



# DataONE

**DataONE Data Management Training -  
the different approaches to data  
management training**

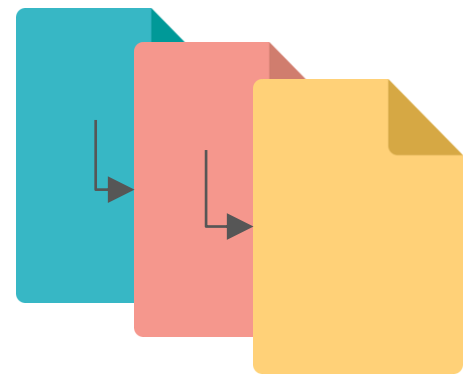
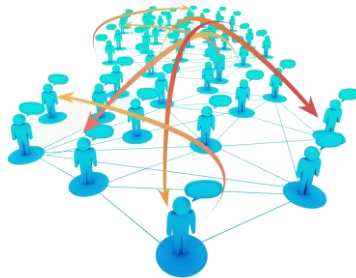
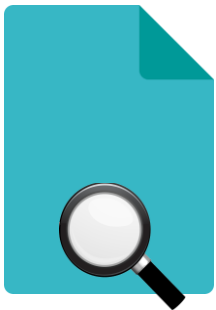


Rebecca Koskela  
Belmont Forum Digital Skills and Curricula  
Development Workshop  
April 28, 2017

