

# Data Help Desk

Get Answers to Your Data Management Questions:  
Workshops, Demos and More ...



Booth 219-223 <http://bit.ly/DHDESA19> #datahelpdesk

# Data Management Skillbuilding Hub

[dataoneorg.github.io/Education/](https://dataoneorg.github.io/Education/)

# Data Management Skillbuilding Hub



[Home](#) [Contribute](#) [FAQ](#) [GitHub](#)



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

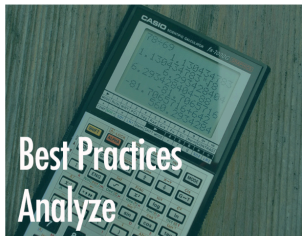
The resources presented on the Data Management Skillbuilding Hub can be updated by users to promote a current, well-maintained, and sustainable educational tool. Learn more about how you can [contribute](#).

### 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](#).

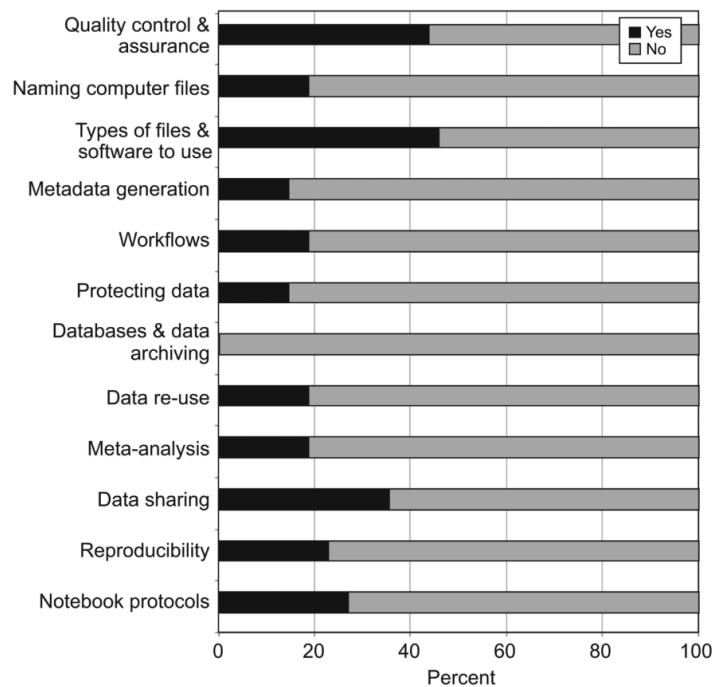
» Filter by content type:

» Filter by stage of the Data Life Cycle:



# Status of Data Management Education

## Survey of Ecology Courses

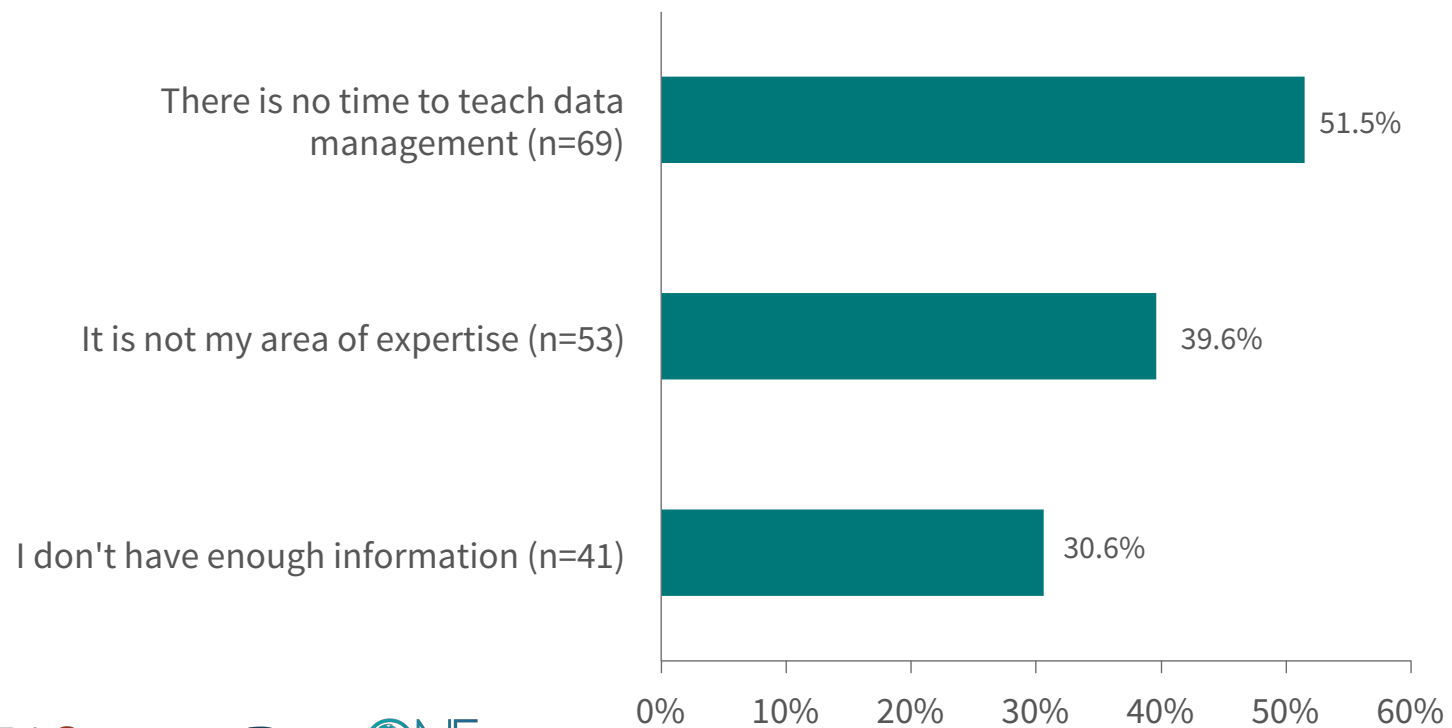


Percent of ecology courses that address and/or teach the data management topics listed

Strasser and Hampton 2012

# Challenges in Data Management Training

## Survey of Educators



# Training

## Leadership in Data Management Education

The collage features several educational resources from DataONE:

- Primer on Data Management: What you always wanted to know\*** by Carly Strasser, Robert Cook, and William Michener. It includes sections on objectives, why manage data, the data life cycle, and best practices.
- Lesson 10: Analysis and Workflows** slide, detailing typical data analyses (data processing, graphical analysis, statistical analysis) and formal workflows (analytical pipeline, Kepler software example).
- Lesson 5: Data Quality Control and Assurance** video player interface, showing a presentation slide with a scatter plot and a workflow diagram.

40  
Webinars

3421  
Unique Webinar  
Attendees



Greg Wilson



Liz Ferguson



Stephanie Hampton



Fernando Pérez

# Training

## Training Events



“Thanking my lucky stars for @DataONEorg education modules this week! Data entry exercise adapted nicely for online participation.”

“Thank you! I love these materials.”

“@DataONEorg website has great education resources on managing data w/ best practices.”

“Great collection of free education material on data management by @DataONEorg.”

“Thanks @DataONEorg for making these education modules available CCO!”

“Awesome presentations about data management. The Metadata module is great!! I like them. Thanks @DataONEorg”

## Community Perspectives





# Education Materials

## The Evolution of Resources

**DataONE** www.dataone.org  
Primer on Data Management: What you always wanted to know\*  
\* but were afraid to ask  
Carly Strasser, Robert Cook, William Michener

**Contents**  
1. Objective of This Primer  
2. Why Manage Data?  
2.1. It will benefit you and your colleagues  
2.2. It will benefit the scientific community  
2.3. Journals and sponsors want you to  
3. How To Use This Primer  
4. The Data Life Cycle: An Overview  
5. Data Management Throughout the Data Life Cycle  
5.1 Plan  
5.2 Collect  
5.3 Assure  
5.4 Describe: Data Documentation  
5.5 Preserve  
5.6 Discover, Integrate, and Analyze  
6. Conclusion  
7. Acknowledgements  
8. References  
9. Glossary

**1. Objective of This Primer**  
The goal of data management is to produce self-colleague who has not been involved with your able to use it effectively and properly? This primer practices that will enable you to develop a data organize, manage, describe, preserve and share

**2. Why Manage Data?**  
2.1. It will benefit you and your collaborators  
Establishing how you will collect, document, use your research project has many benefits. You will research by investing the time and energy before easier for you to find, use, and analyze, and it will your data. In the long term, following good data with the project can find, understand, and use it recommending appropriate ways to cite your data and their use [1].

