Community Resources for Data Management

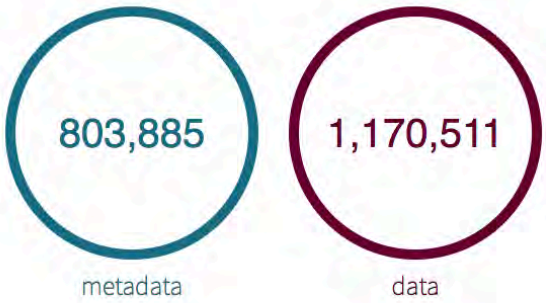
Megan Mach and Amber Budden
DataONE

Nancy Hoebelheinrich
Knowledge Motifs

DataONE Cyberinfrastructure

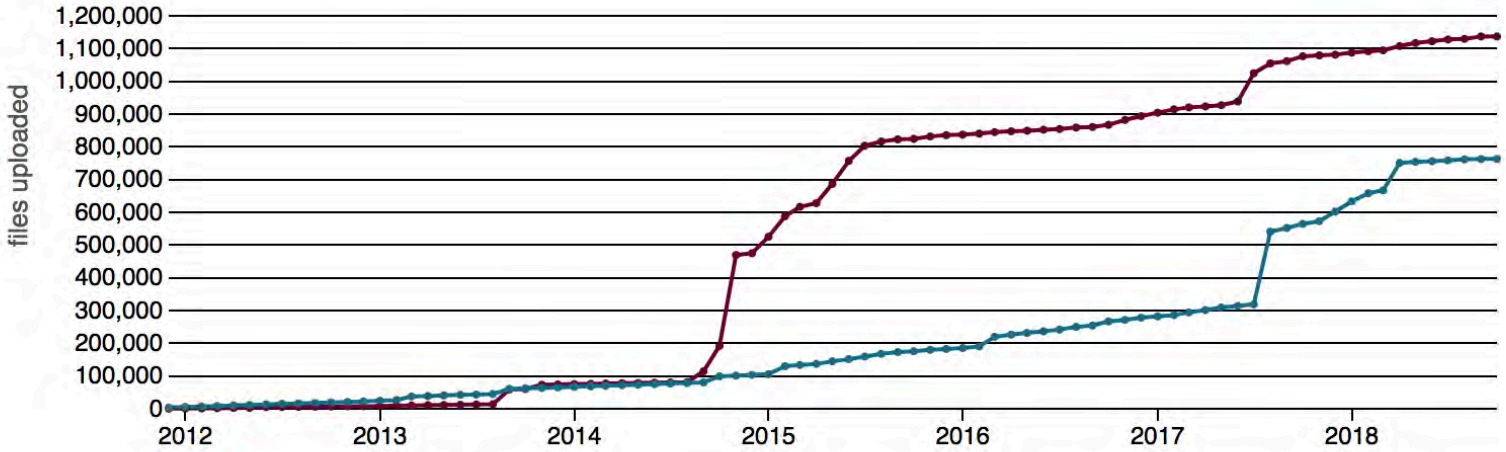
Data Holdings

dataone.org/current-member-nodes



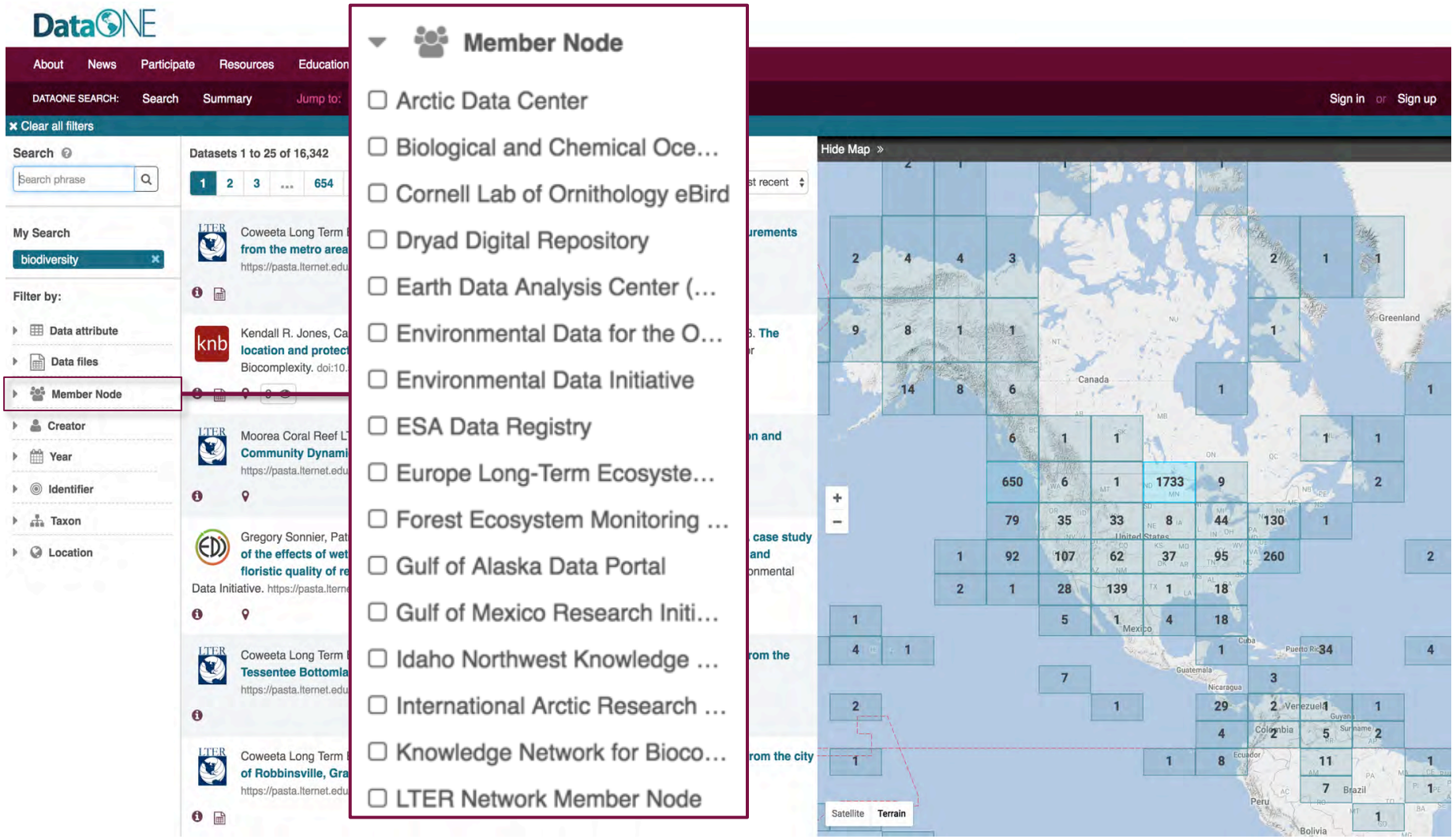
Uploads

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

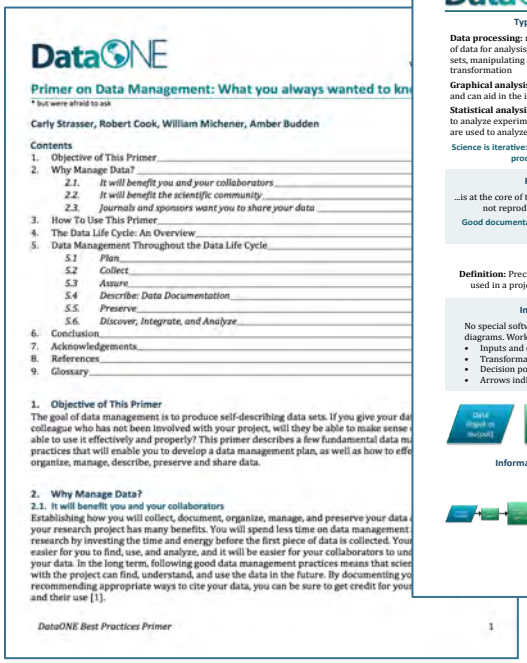


DataONE Search

search.dataone.org



Data Management Education


DataONE

Primer on Data Management: What you always wanted to kn

* but were afraid to ask

Carly Strasser, Robert Cook, William Michener, Amber Budden

Contents

1. Objective of This Primer
2. Why Manage Data?
   - 2.1. It will benefit you and your collaborators
   - 2.2. It will benefit the scientific community
   - 2.3. Journals and sponsors want you to share your data
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

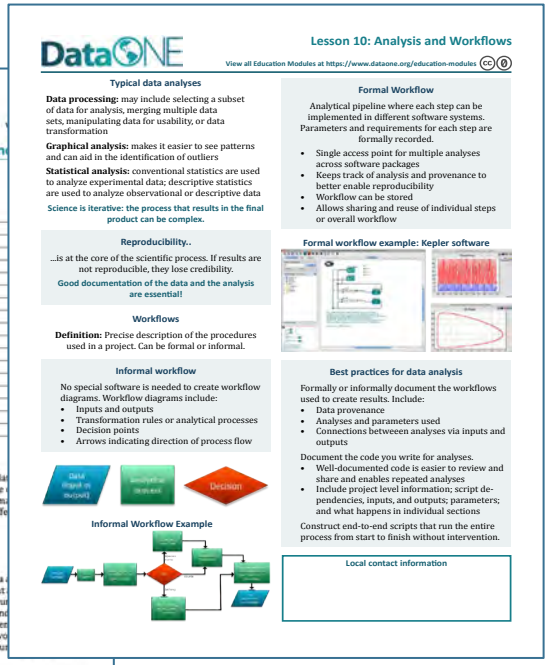
The goal of data management is to produce self-describing data sets. If you give your da colleague who has not been involved with your project, will they be able to make sense able to use it effectively and properly? This primer describes a few fundamental data m practices that will enable you to develop a data management plan, as well as how to effe 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 your research project has many benefits. You will spend less time on data management research by investing the time and energy before the first piece of data is collected. Your easier for you to find, use, and analyze, and it will be easier for your collaborators to und your data. In the long term, following good data management practices means that scien with the project can find, understand, and use the data in the future. By documenting yo recommending appropriate ways to cite your data, you can be sure to get credit for your and their use [1].

