Growing the Next Generation Data Management Experts: Survey of Data Management Training Modules

Chung-Yi Hou (hou@Illinois.edu)[1], Matthew Mayernik (mayernik@ucar.edu)[2]
1 - Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign; 2 – NCAR Library, National Center for Atmospheric Research (NCAR)

Introduction

Using the Federation of Earth Science Information Partners' (ESIP) "Data Management Short Course for Scientists" (Short Course) as the baseline, the Short Course was compared with nine different training modules created by various types of organizations in order to understand the current needs for data management training (DMT) and to build effective training resources.

The poster summarizes and presents the following:

- Common topics presented in the training modules.
- Summary of the effective content presentation styles for the training modules.
- Topics that could be considered when designing the ESIP's next generation training resource.

Background

From 2011-2103, ESIP partnered with the National Oceanic and Atmospheric Administration and the Data Conservancy to produce the current version of the ESIP Short Course. Particularly, the members of the ESIP Data Stewardship Committee created 35 unique modules addressing a variety of basic but broad range of topics pertaining to data management guidelines and best practices.

The current version of the Short Course can be accessed free of charge through two methods:

1) ESIP Commons: http://commons.esipfed.org/datamanagementshortcourse
2) ESIPFED Vimeo: http://vimeo.com/album/2142831

Research Objectives

The initial survey focuses on the DMT resources that are made publicly available and are developed for the Earth and geo science domains. The survey is designed to assess the resources and to meet the following goals:

1) Understand the structure, presentation style, and content of the training topics from the selected DMT resources.
2) Summarize the gaps and the overlaps in the training topics between the different DMT resources and the Short Course.
3) Provide recommendations regarding how to improve and enhance the effectiveness of the DMT resources.

DMT Resources Compared

1. DataONE Education Modules: https://www.dataone.org/education-modules
2. U.S. Geological Survey Data Management Training Modules: http://www.usgs.gov/datamanagement/training/modules.php
3. University of Minnesota Data Management Course (as part of the Data Information Literacy (DIL)): https://sites.google.com/a/umn.edu/data-management-course_structures/home-1
4. MANTRA - Research Data Management Training from EDINA: http://datalib.edina.ac.uk/mantra/
5. University of Illinois at Urbana-Champaign - Research Data Service Workshop: http://researchdataservice.illinois.edu/past-workshops/
6. University of Washington Libraries Research Data Management: https://canvas.uw.edu/courses/889213
7. New England Collaborative Data Management Curriculum: http://library.umassmed.edu/necdmc/index
8. Training for Data Management (TraD) at University of East London: http://www.uel.ac.uk/trad/outputs/resources/
9. Essential 4 Data Support - Research Data Netherlands: http://datasupport.researchdata.nl/en/start-de-cursus/

Method

Select nine unique DMT resources based on the recommendations provided by the ESIP Data Stewardship Committee members → Manually review the content of each resource and document the comparison of the resources in matrix format → Analyze the DMT resources based on the information summarized in the comparison matrix

Results: Overlap between ESIP Short Course and Other DMT Resources (n = 9)*

Percent (n)	Reasons for Data Management
77.8% (7)	• **Institutional Mandates and Data Policies**
77.8% (7)	• **Legal, Ethical, Scientific, and Common Sense Obligations**
77.8% (7)	• **Personal/Professional Achievements**
66.7% (6)	• **Data Management Plan Requirements**

What to Preserve for Data Management

- 77.8% (7) **Definition Discussion of "Scientific Record" and the Different types of Scientific Output**
- 11.1% (1) **Data Management Examples of Different Types of Scientific Output**

Percent (n)	Training for Data Management Plan (DMP)
100.0% (9)	• **Explanation of the Components within a DMP**
88.9% (8)	• **Objects to Include in a DMP**
88.9% (8) to 55.6% (5)	• **Use of Standard/Consistent Format for DMP Content Structure and Organization** • File Naming Conventions (8) • Data Formats (7) • Organization and Standards for Data Management Plan (5)

Recognition and Acknowledging Data Management

- 88.9% (8) **Citation**

Percent (n)	Training for Documentation
88.9% (8)	• **Metadata and Data Documentation**

Ways to Work with Data Repositories/Archives

- 88.9% (8) **Preservation Planning**
- 88.9% (8) to 77.8% (7) **Provide Access to Managed Data**
  - Data Access Topic Discussions (8)
  - Data Access Mechanisms (7)
- 77.8% (7) **Clarification of Policies, Agreements, and Roles/Responsibilities**
- 66.7% (6) to 33.3% (3) **User Outreach**
  - Data Advertisement(6)
  - Determination of Data Audience (5)
  - Providing Access to Your Data: Tracking Data Usage (4)
  - Broadening Data User Community (3)
  - Data Portals and Metadata Registries (3)

Percent (n)	Other Key Issues
100.0% (9) to 77.8% (7)	• **Legal, Privacy, and Confidentiality** • Handling Sensitive Data (9) • Data Restrictions (9) • Data Access and Rights (8) • Copyright and Data (7)

***Please see the additional handout for the specific ESIP Short Course titles that are related to each of the topics.**

Note:
The digital version of the handout and the full comparison matrix that was used for the result analysis are available from the following link: http://hdl.handle.net/2142/88826

Lessons Learned / Recommendations

The following key areas have been observed to be helpful in delivering the content of the DMT resources to its users:

- **Explain** clearly the objectives of the modules and who the intended users are for the modules.
  - Different user types might need different narratives to help in making the content relevant and relatable.
- **Organize** the modules with an overall, encompassing theme or structure could help the users in orienting themselves between modules and to visualize the relationships between the module topics.
- **Supplement** the modules with information that can be just as instructive and effective as the modules' main content.
  - **Examples** of various forms of additional resources include: annotated bibliography, relevant software/tools, case studies, discussion questions, exercises, and glossary of terms.
- **Test the technology or platform** used to build and publish the modules.
  - **Elements** that could affect the effectiveness of the modules include: the operation/functions available to interact with the modules, the audio volume, the display design (font size, color, quantity of text per page), and availability for personalization (adding personal notes, tags, resources, etc).
- **Expand** on the current frequently discussed topics to broaden the scope and coverage of data management.
  - Possible discussion areas include: policy making, training/continued education, and financial support and sustainability.

Future Work

- Add to the comparison a selection of syllabi from academic curriculums or training courses that provide degrees or certifiable credentials with data management as focus.
- Provide a summary of a set of topics that could be considered as the "core knowledge" of data management based on the consistency and frequency of the topics' coverage and discussion among the DMT resources.
- Identify potential partners and collaborators who can work together to provide funding and create content for the next generation of DMT resources.
- Assist in evaluating the effectiveness of the ESIP's Short Course.

Acknowledgment

The authors would like to thank the Foundation for Earth Science, the Federation of Earth Science Information Partners (ESIP), and the ESIP Data Stewardship Committee for their support for this project.