JESLIB 2012; 1(1): 46-50
doi:10.7191/jeslib.2012.1003

Journal of eScience Librarianship
putting the pieces together: theory and practice

Teaching Research Data Management: An Undergraduate/Graduate Curriculum

Mary Piorun,[1] Donna Kafel,[1] Tracey Leger-Hornby,[2] Siamak Najafi,[2] Elaine R. Martin,[1] Paul Colombo,[3] Nancy LaPelle[3]

[1] University of Massachusetts Medical School, Worcester, MA, USA;
[2] Worcester Polytechnic Institute, Worcester, MA, USA;
[3] Consultant

Abstract

With funding from the Institute of Museum and Library Services, the Libraries of the University of Massachusetts Medical School and Worcester Polytechnic Institute collaborated on a plan to expand the scope of science library practices and promote among medical, graduate, and undergraduate science students the preservation of scientific data in relevant repositories and archives. This paper outlines curriculum frameworks and learning needs for research data management instruction that can be delivered through a variety of methods. Individual modules are based on faculty and student interviews, as well as a comprehensive literature review.

Background

Student research projects are part of the core educational missions at both the University of Massachusetts Medical School (UMMS) and Worcester Polytechnic Institute (WPI). Both institutions struggle to manage massive data sets generated by students. The need to manage large data sets is well documented (Carlson 2006, Doctorow 2008, National Science Board 2005), and librarians are working to apply their skills in information management to this issue. The two libraries came together, with funding from the Institute of Museum and Library Services and the National Library of Medicine, to develop the preliminary tools for a full data management curriculum of open access educational tutorials that can be woven into any science, health science, or engineering curriculum. This article outlines the development of the curriculum, a curriculum framework, and competencies for instructing students on how to incorporate data management into the research process.

Curriculum Development

An education committee with representatives from the library, information technology, and the faculty from each campus was assembled to oversee the development of the curriculum. National experts on data manage-

Correspondence to Mary Piorun: mary.piorun@umassmed.edu
Keywords: Data management, curriculum, frameworks

ment were invited to speak to the education committee to discuss various issues, identified needs, and analyze future trends. The first phase of curriculum development involved an extensive literature search and inventory of existing data management curricula. It was learned that most data management-related curricula are not openly accessible and are not targeted for students outside of information science programs (library science or computer science). In the second phase, project consultants were used to conduct 50 interviews with students (30 freshmen at WPI, 10 UMMS, and 10 WPI graduate level) about their current data management practices. It was learned that students maintain data in a variety of formats: Excel, Sims 3, Word, PowerPoint, Adobe Illustrator, Photoshop, GraphPad Prism, Filemaker Pro, SPSS, SAS, and NVivo, to name a few. Often data is stored in e-mails, in the cloud (Dropbox, Google Docs), local drives, network drives, or on external drives attached to their workstation. Students typically did not use any standard naming conventions for directories and files unless instructed to do so by faculty or a supervisor.

The following learning objectives for the curriculum have been developed based on the information gathered in these early phases:

By participating fully in this curriculum, the student should be able to:

1. Explain the need for managing/sharing research data, relevant public policies, and the lifecycle continuum for managing and preserving research data;

2. Identify potential re-users, the value of your research data for re-use, and a dissemination strategy;

3. Use an abbreviated data management plan or data curation profile to manage your research project data and define roles/responsibilities of research staff;

4. Explain the range of research data types, stages, formats, and relevant software that may need to be managed and preserved in your future research efforts;

5. Identify what descriptive data needs to be documented in a standard way via metadata to allow your research data sets to be managed and preserved;

6. Plan how to handle issues involved in securely storing research data in central databases, archives and/or repositories, backing it up, and managing access to your data;

7. Explain legal (ownership) and ethical considerations related to data-sharing;

8. Plan for issues related to long-term preservation, discovery, and re-use.

These learning objectives were then translated into seven modules (subsets of subject-related teaching materials containing objectives, directions for use, and test items; see Table 1) linked to the National Science Foundation data management plan requirements and associated competencies (http://www.nsf.gov/eng/general/dmp.jsp).

The curriculum is designed to be flexible enough to be used with students at various educational (undergraduate and master's/PhD) and experience levels. For example, an undergraduate working on their first research project may be required to review all eight modules, whereas a graduate student working in a new lab may be required to review modules 0-3. The curriculum is also designed to be delivered in a variety of methods: video, online self-paced, and one-on-one instruction. Together, these options allow faculty to choose various combinations of modules and delivery methods to be incorporated into a range of learning environments.