DataONE Best Practices Primer 1


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.

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

Decision

Informal Workflow Example

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

Best practices for data analysis

Formally or informally document the workflows used to create results. Include:

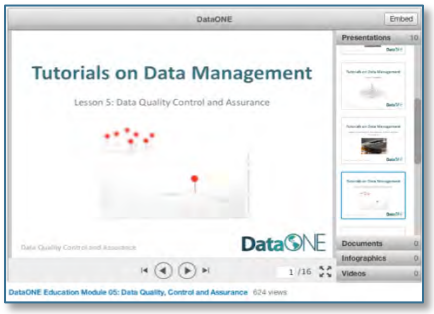
- Data provenance
- Analyses and parameters used
- Connections betweeen analyses via inputs and outputs

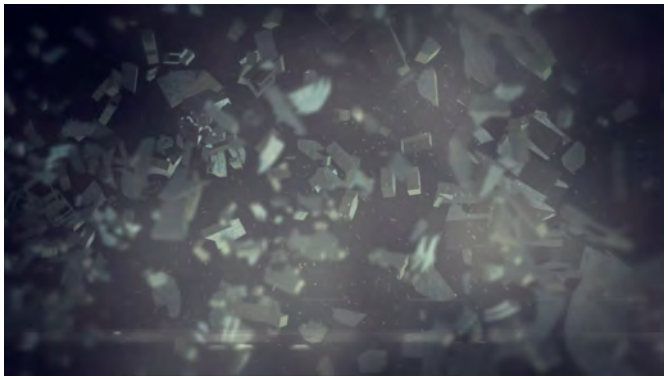
Document the code you write for analyses.

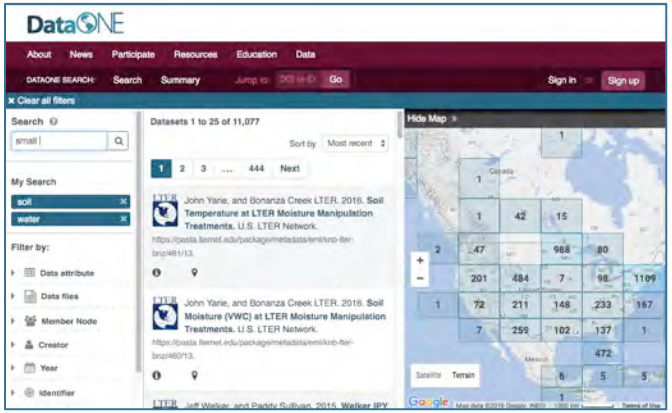
- Well-documented code is easier to review and share and enables repeated analyses
- Include project level information; script dependencies, inputs, and outputs; parameters; and what happens in individual sections

Construct end-to-end scripts that run the entire process from start to finish without intervention.

