needed to Make Data Meaningful to Others	Contextual Details Needed to Make Data Meaningful to Others (aka, Metadata)	<i>(combined with above)</i>
Session 4	Data Storage, Backup and Security	Data Storage, Backup and Security	3. Data security, legal and ethical considerations
Session 5	Legal and Ethical Considerations for Research Data	Legal and Ethical Con- siderations for Research Data	<i>(combined with above)</i>
Session 6	Data Sharing & Re- use Policies	Data Sharing & Reuse Policies; Archiving and Preservation	4. Data Sharing & Reuse; Archiving and Preservation
Session 7	Repositories, Archiv- ing and Preservation	Creating a Data Man- agement Plan with DMPTool, Introduction to EZID and ORCID	<i>(combined with above)</i>

between the librarians at both locations and video capture of the lecture.

The classroom management software Canvas was used to create an accompanying site for the course. On the course page we were able to post the syllabus and all the materials used for the class: reading lists, lecture content and PowerPoint slides, audio recordings of the lectures, homework, and case studies were all accessible. For this pilot project we did not use any features of Canvas to interact with students enrolled in the course.

### Lectures

The lectures as provided by NECDMC are intended to be used out-of-the-box. Each module includes a lecture via PowerPoint, an accompanying text document with background information used to develop the

slides (frequently from more than one contributor), example case studies, data management plans, and associated activities designed to illustrate the module's content.

The co-teachers of the UW course decided the seven-week course would be best served by enhancing the provided NECDMC content by editing for importance, clarity, cohesion, local emphasis and convention, and institutional procedures. The first module is a broad overview of research data management concepts, and as such did not need significant editing, but the following six modules were each edited by one or more of the instructors.

Editing for clarity of content required several hours of work for each module. For each class session, one of the two teaching librarians took the lead, with the other helping with revisions and serving as technical support

during class time. Time needed to revise each module varied, but a minimum of three hours and a maximum of six hours were needed for each module. Additionally, for some modules (metadata, data storage, legal, and ethical considerations), local experts were brought in to consult on the content, and in some cases suggested additions, subtractions, and other modifications. This allowed the lectures to be customized for maximum effectiveness from both a timeliness and local perspective. Local experts were also in attendance for two sessions, so they could participate in the lecture and answer questions from those in attendance.

Overall, a significant amount of time was spent on the modification of each module. Although the lectures are intended to be used out-of-the-box, edits for currency and continuity between modules and local information enhance the effectiveness of the lecture series. This is in no way meant to minimize the helpfulness of the NECDMC content; it is a well-organized curriculum that is suited to wide use. The customization piece is a large and time-consuming one, however, and other instructors using the curriculum should take that time into consideration when planning to teach beyond Module 1. Examples of the UW modifications are seen in Table 2.

During the last weeks of the pilot, the instructors realized the last two sessions (data sharing and reuse, and repositories, archiving and preservation) were both the lightest in terms of out-of-the-box content, and were so closely related thematically that they could be easily combined. The two were combined into one lecture, which left the final week's session open for a hands-on period where students explored some of the tools discussed during the lectures, such as EZID, ORCID, and DMPTool.

### **Readings & Hands-on Activities**

NECDMC provides a list of reading materials and videos to accompany each module.

Since this was a non-credit class, however, the instructors wanted to keep the homework load light. To do this, each reading that was suggested in the curriculum was examined, and instructors selected the most concise and current readings and/or videos. Additional readings or substitutions were included.

The instructors also utilized the exercises and case studies from the provided materials. However, the NECDMC modules are intended to be presented during a 90 minute session. Since this class was designed to have 60 minute sessions, the instructors pared down the number of exercises to just 1 or 2 during each class period. For the case studies used to complete the exercises, the teachers selected from NECDMC the ones we thought were the most easily understood by students from a wide range of disciplinary backgrounds. The same case studies were used to illustrate concepts in more than one module, figuring the students would already have command of what happened in the case study, and could save class time by not having to read new material.

### **Successes**

The Canvas site created for the class proved worthwhile, providing a good central location for all class materials. Holding the class concurrently at two sites generally went well, and live-streaming of the lecture was effective for 6/7 of the sessions. This is no small feat, considering we were our own for technical support for software and were in charge of many pieces of hardware (two computers, projectors, campus WIFI, microphone) and software (Adobe Connect, PowerPoint, Google Docs).

