Community Resources for Data Management

Megan Mach and Amber Budden
DataONE

Nancy Hoebelheinrich
Knowledge Motifs
DataONE Cyberinfrastructure
Data Holdings

dataone.org/current-member-nodes

**Uploads**

The number of individual metadata and data files uploaded over time. Only the first version of each file is counted.
DataONE Search

search.dataone.org
Typical data analyses:
- **Data processing**: may include selecting a subset of data for analysis, merging and aligning data from different sources, and cleaning data for analysis or data transformation.
- **Graphical analysis**: makes it easier to see patterns and can aid in the identification of outliers.
- **Statistical analysis**: conventional statistics are used to analyze experimental data; descriptive statistics are used to analyze observational or descriptive data.

Science is iterative: the process that results in the final product can be complex.

**Reproducibility**: is at the core of the scientific process. If results are not reproducible, they lose credibility. Good documentation of the data and the analysis are essential!

**Formal Workflow**
- Analytical pipeline where each step can be implemented with different software systems.
- Parameters and requirements for each step are formally recorded.
- Single access point for multiple analyses across software packages.
- Keeps track of analysis and provenance to better enable reproducibility.
- Workflow can be stored.
- Allows sharing and reuse of individual steps or overall workflow.

**Local contact information**

**Lesson 10: Analysis and Workflows**

**Definition**: Formal description of the procedures used to create outputs. Can be formal or informal.

**Informal workflow**
- No special software is needed to create workflow diagrams. Workflow diagrams include:
  - Inputs and outputs
  - Transformation rules or analytical processes
  - Decision points
  - Arrows indicating direction of process flow

**Informal Workflow Example**

**Best practices for data analysis**
- Formally or informally document the workflows used to create outputs, include:
  - Analyses and parameters used
  - Connections between analyses and steps in the workflow.
- Document the code you wrote for the analyses.
  - Save documentation code to version control and share and enable repeated analyses.
  - Include project level information, complete with data provenance, and what happens in individual analyses.

**Construct end-to-end scripts that exercise entire processes from start to finish without intervention.**
The Fractured Lab Notebook

Strasser & Hampton (2012) Ecosphere 3:12 DOI: 10.1890/ES12-00139.1
DataONE Education Modules

dataone.org/education-modules
Hands-on Activity 3: Data Management Planning

**Objectives:**
- Understand the challenges of creating a data management plan (DMP).
- Develop a simple DMP using key components.

**Instructions:**
- Provide a template:
  - Title
  - Introduction
  - Data sources
  - Data quality
  - Data format
  - Data storage
  - Data access
  - Data sharing
  - Data security
  - Data preservation
- Develop a simple DMP using the template.

**Outcomes:**
- Identify the basic components of a data management plan (DMP) and draft a simple DMP using key components.

**Additional Resources:**
- DMPTool: http://www.dmp-tool.org/
- Example DMPs from DataONE: http://www.dataone.org/data-management-planning

**Example DMPs from DataONE:**
- NSF-Generic: A good example of a basic template.

**Data Management Planning**

**Time Needed:**

**Key Readings:**
- Data Management Planning: http://dmptool.org/dm_guidance
- Good practices for data management planning.

**Additional Files Needed**
- The DMPTool website, especially the resources.
- Example DMPs from DataONE.

**Planning**

**Hands-on Activity 3: Data Management**

- Allow students plenty of time to explore the DMPTool website, especially the resources.
- Example DMPs from DataONE: http://www.dataone.org/data-management-planning

**Outcomes:**
- Identify the basic components of a data management plan (DMP) and draft a simple DMP using key components.

**Additional Resources:**
- DMPTool: http://www.dmp-tool.org/
- Example DMPs from DataONE: http://www.dataone.org/data-management-planning

**Example DMPs from DataONE:**
- NSF-Generic: A good example of a basic template.

**Data Management Planning**

**Time Needed:**

**Key Readings:**
- Data Management Planning: http://dmptool.org/dm_guidance
- Good practices for data management planning.

**Additional Files Needed**
- The DMPTool website, especially the resources.
- Example DMPs from DataONE.

**Planning**

**Hands-on Activity 3: Data Management**

- Allow students plenty of time to explore the DMPTool website, especially the resources.
- Example DMPs from DataONE: http://www.dataone.org/data-management-planning

**Outcomes:**
- Identify the basic components of a data management plan (DMP) and draft a simple DMP using key components.

**Additional Resources:**
- DMPTool: http://www.dmp-tool.org/
- Example DMPs from DataONE: http://www.dataone.org/data-management-planning

**Example DMPs from DataONE:**
- NSF-Generic: A good example of a basic template.

**Data Management Planning**

**Time Needed:**

**Key Readings:**
- Data Management Planning: http://dmptool.org/dm_guidance
- Good practices for data management planning.

**Additional Files Needed**
- The DMPTool website, especially the resources.
- Example DMPs from DataONE.

**Planning**

**Hands-on Activity 3: Data Management**

- Allow students plenty of time to explore the DMPTool website, especially the resources.
- Example DMPs from DataONE: http://www.dataone.org/data-management-planning

**Outcomes:**
- Identify the basic components of a data management plan (DMP) and draft a simple DMP using key components.