Local contact information

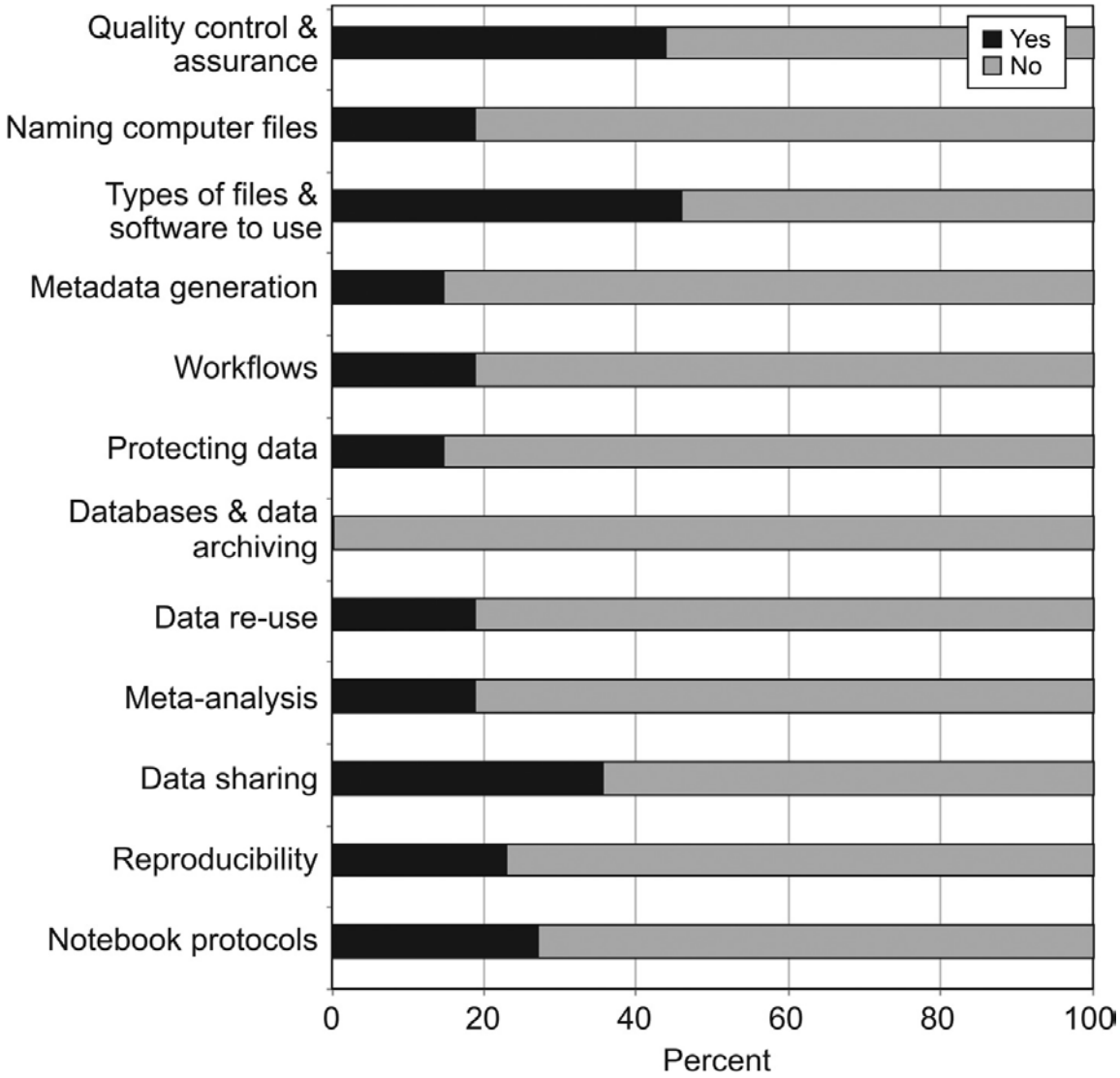






DataONE

The Fractured Lab Notebook

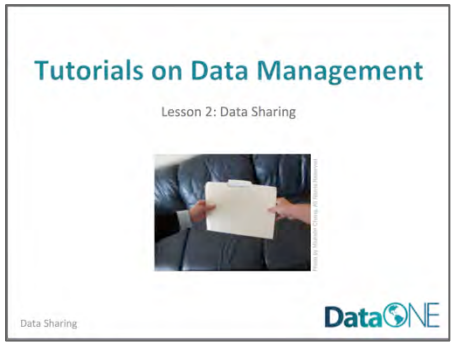


Strasser & Hampton (2012) Ecosphere 3:12 DOI: 10.1890/ES12-00139.1

DataONE

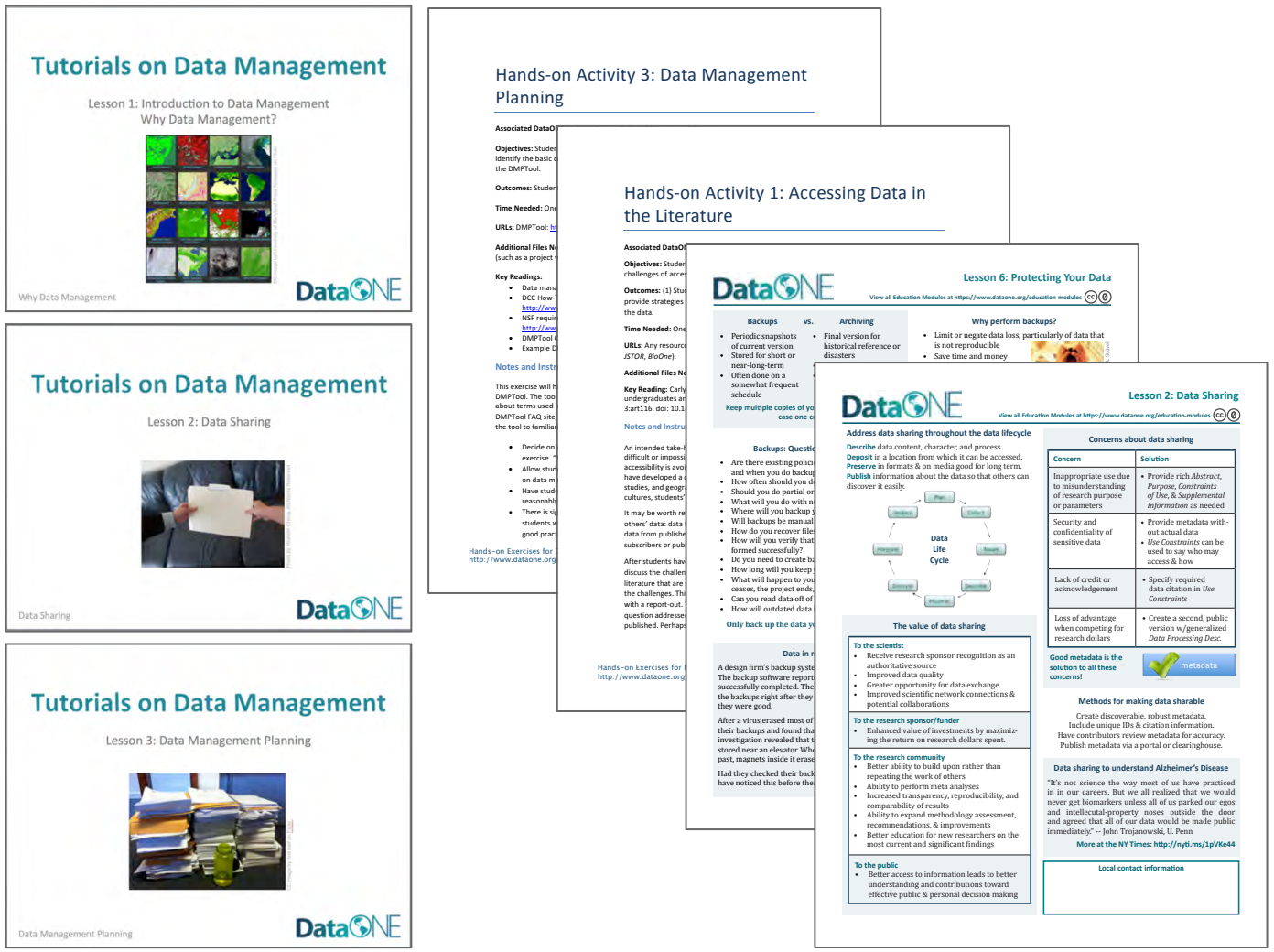
DataONE Education Modules

dataone.org/education-modules



DataONE Education Modules

dataone.org/education-modules



Community Use

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

Community Use

Page	Event Label	Total Events
		14,947 % of Total: 6.42% (232,749)
1. /education-module	1. /sites/all/documents/L01_DataManagement.pptx	**2,115** (14.15%)
2. /best-practices/cre up-policy	2. /sites/all/documents/education-modules/pptx/L01_DataManagement.pptx	**1,364** (9.12%)
3. /data-managemen	3. /sites/all/documents/L07_Metadata.pptx	**1,029** (6.88%)
4. /best-practices	4. /sites/all/documents/DataONE_Education_Modules_Full_Set.pptx	**913** (6.11%)
5. /software-tools/fu	5. /sites/all/documents/L03_DataManagementPlanning.pptx	**793** (5.30%)
6. /current-member-r	6. /sites/all/documents/education-modules/pptx/L03_DataManagementPlanning.pptx	**762** (5.10%)
7. /find-data	7. /sites/all/documents/L02_DataSharing.pptx	**740** (4.95%)
8. /	8. /sites/all/documents/L04_DataEntryManipulation.pptx	**725** (4.85%)
9. /software-tools/db	9. /sites/all/documents/education-modules/pptx/L07_Metadata.pptx	**634** (4.24%)
10. /investigator-toolk	10. /sites/all/documents/L08_WriteQualityMetadata.pptx	**600** (4.01%)

