



Data Management Training Resources Survey Report

Sophie Hou (hou@illinois.edu) - ESIP Data Stewardship Committee Student Fellow
Matthew Mayernik (mayernik@ucar.edu) – National Center for Atmospheric Research
ESIP Winter Meeting
January 2016



Agenda

- Introduction: Data Management Training (DMT) Resources Survey Project
- Background: Data Management for Scientists Short Course
- Methodology: Selection of DMT Resources and the Comparison Matrix
- Results: Overlaps, Gaps, and Recommendations
- Future Work: Short Course Roadmap, Potential Collaborators, Clearinghouse



Introduction:

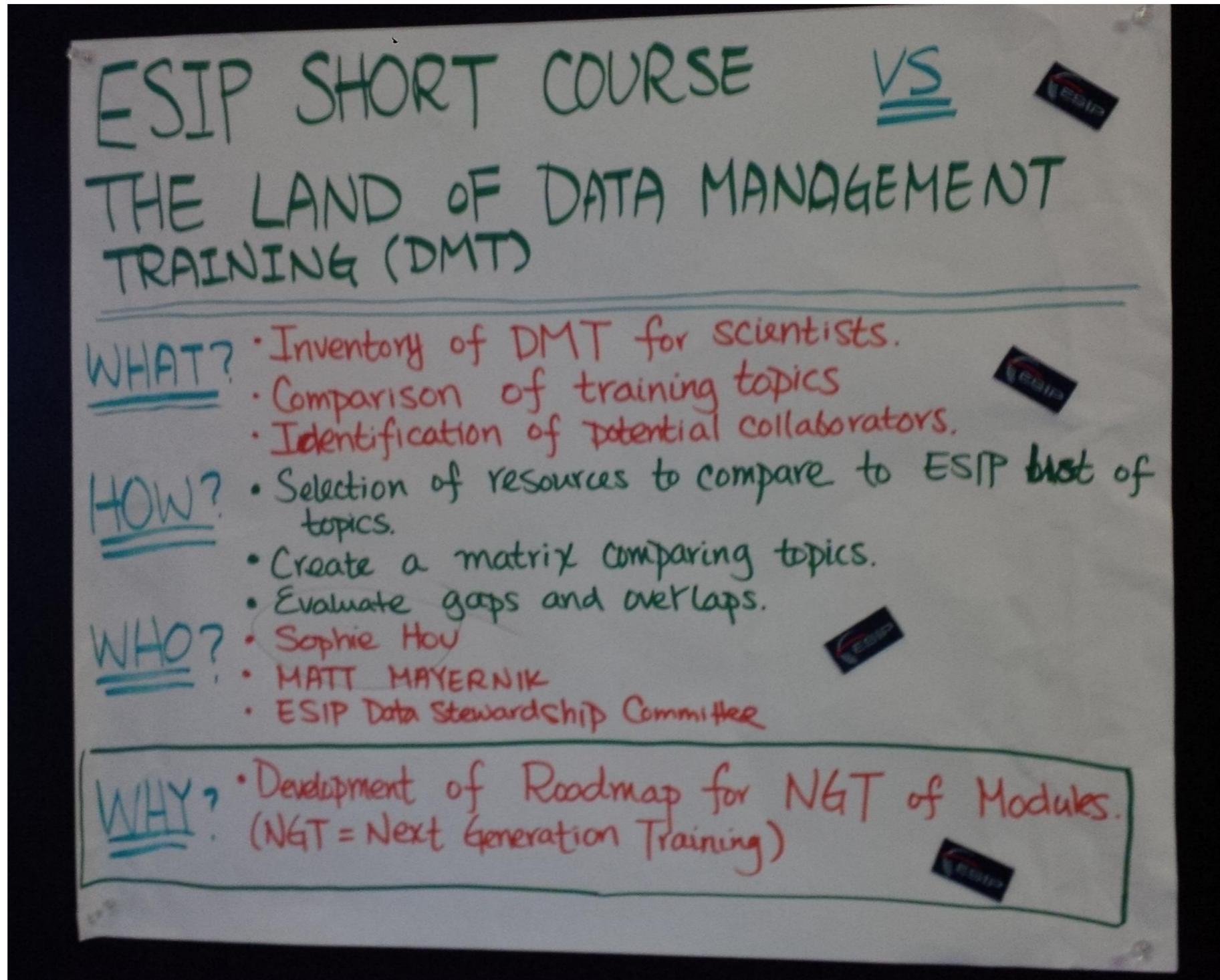
Data Management Training (DMT) Resources Survey Project