**Additional Resources:**
- DMPTool: http://www.dmp-tool.org/
- Example DMPs from DataONE: http://www.dataone.org/data-management-planning

**Example DMPs from DataONE:**
- NSF-Generic: A good example of a basic template.

**Data Management Planning**

**Time Needed:**

**Key Readings:**
- Data Management Planning: http://dmptool.org/dm_guidance
- Good practices for data management planning.

**Additional Files Needed**
- The DMPTool website, especially the resources.
- Example DMPs from DataONE.

**Planning**

**Hands-on Activity 3: Data Management**

- Allow students plenty of time to explore the DMPTool website, especially the resources.
- Example DMPs from DataONE: http://www.dataone.org/data-management-planning

**Outcomes:**
- Identify the basic components of a data management plan (DMP) and draft a simple DMP using key components.

**Additional Resources:**
- DMPTool: http://www.dmp-tool.org/
- Example DMPs from DataONE: http://www.dataone.org/data-management-planning

**Example DMPs from DataONE:**
- NSF-Generic: A good example of a basic template.

**Data Management Planning**

**Time Needed:**

**Key Readings:**
- Data Management Planning: http://dmptool.org/dm_guidance
- Good practices for data management planning.

**Additional Files Needed**
- The DMPTool website, especially the resources.
- Example DMPs from DataONE.

**Planning**

**Hands-on Activity 3: Data Management**

- Allow students plenty of time to explore the DMPTool website, especially the resources.
- Example DMPs from DataONE: http://www.dataone.org/data-management-planning

**Outcomes:**
- Identify the basic components of a data management plan (DMP) and draft a simple DMP using key components.

**Additional Resources:**
- DMPTool: http://www.dmp-tool.org/
- Example DMPs from DataONE: http://www.dataone.org/data-management-planning

**Example DMPs from DataONE:**
- NSF-Generic: A good example of a basic template.
## Community Use

<table>
<thead>
<tr>
<th>Page</th>
<th>Total Events</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>232,749</td>
</tr>
<tr>
<td></td>
<td>% of Total: 100.00% (232,749)</td>
</tr>
<tr>
<td>1. /education-modules</td>
<td>30,854</td>
</tr>
<tr>
<td></td>
<td>(13.24%)</td>
</tr>
<tr>
<td>2. /best-practices/create-and-document-data-back-up-policy</td>
<td>23,989</td>
</tr>
<tr>
<td></td>
<td>(10.30%)</td>
</tr>
<tr>
<td>3. /data-management-planning</td>
<td>23,604</td>
</tr>
<tr>
<td></td>
<td>(10.13%)</td>
</tr>
<tr>
<td>4. /best-practices</td>
<td>10,892</td>
</tr>
<tr>
<td></td>
<td>(4.67%)</td>
</tr>
<tr>
<td>5. /software-tools/fusion-lidar-software</td>
<td>8,562</td>
</tr>
<tr>
<td></td>
<td>(3.67%)</td>
</tr>
<tr>
<td>6. /current-member-nodes</td>
<td>8,198</td>
</tr>
<tr>
<td></td>
<td>(3.52%)</td>
</tr>
<tr>
<td>7. /find-data</td>
<td>7,635</td>
</tr>
<tr>
<td></td>
<td>(3.28%)</td>
</tr>
<tr>
<td>8. /</td>
<td>6,504</td>
</tr>
<tr>
<td></td>
<td>(2.79%)</td>
</tr>
<tr>
<td>9. /software-tools/dbdesigner-4</td>
<td>6,409</td>
</tr>
<tr>
<td></td>
<td>(2.75%)</td>
</tr>
<tr>
<td>10. /investigator-toolkit</td>
<td>5,756</td>
</tr>
<tr>
<td></td>
<td>(2.47%)</td>
</tr>
</tbody>
</table>
# Community Use

<table>
<thead>
<tr>
<th>Page</th>
<th>Event Label</th>
<th>Total Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>/education-modules</td>
<td>14,947 (6.42%)</td>
</tr>
<tr>
<td>2.</td>
<td>/best-practices/create-policy</td>
<td>2,115 (14.15%)</td>
</tr>
<tr>
<td>3.</td>
<td>/data-management</td>
<td>1,364 (9.12%)</td>
</tr>
<tr>
<td>4.</td>
<td>/best-practices</td>
<td>1,029 (6.88%)</td>
</tr>
<tr>
<td>5.</td>
<td>/software-tools/full</td>
<td>913 (6.11%)</td>
</tr>
<tr>
<td>6.</td>
<td>/current-member-reports</td>
<td>793 (5.30%)</td>
</tr>
<tr>
<td>7.</td>
<td>/find-data</td>
<td>762 (5.10%)</td>
</tr>
<tr>
<td>8.</td>
<td>/</td>
<td>740 (4.95%)</td>
</tr>
<tr>
<td>9.</td>
<td>/software-tools/developers</td>
<td>725 (4.85%)</td>
</tr>
<tr>
<td>10.</td>
<td>/investigator-toolkit</td>
<td>634 (4.24%)</td>
</tr>
<tr>
<td>11.</td>
<td>/</td>
<td>600 (4.01%)</td>
</tr>
</tbody>
</table>
Using Peer Review to Support Development of Community Resources for Research Data Management

Heather Soyka  
Kent State University

Amber Budden  
DataONE/University of New Mexico

Viv Hutchison  
US Geological Survey
Solutions

GitHub
Versioned work space
Versioned work space
Structured repository

The image shows a screenshot of a repository named `hub_lessons` on a platform named DataONE. The repository contains various subdirectories and files, each with a commit history indicating updates to the repository. The files are structured under the following directories:

- `00_markdown`
- `01_management`
- `02_datasharing`
- `03_planning`
- `04_entry`
- `05_gaqc`
- `06_protect`
- `07_metadata`
- `08_citation`
- `09_analysis`
- `10_policy`

Each file contains the same message: "update org info and title/author in front matter". The latest commit is dated Jul 23, and the repository is 3 months ago.
Structured repository
The Data Management Skillbuilding Hub contains resources for better data management and is open to community input and update. These resources are adaptable across a range of contexts and intended for use by researchers, teachers, librarians, or anyone who wants to learn better data management practices. Each tile below links to community contributed education materials, such as best practices and lesson plans.