DataONE Best Practices Primer

**DataONE Lesson 10: Analysis and Workflows**  
View all Education Modules at <https://www.dataone.org/education-modules>

**Typical data analyses**  
**Data processing:** may include selecting a subset of data for analysis, merging multiple data sets, manipulating data for usability, 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.

**Formal Workflow**  
Analytical pipeline where each step can be implemented in 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

**Formal workflow example: Kepler software**  
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!

**Workflows**  
**Definition:** Precise description of the procedures used in a project. 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**

**DataONE Tutorials on Data Management**  
Lesson 5: Data Quality Control and Assurance

DataONE Education Module 05: Data Quality, Control and Assurance 624 views



# Education Materials

## The Evolution of Resources

**DataONE** www.dataone.org  
**Primer on Data Management: What you always wanted to know\***  
\* but were afraid to ask  
Carly Strasser, Robert Cook, William Michener

**Contents**  
1. Objective of This Primer  
2. Why Manage Data?  
2.1. It will benefit you and your colleagues  
2.2. It will benefit the scientific community  
2.3. Journals and sponsors want you to  
3. How To Use This Primer  
4. The Data Life Cycle: An Overview  
5. Data Management Throughout the Data Life Cycle  
5.1. Plan  
5.2. Collect  
5.3. Assure  
5.4. Describe: Data Documentation  
5.5. Preserve  
5.6. Discover, Integrate, and Analyze  
6. Conclusion  
7. Acknowledgements  
8. References  
9. Glossary

**1. Objective of This Primer**  
The goal of data management is to produce self-colleague who has not been involved with your able to use it effectively and properly? This primer practices that will enable you to develop a data organize, manage, describe, preserve and share

**2. Why Manage Data?**  
2.1. It will benefit you and your collaborators  
Establishing how you will collect, document, use your research project has many benefits. You will research by investing the time and energy before easier for you to find, use, and analyze, and it will your data. In the long term, following good data with the project can find, understand, and use it recommending appropriate ways to cite your data and their use [1].

**DataONE Best Practices Primer**

**DataONE** Lesson 10: Analysis and Workflows  
View all Education Modules at <https://www.dataone.org/education-modules>

**Typical data analyses**  
**Data processing:** may include selecting a subset of data for analysis, merging multiple data sets, manipulating data for usability, 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.

**Formal Workflow**  
Analytical pipeline where each step can be implemented in 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

**Formal workflow example: Kepler software**

**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!

**Workflows**  
**Definition:** Precise description of the procedures used in a project. 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**

**Tutorials on Data Management**  
Lesson 5: Data Quality Control and Assurance

DataONE Education Module 05: Data Quality, Control and Assurance 624 views

**Data Management Skillbuilding Hub** DataONE

Home Contribute FAQ GitHub

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

- 01. Why Data Management
- 02. Data Sharing
- 03. Data Management Planning
- 04. Data Entry and Manipulation
- 05. Data Quality Control and Assurance
- 06. Protecting Your Data
- 07. Metadata
- 08. Data Citation
- 09. Analysis and Workflows
- 10. Legal and Policy Issues
- Best Practices Analyze
- Best Practices Assure
- Best Practices Collect
- Best Practices Describe
- Best Practices Discover
- Best Practices Integrate