JESLIB 2012; 1(1): 46-50
doi:10.7191/jeslib.2012.1003

Table 1: Seven Modules of Learning Objectives

Module # 1- Overview of Research Data Management
Explain what research data is Explain the need for managing/sharing research data and identify relevant public policies Explain the lifecycle continuum to manage and preserve research data Understand that data should be managed differently in different phases of the life cycle Be familiar with data management plan (DMP) requirements used to characterize and plan for the lifecycle of research data. Identify the value and relative importance of data management to the success of a research project.
Module #2 - Types, Formats, and Stages of Data
Explain what a research data set is and the range of data types Identify stages of research data Identify common potential storage formats for data that will be accessible in the future and non-proprietary where possible (i.e., not related to proprietary or custom software/ instruments used for capturing/analyzing data) Identify relevant quality control techniques/technical standards Identify methods of recording data that are specific to student's discipline and research interests. Define data collection recording policies/procedures for student's research.
Module #3 - Contextual Details Needed to Make Data Meaningful to Others
Understand what metadata is Understand why metadata is important Identify applicable standards for documenting and capturing metadata Understand disciplinary practices associated with the collection and sharing of metadata Identify an approach to creating metadata for a project
Module #4 - Data Storage, Backup and Security
Understand why data storage, backup, and security of research data are important Understand data storage, backup, and security methods for research data Understand best practices for research data storage, access control, migration to newer storage media, and security of research data Identify an approach to creating a data storage, backup, and security plan for a project

Table 1 Continued: Seven Modules of Learning Objectives

Module #5 - Legal and Ethical Considerations for Research Data
Explain ownership considerations related to data sharing Explain and evaluate potential legal issues connected to your data; intellectual property, copyright claims, licenses needed for use, monetary charges for data Explain ethical considerations related to data sharing Understand privacy levels for research data as required by potential funding agencies Recognize the importance of privacy with some forms of research data (HIPAA) Understand the importance of removing key personal identifiers to facilitate confidentiality Understand the need for data attribution and citation
Module #6 - Data Sharing & Re-Use Policies
Identify issues related to discovery and re-use and be able to establish relevant policies Identify issues/obstacles related to re-use and sharing Understand publisher's and licensing restrictions on re-use of data and analysis software and instrumentation Understand Open Access requirements Understand controversies surrounding open science, open data Address re-use/sharing requirements from granting agencies or sponsors Address the need for conversion to standard formats needed for re-use Understand different types of collaborative workspaces for sharing data Identify who can share/access your data and for what purpose Determine requirements for pre/post publication access for project phases of the research Determine temporary or permanent access policy Define process steps and access levels for gaining access Understand options for maximizing data reuse
Module #7 - Plan for Archiving and Preservation of Data
Explain options for a long-term sustainable preservation strategy/policy for your data (eg, discipline specific, institutional, departmental). Identify types of repositories/archives (discipline-based, institutional, etc.) Choose appropriate subject repository for long term storage of data Understand process issues for depositing data in repository Identify issues related to discovery of relevant data sets for re-use Understanding the need for querying and retrieval methods - discovery aids for multiple user communities to find the data they want to re-use Explain data management tools and services available Understand costs for data storage, management tools and services

Case Studies

In discussions with faculty there was a strong preference for the curriculum to incorporate case studies based on real data management problems. After interviewing faculty members at each institution, Education Committee members developed five case studies and discussion questions. The cases are constructed so that they can be excerpted into short scenarios and applied to a single module or can be used as a more comprehensive concluding assignment that incorporates all the modules.

Readings and Exercises

A collection of readings targeting different levels were selected and matched to individual modules. In addition, a general bibliography of readings and resources was developed for faculty who wish to assign additional readings, or for graduate level students who may be assigned additional readings. Besides readings, exercises are being developed for each module; these include quizzes, short tasks, and group work.

Next Steps

The need for research data management curricula was confirmed by students, literature review results, and the external experts consulted by the education committee. The curriculum, case studies, readings, and exercises have all been outlined according to the scope of the initial grants. As a next step, UMMS and WPI are developing the lecture content and packaging one module (Module 5) for online delivery as a pilot to be used as additional funds and partners are identified.

References

Carlson, Scott. 2006. “Lost in a sea of science data.” *The Chronicle of Higher Education* 52: A35.

Doctorow, Cory. 2008. “Big data: Welcome to the petacentre.” *Nature* 455: 17.

National Science Board. (2005). “Long lived data collections: Enabling research and education in the 21st century.” Accessed March 2010. http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf.

Acknowledgement

The authors gratefully acknowledge those who have dedicated their time and effort to this project: Steering Committee Members: Elaine Martin, D.A., MSLS, Co-Chair; Tracey Leger-Hornby, PhD, Co-chair; Sia Najafi, MS; Mary Piorun, MSLS, MBA; Donna Kafel, MLIS; Education Committee Members: Patricia Franklin, MD, MBA, MPH; Christine Drew, MLS; Glenn Gaudette, Ph.D.; David Lapointe, Ph.D.; Laura Hanlan, MLIS; Myrna Morales, MAT, MSLIS; Erica Stults, MS, Ph.D. candidate; Lisa Palmer, MSLS; John Sullivan, D.E.; Project Consultants: Curriculum Design: Paul Colombo, MS; Evaluation Expert: Nancy LaPelle, Ph.D.; Instructional Design: Heather McMorrow, MA.

Funding Statement

This project is made possible by a grant from the U.S. Institute of Museum and Library Services and with funds from the National Library of Medicine under Contract No. N01-LM-6-3508.

Disclosure: The authors report no conflicts of interest.