# Development of Resources

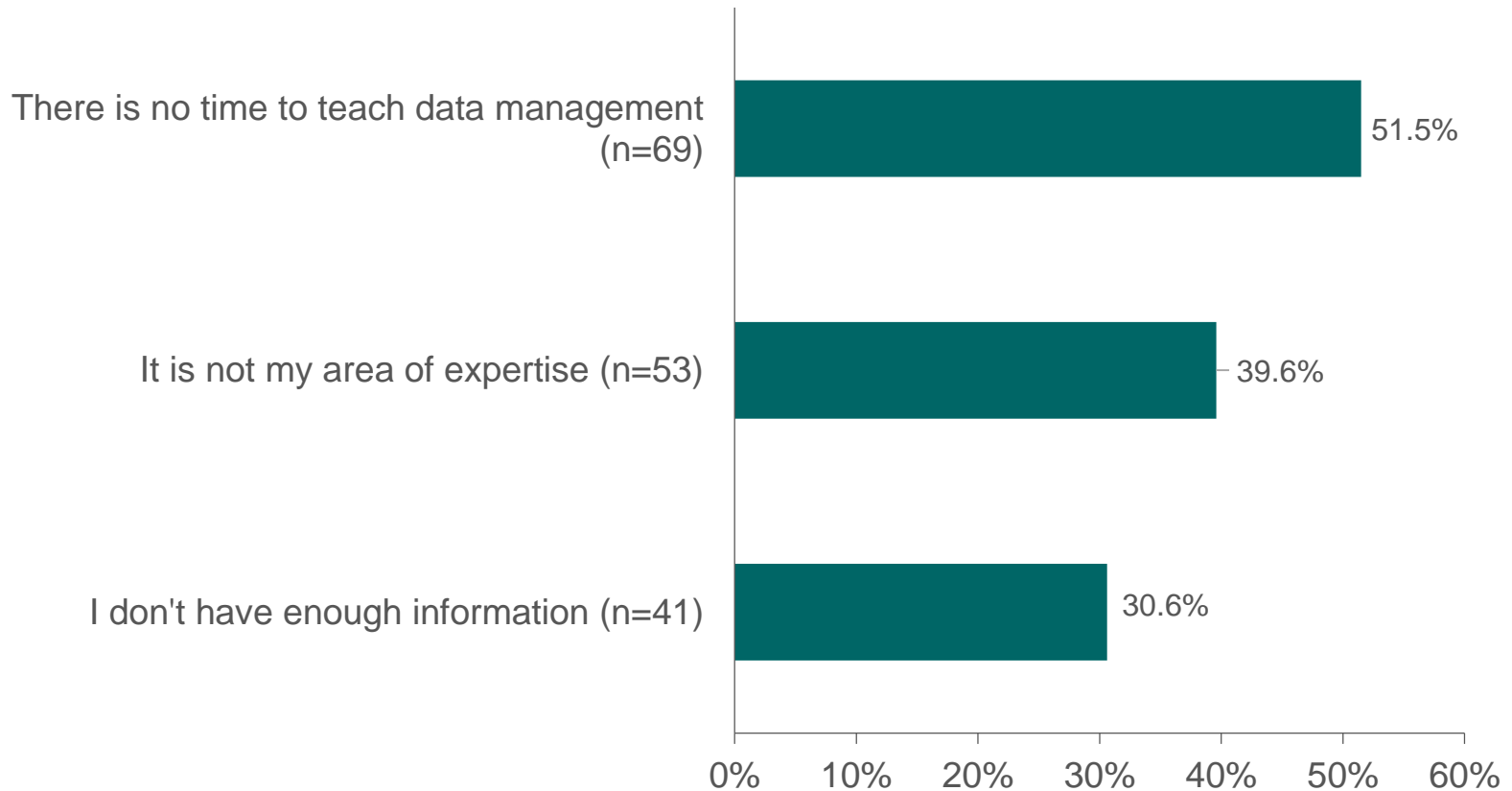
## Approach

- Solicit community feedback
- Build partnerships
- Append, create, revise offerings



# Challenges Educators

Barriers to Teaching Data Management (n=134)





# Best Practices Database and Primer

## Best Practices

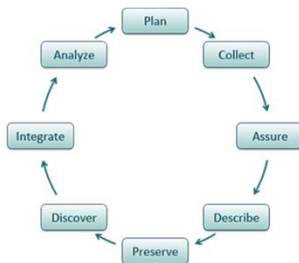
The DataONE Best Practices database provides individuals with recommendations on how to effectively work with their data through all stages of the data lifecycle. Users can access best practices within the database by either clicking on a stage of the lifecycle or selecting keywords under search.

### Best Practices Primer

For students and others new to data management, we provide a **Best Practices Primer** as an introduction to the DataONE Best Practices database and data management in general.

### Public Participation in Science Research Data Management Guide

We also provide a **Data Management Guide** written specifically for the Citizen Science community that takes the users through the steps of the data lifecycle and links to various DataONE Best Practices online.



## Search Best Practices

### Search by Keyword in title

### Search by Keyword in Body

### Filter by tag

access  
analyze  
annotation  
assure  
backup  
calibration  
citation  
coding  
collect

You may enter multiple tags by holding down command (control) and making your selection

### Filter by Data Life Cycle

- Any -



www.dataone.org

## Primer on Data Management: What you always wanted to know\*

\* but were afraid to ask

Carly Strasser, Robert Cook, William Michener, Amber Budden

### Contents

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

### 1. Objective of This Primer

The goal of data management is to produce self-describing data sets. If you give your data to a scientist or colleague who has not been involved with your project, will they be able to make sense of it? Will they be able to use it effectively and properly? This primer describes a few fundamental data management practices that will enable you to develop a data management plan, as well as how to effectively create, organize, manage, describe, preserve and share data.

### 2. Why Manage Data?

#### 2.1. It will benefit you and your collaborators

Establishing how you will collect, document, organize, manage, and preserve your data at the beginning of your research project has many benefits. You will spend less time on data management and more time on research by investing the time and energy before the first piece of data is collected. Your data also will be easier for you to find, use, and analyze, and it will be easier for your collaborators to understand and use your data. In the long term, following good data management practices means that scientists not involved with the project can find, understand, and use the data in the future. By documenting your data and recommending appropriate ways to cite your data, you can be sure to get credit for your data products and their use [1].

DataONE Best Practices Primer

1

# Best Practices Metrics

**Best Practices**

The DataONE Best Practices database provides individuals with recommendations on how to effectively work with their data through all stages of the data lifecycle. Users can access best practices within the database by either clicking on a stage of the lifecycle or selecting keywords under search.

**Best Practices Primer**  
For students and others new to data management, we provide a **Best Practices Primer** as an introduction to the DataONE Best Practices database and data management in general.

**Public Participation in Science Research Data Management Guide**  
We also provide a **Data Management Guide** specifically for the Citizen Science community that takes the users through the steps of the data lifecycle and links to the **Best Practices Primer**.