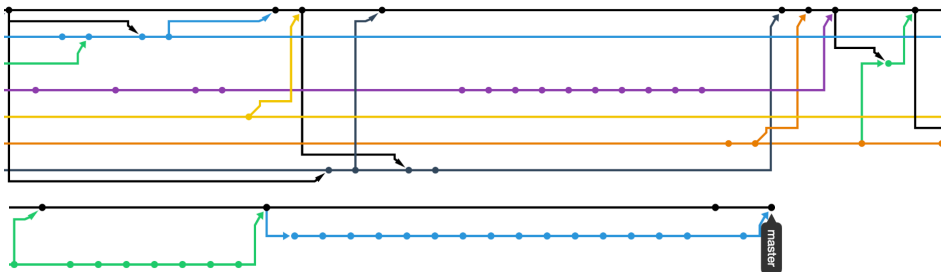
# Open Community Platform



# GitHub

Clone or download ▾

Fork 11



### Data Management Skillbuilding Hub

Home Contribute FAQ GitHub

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

01. Why Data Management	02. Data Sharing	03. Data Management Planning	04. Data Entry and Manipulation
05. Data Quality Control and Assurance	06. Protecting Your Data	07. Metadata	08. Data Citation
09. Analysis and Workflows	10. Legal and Policy Issues	Best Practices Analyze	Best Practices Assure
Best Practices Collect	Best Practices Describe	Best Practices Discover	Best Practices Integrate

# Data Management Skillbuilding Hub



[Home](#) [Contribute](#) [FAQ](#) [GitHub](#)



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

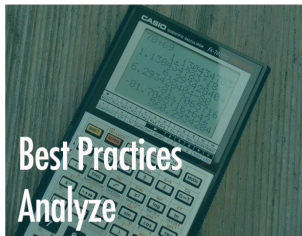
The resources presented on the Data Management Skillbuilding Hub can be updated by users to promote a current, well-maintained, and sustainable educational tool. Learn more about how you can [contribute](#).

## 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:

» Filter by stage of the Data Life Cycle




Browser window showing the URL: [https://dataoneorg.github.io/Education/lessons/02\\_datasharing/index.html](https://dataoneorg.github.io/Education/lessons/02_datasharing/index.html)

# Data Management Skillbuilding Hub

Home   Contribute   FAQ   GitHub

**TEACHING MODULE**



## Presentation View

*Quick tips: Press p for presentation; f for full screen*

**Supporting downloads:**

- PDF Download
- PPT Download
- Handout
- Hands-on Exercise

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:**  
DataONE Community Engagement & Outreach Working Group (2017) "Data Sharing". Accessed through the Data Management Skillbuilding Hub at [https://dataoneorg.github.io/Education/lessons/02\\_datasharing/index](https://dataoneorg.github.io/Education/lessons/02_datasharing/index) on May 24, 2019

**DataONE**

[Home](#)

Version date: Apr 06, 2017

**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.

Data Life Cycle: describe x +

https://dataoneorg.github.io/Education/bp\_step/describe/

# Data Management Skillbuilding Hub

Home Contribute FAQ GitHub

**BEST PRACTICE**

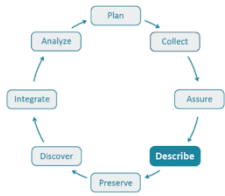
## Best Practice: Describe

Select a Best Practice below to learn more about the "Describe" stage in the *Data Life Cycle*.

### What is the "Describe" stage?

Document data by describing the why, who, what, when, where, and how of the data. Metadata, or data about data, are key to data sharing and reuse, and many tools such as standards and software are available to help describe data.

More information can be found in the [Best Practices Primer](#).



```

graph TD
    Plan --> Collect
    Collect --> Assure
    Assure --> Describe
    Describe --> Preserve
    Preserve --> Discover
    Discover --> Integrate
    Integrate --> Analyze
    Analyze --> Plan
  
```

### Assign descriptive file names

File names should reflect the contents of the file and include enough information to uniquely identify the data file. File names may contain information such as project acronym, study title, location, investigator, year(s) of study, data type, version n... [\(click for more\)](#)

Tags: [access](#) [describe](#) [discover](#) [format](#)