One of the highlights of the pilot was the discussion time amongst the students. They were able to learn from one another's past experiences, share strategies they are currently using for various data management related tasks, discuss priorities in the research data lifecycle, and share stories of

**Table 3:** Class attendees.

Date	Attendees @ OUGL	Attendees @ HSL	Total Attendees
1/22/2014	14	15	29
1/29/2014	13	9	22
2/5/2014	9	2	11
2/12/2014	8	10	18
2/19/2014	3	4	7
2/26/2014	3	3	6
3/5/2014	4	2	6

RDM gone wrong. The discussions also provided the librarians with insight into the data management issues that interested or troubled the students the most, thus providing material for improving upon the curriculum.

Overall, student feedback surveys indicated medium-to-high satisfaction with all of the sessions. The greatest number of suggestions emerged around the sessions on metadata and security, with students asking for more content.

### Issues & Problems

One entire lecture did not stream or record audio. Without dedicated technical support from someone who was not also teaching the class, extra time was needed before each session to make sure everything was going to run.

Retention was another issues for the class. As previously stated, 78 students registered for the class by the end of December 2013. By mid-January 2014, when the class began, 35 students attended the first session (combined between the two locations). Subsequent attrition was a problem (Table 3). This retention issue follows pretty closely the retention for MOOC's: Educause states that around 5% of students who enroll for a MOOC actually complete the class (<http://www.educause.edu/ero/article/retention-and-intention-massive-open-online-courses-depth-0>, accessed 7/14/14). Although this

class wasn't a traditional MOOC, multiple students expressed that they believed they didn't have to come to class because the materials were online. The intention with Canvas was that those materials would be supplemental to in-class learning, but it was discovered that that is not how the students perceive a Canvas site. A minimum of 10 students would better foster conversation around personal experience, the case studies, and Q&A on lecture topics.

### Suggestions for Improvement

The one lecture that would not stream threw the offsite instructors into a minor scramble. Luckily, all documents were shared on a google docs site so the remote instructors could access the materials. Since only one presenter was experienced in the content delivery, it left the entire remote class session without a lecture. Ideally, all instructors should be conversant with the lecture material.

There were also redundancies among some of the module lectures. While believing that repetition is a good thing, instructors also decided to combine two lectures (storage and backup, and archiving and preservation) into one lecture. For future iterations, additional combinations would work well, such as combining modules 2-7 (skipping the introductory material) into three, 90-minute sessions, potentially with a fourth for hands-on time with some of the tools described in the class.

## Future Plans

UW will definitely offer the course again, likely in different formats, including another session for graduate students (reduced to three 90-minute sessions), as well as a session geared toward librarians at UW, with a focus on case studies and DMP creation.

For all future sessions, instructors would like to see better retention, which will require some effort on the part of the staff. One way to do this would be to offer the class as a one-credit course through a department with broad reach such as the graduate school. Additional effort could also be put into marketing, which would not only get the word out that the libraries offer help in data management education, but hopefully increase enrollment.

HSL is also experimenting with developing online multimedia content on discrete topics of higher interest to the health sciences community to supplement the curriculum. Examples of specialized topics include human subjects perspectives and actual case studies from clinical research. Because of its widely dispersed community of users, future HSL curricula will likely be provided online only.

## Conclusion

There are many RDM curriculum plans available online, several of which are annotated on the UW Research Data Management Libguide, <http://guides.lib.washington.edu/rdm>. A quick look at these curricula shows that there is significant overlap in content. Rather than starting a curriculum from scratch, it seems prudent to use existing curricula as a starting point. This allows instructors to save time by using a pre-existing course as a backbone, and by giving them time to customize material to their particular audience and university. The NECDMC curriculum provides a strong foundation for this type of customization, and a growing community of users with which to share experience.

## Electronic Content

Appendices A and B

An online supplement to this article can be found at <http://escholarship.umassmed.edu/jeslib/vol3/iss1/> under "Appendices A and B".

*Disclosure:* The authors report no conflicts of interest.

All content in Journal of eScience Librarianship, unless otherwise noted, is licensed under a Creative Commons Attribution-Noncommercial-Share Alike License <http://creativecommons.org/licenses/by-nc-sa/3.0/>

ISSN 2161-3974