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

**Primer**

3500+ unique downloads  
(most downloaded resource)

**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, Amber Budden

**Contents**

1. Objective of This Primer	1
2. Why Manage Data?	1
2.1. It will benefit you and your collaborators	1
2.2. It will benefit your research project	2
2.3. Journals and sponsors want you to share your data	2
3. How To Use This Primer	2
4. The Data Life Cycle: An Overview	3
4.1. Plan	3
4.2. Collect	4
4.3. Assure	4
4.4. Describe: Data Documentation	4
4.5. Preserve	4
4.6. Discover, Integrate, and Analyze	4
5. Conclusion	7
7. Acknowledgements	8

1. **Objective of This Primer**  
The primary objective of this primer is to provide an overview of the DataONE Best Practices database and its resources. It is intended for individuals who are new to data management or who are looking for best practices to follow. The primer describes a few fundamental data management practices that will enable you to develop a data management plan, as well as how to effectively create, organize, manage, describe, preserve, and share data.

2. **Why Manage Data?**  
**2.1. It will benefit you and your collaborators**  
Establishing how you will collect, document, organize, manage, and preserve your data at the beginning of your research project has many benefits. You will spend less time on data management and more time on research by investing the time and energy before the first piece of data is collected. Your data also will be easier for you to find, use, and analyze, and it will be easier for your collaborators to understand and use your data. In the long term, following good data management practices means that scientists not involved with the project can find, understand, and use the data in the future. By documenting your data and recommending appropriate ways to cite your data, you can be sure to get credit for your data products and their use [1].

*DataONE Best Practices Primer* 1

**Web page**

5800+ unique visits  
(4<sup>th</sup> most visited page, 2<sup>nd</sup> most visited area)

**Search Best Practices**

Search by Keyword in title

Search by Keyword in description

**Filter by tag**

- access
- analyze
- annotation
- assure
- collection
- documentation
- citizen science
- coding
- collect

You may enter multiple tags by holding down command (control) and making your selection

Search  Reset

# Data Management Modules Overview

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

DataONE Education Module 09: Data Citation 457 views

**DataONE** Lesson 10: Analysis and Workflows

**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.

**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 procedure 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 process
- Decision points
- Arrows indicating direction of process flow

**Informal Workflow Example**

Diagram showing: Data (Input or output) → Analytical process → Decision

## Hands-on Activity 1: Accessing Data in the Literature

**Associated DataONE Lecture:** Lesson 1: *Why Data Management*

**Objectives:** Students recognize the value of accessibly archived data, by experiencing the challenges of accessing data from published papers.

**Outcomes:** (1) Students can explain why accessible data archiving is valuable. (2) Students can provide strategies for getting data from published papers, and anticipate challenges to accessing the data.

**Time Needed:** One hour out-of-class, 15 – 30 minutes in-class discussion.

**URLs:** Any resource for searching scientific literature (e.g. *Web of Science*, *Google Scholar*, *JSTOR*, *BioOne*).

**Additional Files Needed:** None

**Key Reading:** Carly A Strasser and Stephanie E Hampton. 2012. The fractured lab notebook: undergraduates and ecological data management training in the United States. *Ecosphere* 3:art116. doi: 10.1890/ES12-00139.1

**Notes and Instructions for Instructors:**

An intended take-home lesson of this activity is that access to valuable original data can become difficult or impossible in a short period of time after a paper is published, but this loss of accessibility is avoidable. How easy it is to access original data depends on the field; some fields have developed a culture of data sharing and data accessibility, including genetics, climate studies, and geography. Others do not have this tradition. Because of these field-specific cultures, students' success at accessing data will depend on the topic and question they chose.

It may be worth reviewing with the students the different ways by which scientists access others' data: data tables or published data appendices within a paper, extracting (estimating) data from published graphs, online data archives or data streams (either restricted to journal subscribers or public), writing the author and requesting the data etc.

After students have completed the exercise (see *Student Instructions*, below), have students discuss the challenges that they faced in figuring out how to access data from the published literature that are relevant to their question, and ways the students came up with to deal with the challenges. This can be done as a 15 to 30 minute whole-class discussion or in small groups with a report-out. Things to note include whether accessibility to data varied depending on the question addressed, and whether accessibility depended on how long ago the paper was published. Perhaps culminate the discussion with questions about why data underlying

Hands-on Exercises for Data Management 1  
<http://www.dataone.org/education-modules>

**DataONE**

# Data Management Modules Enhancement

- Comprehensive peer-review and revision
- Transition to GitHub