The resources presented on the Data Management Skillbuilding Hub can be updated by users to promote a current, well-maintained, and sustainable educational tool. Learn more about how you can contribute.

**Using This Resource**

Click individual tiles to learn more and use each resource. You can limit resources by content type and Data Life Cycle stage. Comprehensive information is available in the FAQ.

- Filter by content type: [ALL] [TEACHING MODULE] [BEST PRACTICE] [VIDEO]
- Filter by stage of the Data Life Cycle: [All]
Data Management Skillbuilding Hub

• Structure: Data life-cycle
• Current holdings: Education modules and best practices
• Citation: Credit where credit is due
• Editing content (forking!)
• Creating new content (in the works)
• Future holdings
The Data Management Skillbuilding Hub contains resources for better data management and is open to community input and update. These resources are adaptable across a range of contexts and intended for use by researchers, teachers, librarians, or anyone who wants to learn better data management practices. Each tile below links to community contributed education materials, such as best practices and lesson plans.

The resources presented on the Data Management Skillbuilding Hub can be updated by users to promote a current, well-maintained, and sustainable educational tool. Learn more about how you can contribute.

Using This Resource

Click individual tiles to learn more and use each resource. You can limit resources by content type and Data Life Cycle stage. Comprehensive information is available in the FAQ.

Filter by content type:  
- ALL  
- TEACHING MODULE  
- BEST PRACTICE  
- VIDEO

Filter by stage of the Data Life Cycle:  
- All
Data life-cycle stages
Data Management Skillbuilding Hub

The Data Management Skillbuilding Hub contains resources for better data management and is open to community input and update. These resources are adaptable across a range of contexts and intended for use by researchers, teachers, librarians, or anyone who wants to learn better data management practices. Each tile below links to community contributed education materials, such as best practices and lesson plans.

The resources presented on the Data Management Skillbuilding Hub can be updated by users to promote a current, well-maintained, and sustainable educational tool. Learn more about how you can contribute.

**Using This Resource**

Click individual tiles to learn more and use each resource. You can limit resources by content type and Data Life Cycle stage. Comprehensive information is available in the FAQ.

- Filter by content type: ALL, Plan, Collect, Assure, Describe, Preserve, Discover, Integrate, Analyze
- Filter by stage of the Data Life Cycle

---

**01. Why Data Management**

**02. Data Sharing**

**07. Metadata**

---

**08. Data Citation**

---

**Hosted by DataONE**

In collaboration with the community, DataONE has developed high quality resources for helping educators and librarians with training in data management, including teaching materials, webinars and a database of best-practices to improve methods for data sharing and management.

If you have a question or concern, please open an issue in this repository on GitHub.
When first sharing research data, researchers often raise questions about the value, benefits, and mechanisms for sharing. Many stakeholders and interested parties, such as funding agencies, communities, other researchers, or members of the public may be interested in research, results and related data. This lesson addresses data sharing in the context of the data life cycle, the value of sharing data, concerns about sharing data, and methods and best practices for sharing data.

Cite this lesson:
The Data Lifecycle

Several stages require critical attention to ensure effective data sharing

<table>
<thead>
<tr>
<th>Step</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Describe</td>
<td>document the data content, character and process</td>
</tr>
<tr>
<td>Deposit</td>
<td>store the data in a location from which it can be accessed</td>
</tr>
<tr>
<td>Preserve</td>
<td>select storage formats and media with long term use in mind</td>
</tr>
<tr>
<td>Discover</td>
<td>publish information about the data so that others can find it</td>
</tr>
</tbody>
</table>

Effective data sharing requires careful thought during each stage of the data development process including:

- description and documentation of the data process, content, and character;
- deposition and storage of the data in a location from which it can be accessed or shared;
- preservation of the data using a format and media that enable long term reuse; and
- making the data discoverable by publishing information about the data in research publications, data clearinghouses and data distribution portals.

Why share data

Data sharing requires effort, resources, and faith in others. Why do it?

For the benefit of:
- the public
- the research sponsor
- the research community
- the researcher

Why expend the extra effort to share data? Because it benefits the public, the research sponsor, the research community and, perhaps most importantly, the researcher.
When first sharing research data, researchers often raise questions about the value, benefits, and mechanisms for sharing. Many stakeholders and interested parties, such as funding agencies, communities, other researchers, or members of the public may be interested in research, results and related data. This lesson addresses data sharing in the context of the data life cycle, the value of sharing data, concerns about sharing data, and methods and best practices for sharing data.