ESIP Summer Meeting, 2015

- The session - “Planning for Next Season's Data Management Short Course Episodes!” - was held during the 2015 Summer Meeting in order to gather people who would be interested in helping and to discuss potential ideas for creating the second generation of the Data Management for Scientists Short Course.
- Many who attended the session noted that several DMT resources were also being developed by other organizations and delivered in different formats
- The group determined that it would be helpful if the current “landscape” of the DMT resources was assessed before starting ESIP’s own effort.
- The “Data Management Training (DMT) Resources Survey” project was proposed and accepted at 2015ESIP Summer Meeting’s FUNding Friday.

Original FUNding Friday Poster

A photograph of a whiteboard with handwritten text in green and red ink. The text is organized into sections: a title, a comparison point, and four questions (WHAT?, HOW?, WHO?, WHY?) each followed by a list of bullet points. There are several small ESIP logos pinned to the board.

ESIP SHORT COURSE VS

THE LAND OF DATA MANAGEMENT TRAINING (DMT)

WHAT?

- Inventory of DMT for scientists.
- Comparison of training topics
- Identification of potential collaborators.

HOW?

- Selection of resources to compare to ESIP list of topics.
- Create a matrix comparing topics.
- Evaluate gaps and overlaps.

WHO?

- Sophie Hou
- MATT MAYERNIK
- ESIP Data Stewardship Committee

WHY?

- Development of Roadmap for NGT of Modules. (NGT = Next Generation Training)



Background:

Data Management for Scientists Short Course

Data Management for Scientists Short Course



- Developed between 2011 and 2013 by ESIP, in cooperation with NOAA and the Data Conservancy.
 - 12 individuals contributed as module authors, and the authors represented 12 different organizations covering federal agencies, academic institutions, information organizations, and data centers.
- Currently, there is a total of 35 modules available.
- The modules can be accessed free of charge through:
 - 1) ESIP Commons (<http://commons.esipfed.org/datamanagementshortcourse>)
 - 2) ESIPFED Vimeo (<http://vimeo.com/album/2142831>)
- The Short Course has been presented to these audiences at events with data management focus in order to collect and review their feedback.

Data Management for Scientists

Short Course - Continued



Titles for: Responsible Data Use Section

- [Citation and Credit](#)
- [Copyright and Data](#)
- [Data Restrictions](#)

Titles for: Data Management Plans Section

- [Why Create a Data Management Plan?](#)
- [Elements of a Data Management Plan](#)
 - [Identifying the materials to be created](#)
 - [Organization and Standards](#)

Titles for: The Case for Data Stewardship Section

- [Agency Requirements](#)
 - [NASA Data Management Plans](#)
 - [NSF Data Management Plans](#)
 - [NOAA Administrative Order 212-15: Management of Environmental and Geospatial Data and Information](#)
- [Enhancing Your Reputation](#)
- [Preserving the Scientific Record](#)
 - [Establishing Relationships with Archives](#)
 - [Preserving a Record of Environmental Change](#)
 - [Case Study 1 – National Snow & Ice Data Center \(NSIDC\) Glacier Photos](#)
 - [Case Study 2 – Arctic Temperature Variability Data](#)