DataONE

Maintenance

Journal of eScience Librarianship

Volume 6 | Issue 2 Article 1

2017-09-08

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

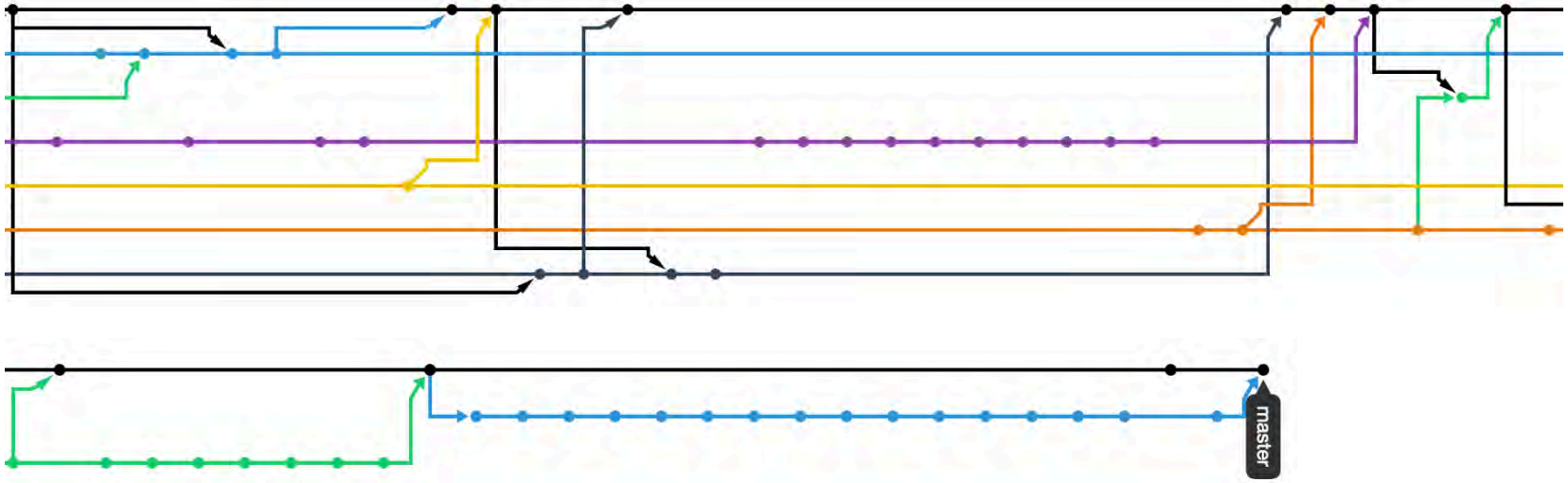
Viv Hutchison
US Geological Survey

Challenges

Solutions

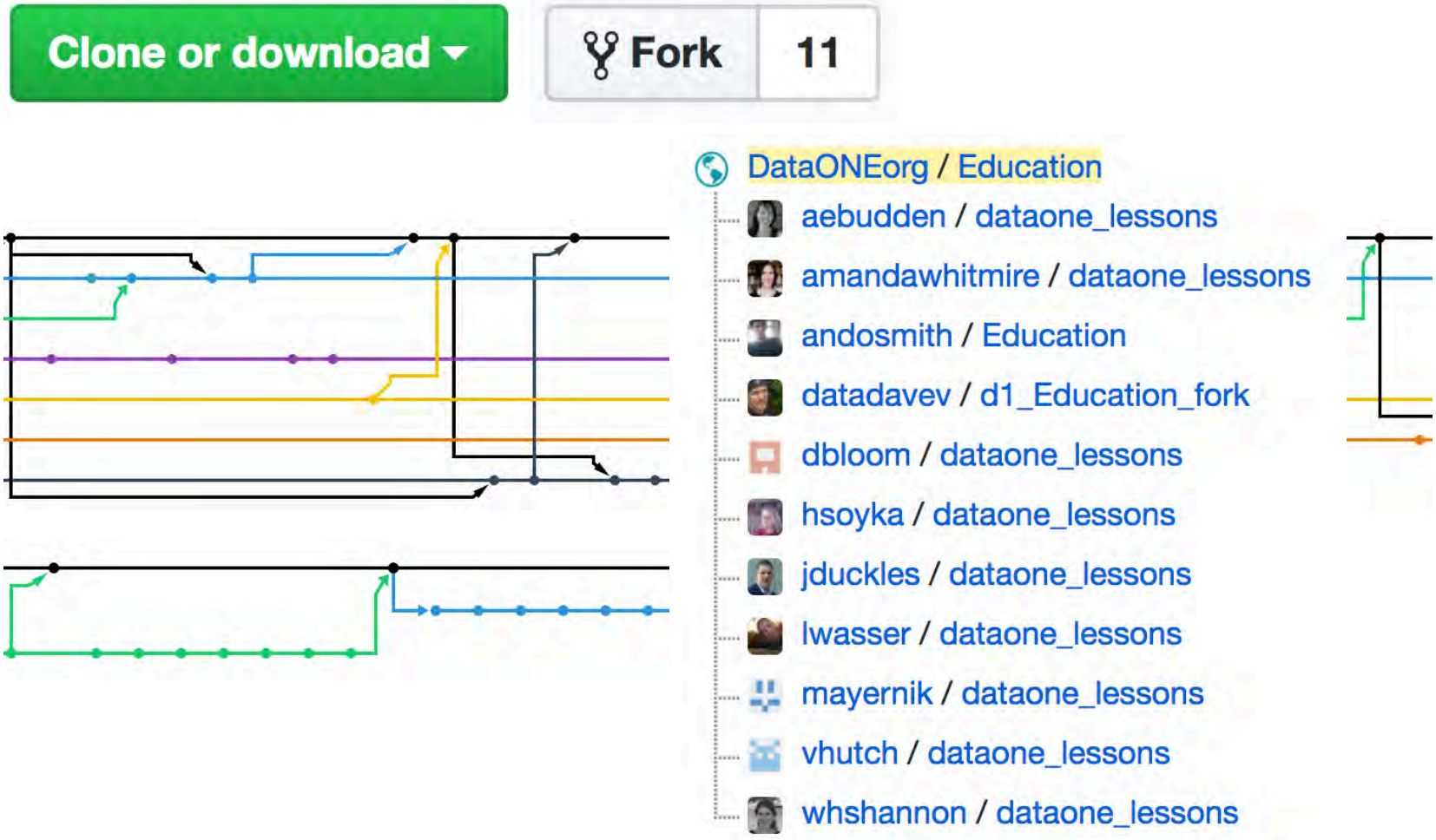
DataONE

Versioned work space