Cite this lesson:
The Data Management Skillbuilding Hub contains resources for better data management and is open to community input and update. These resources are adaptable across a range of contexts and intended for use by researchers, teachers, librarians, or anyone who wants to learn better data management practices. Each tile below links to community contributed education materials, such as best practices and lesson plans.

The resources presented on the Data Management Skillbuilding Hub can be updated by users to promote a current, well-maintained, and sustainable educational tool. Learn more about how you can contribute.

Using This Resource
Click individual tiles to learn more and use each resource. You can limit resources by content type and Data Life Cycle stage. Comprehensive information is available in the FAQ.

» Filter by content type:  ALL  TEACHING MODULE  BEST PRACTICE  VIDEO

» Filter by stage of the Data Life Cycle  All
Best Practice: Assure

Select a Best Practice below to learn more about the “Assure” stage in the Data Life Cycle.

What is the “Assure” stage?

Employ quality assurance and quality control procedures that enhance the quality of data (e.g., training participants, routine instrument calibration) and identify potential errors and techniques to address them.

More information can be found in the Best Practices Primer.
Best Practice: Assure

Select a Best Practice below to learn more about the “Assure” stage in the Data Life Cycle.

What is the “Assure” stage?

Employ quality assurance and quality control procedures that enhance the quality of data (e.g., training participants, routine instrument calibration) and identify potential errors and techniques to address them.

More information can be found in the Best Practices Primer.

Communicate data quality

Information about quality control and quality assurance are important components of the metadata: (click for more)

Tags: assure flag qualify

Confirm a match between data and their description in metadata

To assure that metadata correctly describes what is actually in a data file, visual inspection or analysis should be done by someone not otherwise familiar with the data and its format. This will assure that the metadata is sufficient to describe the data... (click for more)

Tags: assure data consistency describe documentation metadata quality

Consider the compatibility of the data you are integrating

The integration of multiple data sets from different sources requires that they be compatible. Methods used to create the data should be considered early in the process, to avoid problems later during attempts to integrate data sets. Note that just beca... (click for more)

Tags: analyze assure database integrate quality tabular

Develop a quality assurance and quality control plan

Just as data checking and review are important components of data management, so is the step of documenting how these tasks were accomplished. Creating a plan for how to review the data before it is collected or compiled allows a researcher to
Communicate data quality

Data Life Cycle stage(s): Assure

Information about quality control and quality assurance are important components of the metadata:

- Qualify (flag) data that have been identified as questionable by including a flagging column next to the column of data values. The two columns should be properly associated through a naming convention such as Temperature, flag_Temperature.
- Describe the quality control methods applied and their assumptions in the metadata. Describe any software used when performing the quality analysis, including code where practical. Include in the metadata who did the quality control analysis, when it was done, and what changes were made to the dataset.
- Describe standards or test data used for the quality analysis. For instance, include, when practical, the data used to make a calibration curve.
- If data with qualifier flags are summarized to create a derived data set, include the percent flagged data and percent missing data in the metadata of the derived data file. High frequency observations are often downsampled, and it is critical to know how much of the data were rejected in the primary data.

Description Rationale

Data quality and any methods used for quality control should be communicated so others can assess the data independently.

Additional Information


Additional Information (Biblio)

Best Practices for Preparing Ecological and Ground-Based Data Sets to Share and Archive
Search the Data Management Skill Building Hub

Results are listed in order of find, not by best match to search word, and will be alphabetically ordered by result type.

data backup

**Teaching Module:** Protecting Your Data: Backups, Archives & Data Preservation
Life Cycle Step(s): preserve, assure
Authoring Organization: DataONE

**Best Practice:** Create and document a data backup policy
Life Cycle Step(s): plan, preserve
Authoring Organization: DataONE

**Best Practice:** Ensure integrity and accessibility when making backups of data
Life Cycle Step(s): preserve
Authoring Organization: DataONE

*Hosted by DataONE*

In collaboration with the community, DataONE has developed high quality resources for helping educators and librarians with training in data management, including teaching materials, webinars and a database of best-practices to improve methods for data sharing and management.

If you have a question or concern, please open an Issue in this repository on GitHub.
When first sharing research data, researchers often raise questions about the value, benefits, and mechanisms for sharing. Many stakeholders and interested parties, such as funding agencies, communities, other researchers, or members of the public may be interested in research, results and related data. This lesson addresses data sharing in the context of the data life cycle, the value of sharing data, concerns about sharing data, and methods and best practices for sharing data.

Cite this lesson:
Data Management Skillbuilding Hub

- Structure: Data life-cycle stage
- Current holdings: Education modules and best practices
- Citation: Credit where credit is due
- **Editing content (forking!)**
- Creating new content (in the works)
- Future holdings
Guidelines for contributors and content editors

This document details our recommended processes to update current content, suggest changes to content, and fork content for your own use, as well as an introduction to how the content is organized and the tools we use to display content.

This repository was developed by the DataONE Community Engagement and Outreach Working Group and continues to be maintained by members of this team. Thank you for your interest in contributing to these educational materials.
Guidelines for contributors and content editors

This document details our recommended processes to update current content, suggest changes to content, and fork content for your own use, as well as an introduction to how the content is organized and the tools we use to display content.

This repository was developed by the DataONE Community Engagement and Outreach Working Group and continues to be maintained by members of this team. Thank you for your interest in contributing to these educational materials.

Update current content

Want to update a link or method? See a spelling error? Changes can be easily proposed by opening the GitHub Education page and editing content directly. For help, try this brief GitHub tutorial on forking and editing content.