Titles for: Local Data Management Section

- [Managing Your Data](#)
 - [Assign Descriptive File Names](#)
 - [Backing Up Your Data](#)
- [Data Formats](#)
 - [Choosing and Adopting Community Accepted Standards](#)
 - [Using Self-describing Data Formats](#)
- [Creating Documentation and Metadata](#)
 - [Introduction to Metadata and Metadata Standards](#)
 - [Creating a Citation for Your Data](#)
 - [Metadata for Discovery](#)
- [Working with Your Archive: Broadening Your User Community](#)
- [Advertising your data](#)
 - [Agency requirements for submitting metadata](#)
 - [Using data portals and metadata registries](#)
- [Using Data Portals and Metadata Registries: Submitting Metadata to the GCMD](#)
- [Providing Access to Your Data](#)
 - [Determining your audience](#)
 - [Access Mechanisms](#)
 - [Tracking Data Usage](#)
 - [Handling sensitive data](#)
 - [Rights](#)

Data Management for Scientists Short Course - Continued



ESIP Commons Learn Meeting - Help Log In

Why Create a Data Management Plan?

Submitted by administrator on Tue, 03/05/2013 - 15:26

Overview:
In this module, we'll very briefly review what a Data Management Plan is, followed by a discussion of our top three reasons for creating a data management plan. These reasons are:

First, that proper data management planning should make your work easier and possibly even cheaper than if you had handled your data in an ad-hoc fashion throughout your project.

Second, handling your data properly and documenting them well, can actually improve your standing with your users and with your colleagues, most importantly, and just and perhaps least, because your funding agency says that you must.

Hopefully, by the end of this module you will become convinced that while funding agency requirements may be the stick making you create data management plans now, creation of the plans has actually been in your best interest all along.

Why Create a Data Management Plan?
Ruth Duerr
National Snow and Ice Data Center
Version 1.0
February 2013

Attachments:
DMPWhyDoADataManagementPlanDuerr_final.pptx
Media/Video:
<http://vimeo.com/69442862>

Collaboration Area:
Data Management Training
Data Preservation

DOI #E2id:
doi:10.7202/103718425

Target Audience:
Research Scientist

Educational Purpose:
Online course

Time Required:
15m00s

Learning Mode:
Asynchronous

Learning Resource Type:
Online exercise

Language of Learning Resource:
English

Publication Date:
February 2013

Publisher:
Federation of Earth Science Information Partners

Contributor:
Video performed by Nancy J. Hoebebeinrich
Editor:
Ruth E. Duerr
Nancy J. Hoebebeinrich

Version:
1.0

Keywords:
data management
online courses
data literacy
data management plans (DMP)
training

Log in or register to post comments

Twitter +1 LinkedIn Facebook Print PDF

vimeo Join Log In Create Watch Search Upload

The Case for Data Management: Agency Requirements

Ruth Duerr
National Snow and Ice Data Center

Version 1.0
Review Date

NOAA **DataConservancy**

05:07

The Case for Data Management: Agency Requirements

from **ESIPFed** PLUS 2 years ago / NOT YET RATED

This is the Federation of Earth Science Information Partners Data Management Short Course, Section: The Case for Data Management, Module: Agency Requirements

This training module is part of the Federation of Earth Science Information Partners (or ESIP Federation's) Data Management Course for Scientists. The subject of this module is "The Case for Data Management: Agency Requirements." This module was authored by Ruth Duerr from the National Snow and Ice Data Center. Besides the ESIP Federation, sponsors of this Data Management Course are the Data Conservancy and the United States National Oceanographic and Atmospheric Administration (NOAA).

Follow + Collections Stats Download

vimeo.com/45318853



Methodology:

Selection of Data Management Training (DMT)
Resources and the Comparison Matrix



Selection of Resources

- Potential DMT resources were collected based on the recommendations provided by the ESIP Data Stewardship Committee members.
- All of the DMT resources selected were formatted and presented as publicly available training modules. Additionally, the DMT resources selected were developed mainly for science disciplines, especially with Earth and geo science researchers as part of the key, targeted users.
- A total of nine unique DMT resources were selected in order to represent a diversity of the following characteristics:
 - Federal Science Agency vs. Academic Libraries vs. Information Professionals Consortia.
 - National vs. International.