### Choose and use standard terminology to enable discovery

Terms and phrases that are used to represent categorical data values or for creating content in metadata records should reflect appropriate and accepted vocabularies in your community or institution. Methods used to identify and select the proper termin... [\(click for more\)](#)

Tags: [controlled vocabulary](#) [describe](#) [documentation](#) [metadata](#) [ontologies](#) [preserve](#) [standards](#)

### 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 da... [\(click for more\)](#)

Tags: [assure](#) [data consistency](#) [describe](#) [documentation](#) [metadata](#) [quality](#)

### Create a data dictionary

A data dictionary provides a detailed description for each element or variable in your dataset and data model. Data dictionaries are used to document important and useful information such as a descriptive name, the data type, allowed values, units, and ... [\(click for more\)](#)

Tags: [controlled vocabulary](#) [describe](#) [documentation](#) [metadata](#) [terminology](#) [units](#)

### Best Practices by Data Life Cycle

- All
- Plan
- Collect
- Assure
- Describe
- Preserve
- Discover
- Integrate
- Analyze

**Learn more:**  
[BP Primer](#)

Contributing to Lessons x +  
https://dataoneorg.github.io/Education/CONTRIBUTING

# Data Management Skillbuilding Hub

Home Contribute FAQ GitHub

## 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)

Best Practice: How to write a ...

https://dataoneorg.github.io/Education/bestpractices/how-to-write

# Data Management Skillbuilding Hub

Home Contribute FAQ GitHub

DataONE

## BEST PRACTICE

### How to write a Best Practice file

Data Life Cycle stage(s): [Life cycle stage here](#)

#### Filename

Each file is named as the first three words of the title, separated by dashes. This page's title is "How to write a best practice", therefore the filename is `how-to-write.md`.

#### Header

The header (aka: front matter) for best practices has the following fields:

- `title` : the title of the webpage, as it will appear in the lists
- `layout` : is always `bestpractice` (NB this will likely be done automatically using the collections features from jekyll)
- `tags` : a list of (short) keywords describing the content of the best practice text
- `step` : a list of one or more steps of the data lifecycle to which this best practice applies
- `related` : a list of related best practices identifiers – an identifier is the first three words of the title, separated by dashes (optional)
- `update` : the date this best practice was created
- `author` : a list of authors that created the best practice
- `organization` : name of organization that oversaw the creation of the best practice
- `org_url` : website of the organization, organization logo will open this webpage when selected
- `org_logo` : name of the organization's logo file, this must me a png
- `categories` : this must be listed as ["Best Practice"], used for sorting and accessing education materials

**in 'Raw' view, these are the headers you should have**

content

#### Best Practices by Data Life Cycle

- All
- Plan
- Collect
- Assure
- Describe
- Preserve
- Discover
- Integrate
- Analyze

**Learn more:**  
[BP Primer](#)



Edit Content in GitHub x +

https://dataoneorg.github.io/Education/githubdirect

# Data Management Skillbuilding Hub

Home Contribute FAQ GitHub

DataONE

## 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.

[dataoneorg.github.io/Education/](https://dataoneorg.github.io/Education/)



## Reproducible Research Techniques for Synthesis

A five day immersion into widely adopted R-based tools for open science



# NCEAS.UCSB.EDU/LEARNING-HUB/SHORT-COURSE

## Curriculum at a Glance

### Enable data reuse through better data management

Metadata - how to write a quality data description  
Data modeling - tidy data for efficient access and storage  
Data publishing, citation and credit

### Build reproducible scientific workflows

Data munging with R tidyverse  
Working collaboratively - git and GitHub  
Writing functions in R  
Building packages for publishing reproducible research

### Communicate results effectively

Literate analysis with RMarkdown  
Publishing analytical web pages with GitHub pages  
Data visualization with ggplot and leaflet

## Details

### Dates:

November 4-8, 2019  
February, 2020 TBD  
May, 2020 TBD

### Cost:

\$2,100  
*Includes:* 5 days of instruction, coffee, light snacks and lunch.  
*Does not include:* travel, lodging costs, breakfast or dinner.

### Location:

NCEAS, 735 State St. Suite #300  
Santa Barbara, CA