Edit content

1. Create a fork of the lessons or best practices repository into your github account, depending on which content you wish to edit.
2. Modify the files that you want to change (See “Structure” below for tips on making changes).
3. Submit a pull-request against the master branch of this repository.
4. Your changes will be reviewed by the repository admins.
Guidelines for contributors and content editors

This document details our recommended processes to update current content, suggest changes to content, and fork content for your own use, as well as an introduction to how the content is organized and the tools we use to display content.

This repository was developed by the DataONE Community Engagement and Outreach Working Group and continues to be maintained by members of this team. Thank you for your interest in contributing to these educational materials.

Update current content

Want to update a link or method? See a spelling error? Changes can be easily proposed by opening the GitHub Education page and editing content directly. For help, try this brief GitHub tutorial on forking and editing content.

Edit content

1. Create a fork of the lessons or best practices repository into your github account, depending on which content you wish to edit.
2. Modify the files that you want to change (See “Structure” below for tips on making changes).
3. Submit a pull-request against the master branch of this repository.
4. Your changes will be reviewed by the repository admins.

Page not rendering?

Check that the title field of the YAML header (the first line of each lesson) is in quotes.

Suggest changes to content

1. Open an Issue on this repository.
2. Provide your suggested changes with as much detail and guidance as possible. Be specific.
3. Your suggestions will be reviewed by the repository admins.
4. Changes will be pushed to the repository by the repository admins regularly/as needed.
There aren't any open issues.

You could search all of GitHub or try an advanced search.

ProTip! Type `g` on any issue or pull request to go back to the issue listing page.
Guidelines for contributors and content editors

This document details our recommended processes to update current content, suggest changes to content, and fork content for your own use, as well as an introduction to how the content is organized and the tools we use to display content.

This repository was developed by the DataONE Community Engagement and Outreach Working Group and continues to be maintained by members of this team. Thank you for your interest in contributing to these educational materials.

Update current content

Want to update a link or method? See a spelling error? Changes can be easily proposed by opening the GitHub Education page and editing content directly. For help, try this brief GitHub tutorial on forking and editing content.

Edit content
1. Create a fork of the lessons or best practices repository into your github account, depending on which content you wish to edit.
2. Modify the files that you want to change (See "Structure" below for tips on making changes).
3. Submit a pull-request against the master branch of this repository.
4. Your changes will be reviewed by the repository admins.

Page not rendering?

Check that the title field of the YAML header (the first line of each lesson) is in quotes.

Suggest changes to content
1. Open an issue on this repository.
2. Provide your suggested changes with as much detail and guidance as possible. Be specific.
3. Your suggestions will be reviewed by the repository admins.
4. Changes will be pushed to the repository by the repository admins regularly/as needed.

Fork content for your own use

Fork and edit content through GitHub, rather than editing privately, to enable others to use your edited content and to track how these materials are used.
1. Create a fork of the lessons or best practices repository into your github account
2. Modify the files that you want to change (See "Structure" below for tips on making changes)
Edit or fork content for your own use

Select one of the buttons below to open the appropriate GitHub repository:

- Teaching Module
- Best Practice

Hosted by DataONE

In collaboration with the community, DataONE has developed high quality resources for helping educators and librarians with training in data management, including teaching materials, webinars and a database of best-practices to improve methods for data sharing and management.

If you have a question or concern, please open an Issue in this repository on GitHub.
<table>
<thead>
<tr>
<th>File</th>
<th>Description</th>
<th>Last Update</th>
</tr>
</thead>
<tbody>
<tr>
<td>images</td>
<td>adding organization info to bp front matter</td>
<td>3 months ago</td>
</tr>
<tr>
<td>logos</td>
<td>bestpractices citation</td>
<td>3 months ago</td>
</tr>
<tr>
<td>advertise-your-data.md</td>
<td>Add authorship and update date information to BP</td>
<td>2 months ago</td>
</tr>
<tr>
<td>assign-descriptive-file.md</td>
<td>Add authorship and update date information to BP</td>
<td>2 months ago</td>
</tr>
<tr>
<td>backup-your-data.md</td>
<td>Add authorship and update date information to BP</td>
<td>2 months ago</td>
</tr>
<tr>
<td>check-data-and.md</td>
<td>Add authorship and update date information to BP</td>
<td>2 months ago</td>
</tr>
<tr>
<td>choose-and-use.md</td>
<td>Add authorship and update date information to BP</td>
<td>2 months ago</td>
</tr>
<tr>
<td>communicate-data-quality.md</td>
<td>Add authorship and update date information to BP</td>
<td>2 months ago</td>
</tr>
<tr>
<td>confirm-a-match.md</td>
<td>Add authorship and update date information to BP</td>
<td>2 months ago</td>
</tr>
<tr>
<td>consider-the-compatibility.md</td>
<td>Add authorship and update date information to BP</td>
<td>2 months ago</td>
</tr>
<tr>
<td>create-a-data.md</td>
<td>updates to frontmatter and stage blurbs</td>
<td>a month ago</td>
</tr>
<tr>
<td>create-and-document.md</td>
<td>updates to frontmatter and stage blurbs</td>
<td>a month ago</td>
</tr>
<tr>
<td>create-manage-and.md</td>
<td>Add authorship and update date information to BP</td>
<td>2 months ago</td>
</tr>
<tr>
<td>decide-what-data.md</td>
<td>Add authorship and update date information to BP</td>
<td>2 months ago</td>
</tr>
<tr>
<td>define-expected-data.md</td>
<td>Add authorship and update date information to BP</td>
<td>2 months ago</td>
</tr>
<tr>
<td>define-roles-and.md</td>
<td>Add authorship and update date information to BP</td>
<td>2 months ago</td>
</tr>
<tr>
<td>define-the-data.md</td>
<td>Add authorship and update date information to BP</td>
<td>2 months ago</td>
</tr>
</tbody>
</table>
To assure that metadata correctly describes what is actually in a data file, visual inspection or analysis should be done by someone not otherwise familiar with the data and its format. This will assure that the metadata is sufficient to describe the data. For example,
Data Management Skillbuilding Hub