Selection of Resources - Continued

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/>



Results:

Overlaps, Gaps, and Recommendations



Overlaps with the ESIP Short Course

- The poster – “Data Management Training Modules: An Initial Survey and Comparison Result (Funding Friday)” (<http://commons.esipfed.org/node/8763>) - will focus on presenting topics that overlapped with the ESIP Short Course.
- Overall, most of the Short Course’s current topics were also discussed by the selected DMT resources with the following exceptions:
 - Data Management Examples of Different Types of Scientific Output (1 or 11.1%)
 - Discussion of self-describing data formats (1 or 11.1%).
 - Providing Access to Your Data: Tracking Data Usage (4 or 44.4%)
 - Broadening Data User Community (3 or 33.3%)
 - Data Portals and Metadata Registries (3 or 33.3%)



Gaps in the ESIP Short Course

- Out of the 68 planned, future topics by the Short Course, 35 were already covered by the nine selected DTM resources.
 - **Topics that have been covered most frequently are:**
 - “Some available resources to help with developing your plan” (5)
 - “Working with your archive - Planning for longer term preservation” (6)
 - “Advertising Your Data - Journals and Publications” (5)
 - “What archives are out there? (5)
 - Discipline or institutional archives
 - Finding an archive
 - What to do if there is no archive out there”
 - **Topics that have been covered least frequently are (each topic is covered by 1 other resource):**
 - “Data Formats - Building understandable spreadsheets”
 - “Creating documentation and metadata – For your collections as a whole / Creating item level metadata / metadata for tracking data processing”
 - “Writing Sharable Code”
 - “What do long term archives do with my data?”
 - “Other aspects necessary for understanding how to support the data”
 - “Media migration”

Gaps in the ESIP Short Course - Continued



- A total of 104 additional topics that were not in the Short Course were covered by the nine selected DTM resources.
 - **Topics that have been covered most frequently are:**
 - Value of data sharing from a variety of perspectives, including the public, researchers, funders/sponsors, and scientific community (9)
 - Concerns regarding data sharing that are not related to privacy/confidentiality/rights (8)
 - Things to consider and best practices relating to data preservation (7)
 - Value of long term storage and management (6)
 - Legislation, policies, and legal considerations (6)
 - **Sample of the topics that have been covered least frequently are (each topic is covered by 1 other resource):**
 - Tools to help with data entry
 - Spreadsheet versus relational database
 - Data manipulation / data analysis / workflow
 - Data Quality and assurance / data contamination / Data Seal of Approval
 - Differences between various preservation strategies, such as backup versus archive versus file transformation
 - Metadata comparison/crosswalk
 - Quality of metadata
 - Cost model
 - Librarian as data management information partners



Gaps in Training Topics

- Out of the 68 planned, future topics by the Short Course, the following is a sample of the topics that were not covered by any of the nine resources:
 - **The Case for Data Stewardship:**
 - “Return on Investment”
 - “Facilitating science through interoperable discovery and access”
 - “What Not to do when Archiving Data!”
 - **Data Management Plans:**
 - “Elements of a plan - roles and responsibilities”
 - “Elements of a plan - Data Access, sharing, and re-use policies”
 - **Local Data Management:**
 - Working with your archive - Work with your archive early and often
 - Casting your data
 - **Preservation Strategies:**
 - Emerging standards for preservation
 - **Responsible Data Use:**
 - Providing Feedback
 - Collaboration
 - Community participation



Observations / Recommendations

- **Explain** clearly the objectives of the modules and who the intended users are for the modules.
- **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 module's main content.
- **Test** the technology or platform used to build and publish the modules.
- **Expand** on the current frequently discussed topics to broaden the scope and coverage of data management.



Future Work:

Short Course Roadmap, Potential
Collaborators, Clearinghouse



Next Steps - Discussion

- 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 Short Course.



Thank You!

Sophie Hou (hou@illinois.edu)
Matthew Mayernik (mayernik@ucar.edu)