• Structure: Data life-cycle stage
• Current holdings: Education modules and best practices
• Citation: Credit where credit is due
• Editing content (forking!)
• Creating new content (in the works)
• Future holdings
Data Management Skillbuilding Hub

- Structure: Data life-cycle stage
- Current holdings: Education modules and best practices
- Citation: Credit where credit is due
- Editing content (forking!)
- Creating new content (in the works)
- **Future holdings**
Data Management Skillbuilding Hub

- **Structure**: Data life-cycle stage
- **Current holdings**: Education modules and best practices
- **Citation**: Credit where credit is due
- **Editing content**: (forking!)
- **Creating new content**: (in the works)
- **Future holdings**

**Upcoming Webinar:**

**Tutorials:**
Welcome to the DMT Clearinghouse

The Data Management Training (DMT) Clearinghouse is a registry for online learning resources focusing on research data management.

It was created in collaboration between the U.S. Geological Survey's Community for Data Integration, the Earth Sciences Information Partnership (ESIP), and DataONE.

For questions or feedback, please contact clearinghouseEd@esipfed.org

Data Management Training (DMT) Clearinghouse:

A Convenient and Curated Source for Finding Educational Resources on Research Data Management (RDM)

DataONE Webinar
October 9, 2018
Nancy J. Hoebelheinrich

Knowledge Motifs LLC
Mapping sensible data relationships
Research Data Management Training – Sometimes you or your research team need it,

• Why is training needed?
  • Motivators:
    – Funders require
    – Publishers are beginning to require
    – For scientific reproducibility
    – For data re-use by colleagues & collaborators
    – Community culture beginning to expect open data
    – Others...?
How do I learn / teach RDM?

Who else is doing?

Whoa! Too many choices! Which is right for me?

Search
Browse
Submit

Web Portal / Metadata Content Repository / Clearinghouse

Lab / Classroom / 1 on 1

Self – taught Researcher

Excerpted from the SGCI BootCamp “Pitch Deck”
Introducing the ESIP-hosted Data Management Training Clearinghouse!

http://dmtclearinghouse.esipfed.org/
What is the Data Management Training (DMT) Clearinghouse??

What?

- *Metadata registry* for educational resources on research data management
- Capabilities include:
  - Search
  - Browse
  - Submit
- Collaboratively developed & maintained

What kind of training resources?

- Short courses ala “Kahn Academy” (7 – 15 min. modules)
- Videos
- Learning activities to supplement courses
- Presentations & webinars
- Data “recipes”
- Syllabi & curricula

http://dmtclearinghouse.esipfed.org/
Why use the DMT Clearinghouse??

Answers to these Researcher questions:

-- What kind of resources are available on RDM & where did they come from?
-- Do they pertain to my subject domain?
-- Do they fit the data management framework of my organization?
-- Are they appropriate for my role on my research team?
-- What do they cost?

Answers to these Data Specialist questions:

-- I’ve been asked to provide some training to a research team on RDM. What have others done that I can adapt?
-- Where can I find practical, subject-domain targeted exercises for my generic tutorials on RDM?
-- Colleagues keep asking me for the training resources that I’ve created. Where can I share them easily?
Where, how is the DMT Clearinghouse maintained??

Currently, hosted & maintained
- On & by the ESIP Federation Commons – a Drupal based content management system
- Don’t need to register for Search / Browse / Suggest a Resource to add
- Do not need an ESIP Acct to “submit” a resource, unless planning to create full description

Sustainability Plan
- Crowd-sourced submissions
- Domain-knowledgeable reviewers & editors to maintain quality & currency of resources
- Always seeking user interface & functionality feedback
- After initial seed $$, have been funded by IMLS for 3 year National Leadership Grant
- Exploring options for longer term Sustainability from NSF
Let’s take a look!

http://dmtclearinghouse.esipfed.org/
Browse Function
Browse Function

All Learning Resources

DataONE Data Management Module 02: Data Sharing
- DataONE Education Modules
- Analyze, Assess, Collect, Describe, Discover, Integrate, Plan, Preserve
- May 2012

DataONE Data Management Module 03: Data Management Planning
- DataONE Education Modules
- Plan
- May 2012

DataONE Data Management Module 01: Why Data Management
- DataONE Education Modules
- Analyze, Assess, Collect, Describe, Discover, Integrate, Plan, Preserve
- May 2012

DataONE Data Management Module 05: Data Quality Control and Assurance
- DataONE Education Modules
- Assess, Collect
- May 2012

DataONE Data Management Module 07: Metadata
- DataONE Education Modules
- Describe
- May 2012

Mozilla Science Lab’s Open Data Primers

Mozilla Science Lab Open Data Instructor Guides
Browse Function

**All Learning Resources**

<table>
<thead>
<tr>
<th>Framework</th>
<th>DataONE Data Management Module 02: Data Sharing</th>
<th>DataONE Data Management Module 03: Data Management Planning</th>
<th>DataONE Data Management Module 01: Why Data Management</th>
<th>DataONE Data Management Module 05: Data Quality Control and Assurance</th>
<th>DataONE Data Management Module 07: Metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESIP Data Management for Scientists Short Course</td>
<td>Analyze</td>
<td>Assure</td>
<td>Collect</td>
<td>Describe</td>
<td>Discover</td>
</tr>
<tr>
<td>ICSU - World Data System Training Resources Guide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Digital Preservation Network</td>
<td>DataONE Education Modules</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USGS Science Support Framework</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mozilla Science Lab's Open Data Primers

Mozilla Science Lab Open Data Instructor Guides
What is an educational framework?

- An [educational] framework is a plan or set of steps that defines or collects the content using clear, definable standards about what the student should know and understand.

- For purposes of the DMT Clearinghouse, a given learning resource may be associated with a community-defined standard for data management.

- For example, the DataONE framework represents the DataONE’s “Data Life Cycle”.

https://www.dataone.org/data-life-cycle

The DataONE data life cycle was developed ...in collaboration with the broader DataONE community ... [and] serves as an underlying framework for the development of tools, services and education materials by DataONE.
Browse Function

From ~250 ‘published’ resources ➔ ~10
Browse Function -- you found one that looks useful! What next?

You can either click on the **Title** or the **More button** to get more info.
Browse Function -- you found one that looks useful from the brief description!

What next?
Browse Function -- you want to see more about this one! What next?

You can either ...

Click on the **View full description** to look at the full metadata ... or the **View resource** button to go directly to the “landing page” of the resource.

---

**ORNL DAAC Data Management Workshops**

Educational workshops on various scientific data management best practices designed to (1) introduce new data collectors to best practices in data curation and (2) enhance the skillsets of experienced data providers. New workshops are added as they are made available.

**View full description**

---

**ICSU - World Data System Training Resources Guide**

---

---
Search Function
Search Function: 3 approaches

You filter by pre-set categories ("filters")

You enter your own search terms
Search Function: Approach 3: Combining Approach 1 & Approach 2 ...
Search Function: Approach 1 = You search using your own terms
Search Function: Approach 2 = You start by checking boxes within the preset categories ("filters")
Built-in Search Filters

**Filter**
- Framework
  - ESIP Data Management for Scientists Short Course (33)
  - USGS Science Support Framework (18)
  - DataONE Education Modules (10)
  - ICSU - World Data System Training Resources Guide (8)
  - The Digital Preservation Network (7)
- Keywords
  - Data management (32)
  - Data sharing (27)
  - Data management planning (23)
  - Data preservation (15)
  - Data life cycle (11)
  - Data citation (10)
  - Data reuse - Core Trustworthy Data Repositories Requirements (10)
  - Community standards (9)
  - Data access methods (9)
  - Metadata (9)
- Organizations
  - Federation of Earth Science Information Partners (ESIP Federation) (34)
  - U.S. Geological Survey (16)
  - DataONE (10)

**People**
- Nancy J. Hoebelheinrich (33)
- Ruth E. Duerr (33)
- Matthew Mayernik (9)
- Robert R. Downs (7)
- Robert Cook (6)
- Chen Chiu (2)
- Curt Timles (2)
- Jason Kudulis (2)
- Kathy Martinovich (2)
- Linda Olsen (2)
- Mimi Tzeng (2)
- Suresh K.S. Vannan (2)
- Tyler Stevens (2)
- Zannah Marsh (2)
- Ben Wheeler (1)
- Dave Fearon (1)
- Dorothea Salo (1)
- Drew Ignizio (1)
- Elliot Metsger (1)
- Emily Fort (1)

**License**
- Creative Commons Attribution 3.0 United States - CC BY 3.0 US (35)
- Creative Commons 0 - CC0 "No Rights Reserved" (Public Domain) (31)
- Creative Commons Attribution 4.0 International - CC BY 4.0 (15)
- Creative Commons Attribution-NonCommercial 4.0 International - CC BY-NC 4.0 (5)
- Creative Commons 1.0 Universal (Public Domain Dedication) (3)
- Creative Commons Attribution-ShareAlike 4.0 International License - CC BY-SA 4.0 (3)
- Creative Commons Attribution 3.0 Unported - CC BY 3.0 (1)

**Publication Date**
- 2012 (31)
- 2017 (15)
- 2013 (13)
- 2015 (9)
- 2016 (5)
- 2014 (2)
- 2010 (1)

**Cost**
- No Fee (101)
- Fee (3)
Demonstrating Approach 3: Start by entering your own search term...

...then limiting by Keyword filter...
Demonstrating Search: finally, limiting even more by the Framework filter...
Demonstrating Search: however, if too precise, you can either **Clear all** to start over ... or uncheck filters / facets.
### ORNL DAAC Data Management Workshops

**Key Info**

**URL** - the landing page for the learning resource:
http://daac.ornl.gov/workshops/workshops.shtml

**Description** - a brief synopsis, abstract or summary of what the learning resource is about:
Educational workshops on various scientific data management best practices designed to (1) introduce new data collectors to best practices in data curation and (2) enhance the skillsets of experienced data providers. New workshops are added as they are made available.

**Authoring Organization(s) Name:**
NASA ORNL DAAC (Oak Ridge National Laboratory Distributed Active Archive Center)

**License** - link to legal statement specifying the copyright status of the learning resource:
Creative Commons 1.0 Universal (Public Domain Dedication)

**Access Cost:**
No fee

**Primary language(s) in which the learning resource was originally published or made available:**
English

### More info about

**Keywords** - short phrases describing what the learning resource is about:
- Appraisal - Core Trustworthy Data Repositories Requirements
- Data discovery and identification - Core Trustworthy Data Repositories Requirements
- Data integrity and authenticity - Core Trustworthy Data Repositories Requirements
- Data quality - Core Trustworthy Data Repositories Requirements
- Data reuse - Core Trustworthy Data Repositories Requirements
- Expert guidance - Core Trustworthy Data Repositories Requirements
- Preservation plan - Core Trustworthy Data Repositories Requirements
Subject Discipline - subject domain(s) toward which the learning resource is targeted:
- Engineering
- Aerospace Engineering
- Physical Sciences and Mathematics
- Earth Sciences
- Physical Sciences and Mathematics
- Environmental Sciences

Publisher - organization credited with publishing or broadcasting the learning resource:
NASA ORNL DAAC (Oak Ridge National Laboratory Distributed Active Archive Center)

Media Type - designation of the form in which the content of the learning resource is represented, e.g., moving image:
Collection - a group or set of items that comprise a single learning resource, e.g., a PDF version of a slide presentation, an audio file of the presentation and a textual representation of the oral transcription of the presentation.

Contributor Organization(s):
- Name:
  - ICSU - World Data System (WDS)
- Type:
  - Endorser

Contact Organization(s):
NASA ORNL DAAC (Oak Ridge National Laboratory Distributed Active Archive Center)

Educational Info

Purpose - primary educational reason for which the learning resource was created:
Professional Development - Increasing knowledge and capabilities related to managing the data produced, used or re-used, curated and/or archived.

Learning Resource Type - category of the learning resource from the point of view of a professional educator:
Lesson - detailed description of an element of instruction in a course, contained in a unit of one or more lessons, and used by a teacher to guide class instruction.

Target Audience - intended audience for which the learning resource was created:
- Data supporter
- Early-career Research Scientist
- Research Scientist

Intended time to complete - approximate amount of time the average student will take to complete the learning resource:
More than 1 hour (but less than 1 day)

Framework - A community-based organization plan or set of steps for education or training:
- ICSU - World Data System Training Resources Guide
Submit Function, briefly
Submit Function, briefly

Submit new Learning Resource

Your learning resource suggestion will be submitted to the Clearinghouse, but will not be published immediately as the information needs to be reviewed for quality control and relevancy.

If you would like to submit more information, please log in and return to this page. If you don't yet have an ESIP account, you can create one by clicking "Log in" above, then "I want to create an account."

From the Help pages, you can also find out more about how and what kind of information to submit. We do require that you give us your name and email address if you submit without having a user login in case reviewers or editors have questions. Rest assured that your contact information will not be shared publicly without your permission.

Thank you for your interest in making data management training resources widely available!

Title *

URL - the landing page for the learning resource

https://oedbreeze.cr.usgs.gov/dm-value/

Access Cost *
- No fee
- Fee

Submission Contact Name *

Submission Contact Email Address *

Address will not be shared without permission

CAPTCHA

This question is for testing whether you are a human visitor and to prevent automated spam submissions.

[ ] I'm not a robot

[RECAPTCHA]
As a community supported resource we’d love to have you to join us by...

- Submitting your learning resources
- Joining our Working Groups on
  - Assessment Framework
  - Metadata Enhancement
  - Content Diversification
- Editorial assistance
- Usability testing
- Spreading the word
- Jumping in on our crowdsourcing events
Join us!

Join the ESIP Research Data Management Cluster at: esip_dmtraining@lists.esipfed.org
Contact:
Nancy Hoebelheinrich (nhoebel@kmotifs.com) or clearinghouseEd@esipfed.org

http://dmtclearinghouse.esipfed.org/
Data Management Skillbuilding Hub

- One off lessons
- Host/store here
- Indexed at DMT Clearinghouse

DMT Clearinghouse

- Index here
- Metadata
- This is a registry

http://dataoneorg.github.io/Education

http://dmtclearinghouse.esipfed.org

Please take some time and come ask us questions at the help desk

Megan Mach mach@unm.edu
Nancy Hoebelheinrich nhoebel@kmotifs.com
Amber Budden aebudden@dataone.unm.edu
DataONE Webinar Series
www.dataone.org/webinars
Upcoming Webinar Event
www.dataone.org/upcoming-webinar
Previous Webinar Events (Recording and Discussion)
www.dataone.org/previous-webinars

#DWS2018
@DataONEorg
Upcoming Webinar Event
www.dataone.org/upcoming-webinar

Schema.org: Improving access to data through a standardized language
November 13, 2018

Bryce Mecum, Scientific Software Engineer, National Center for Ecological Analysis and Syntheis (NCEAS)
Doug Fils, Data Management Technical Expert, Consortium for Ocean Leadership (COL)