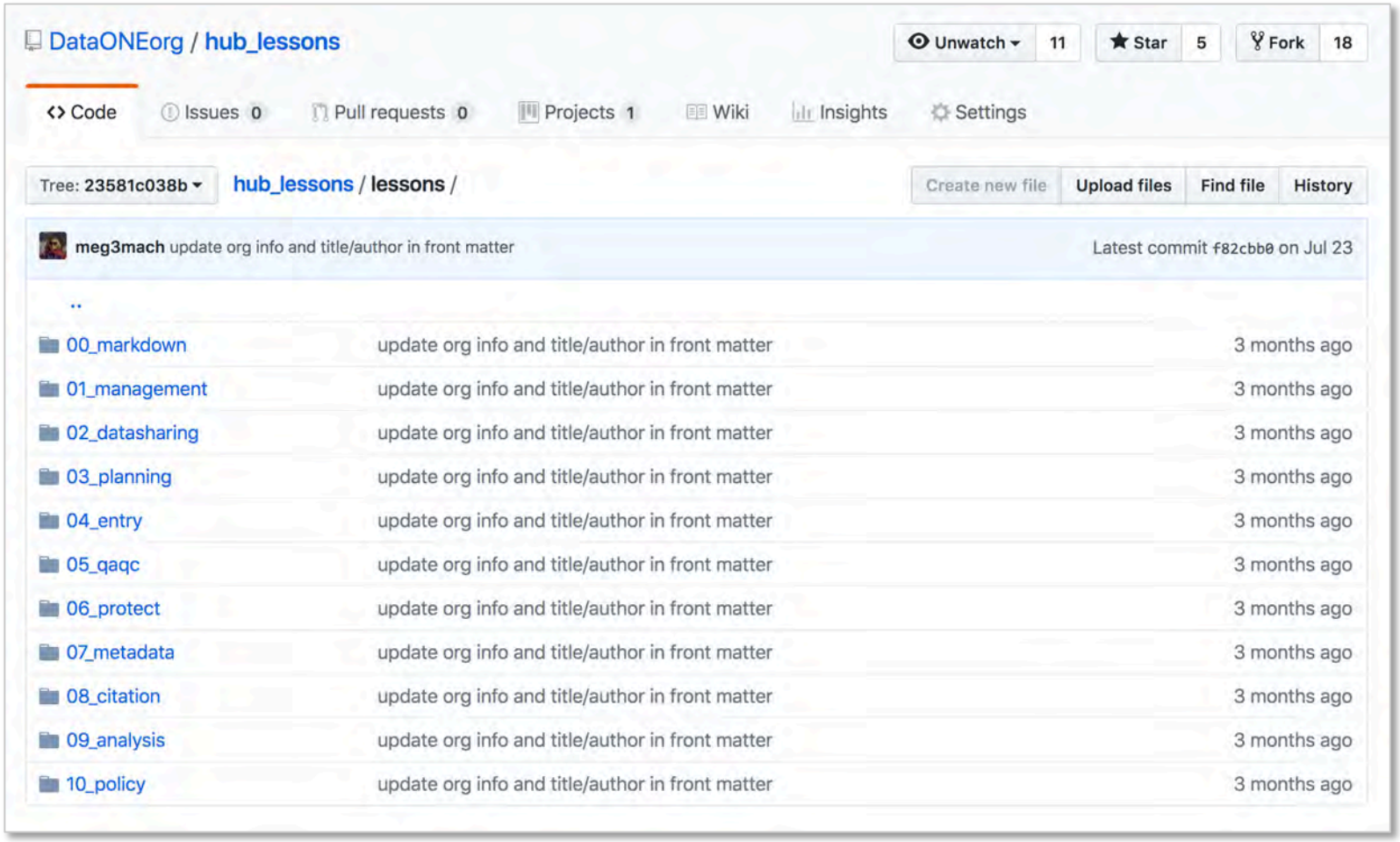


DataONE

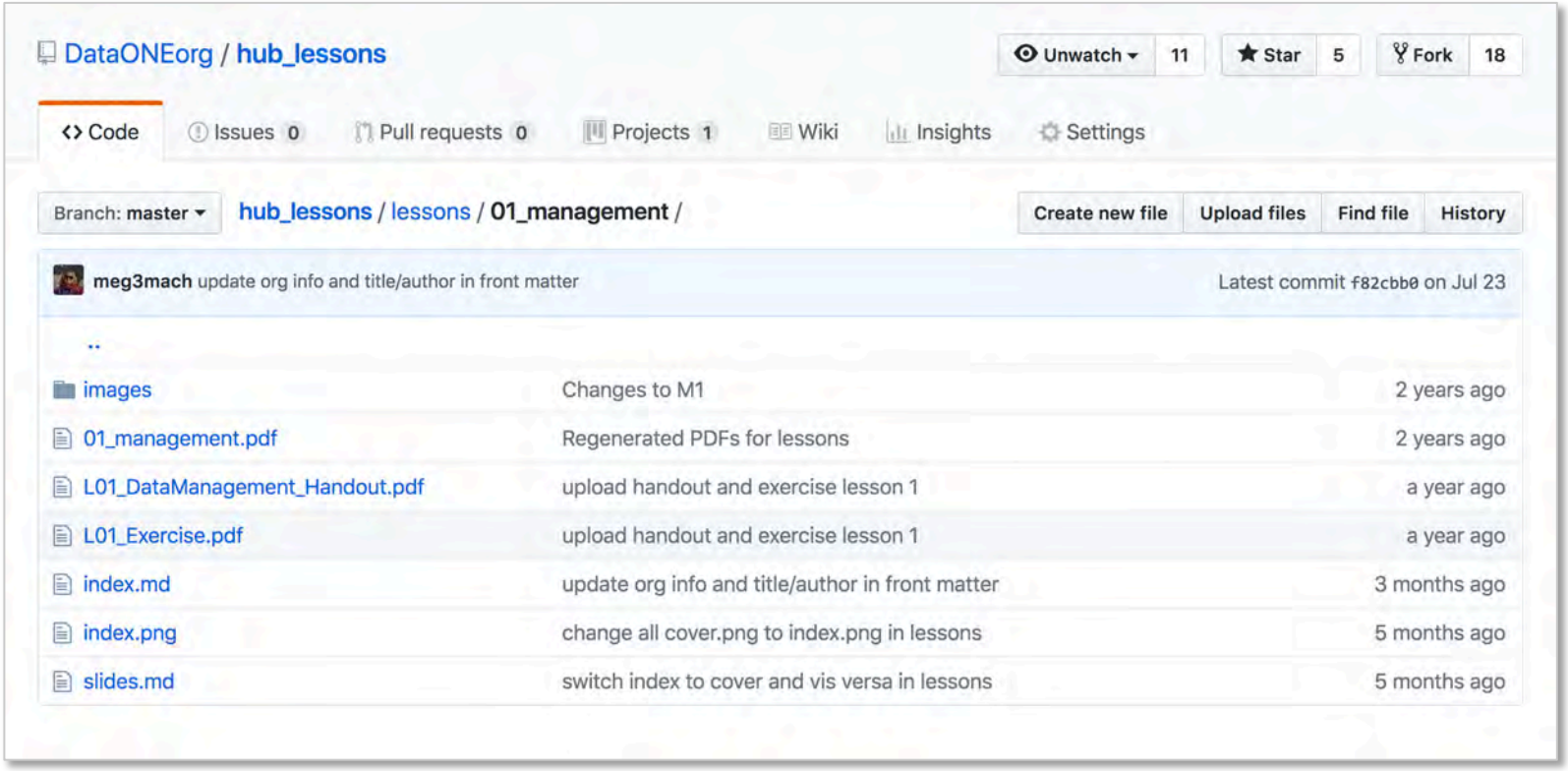
Versioned work space



Structured repository



Structured repository



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

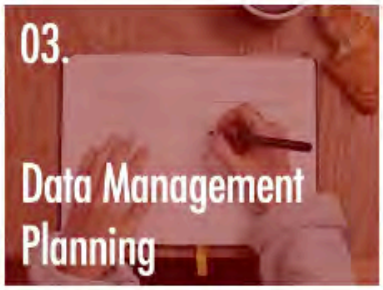
The resources presented on the Data Management Skillbuilding Hub can be updated by users to promote a current, well-maintained, and sustainable educational tool. Learn more about how you can contribute.

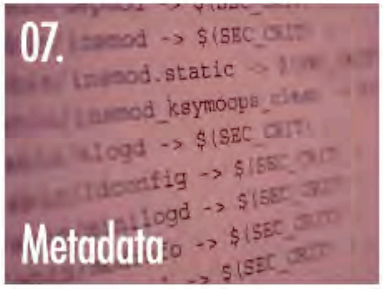
Using This Resource

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





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

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.

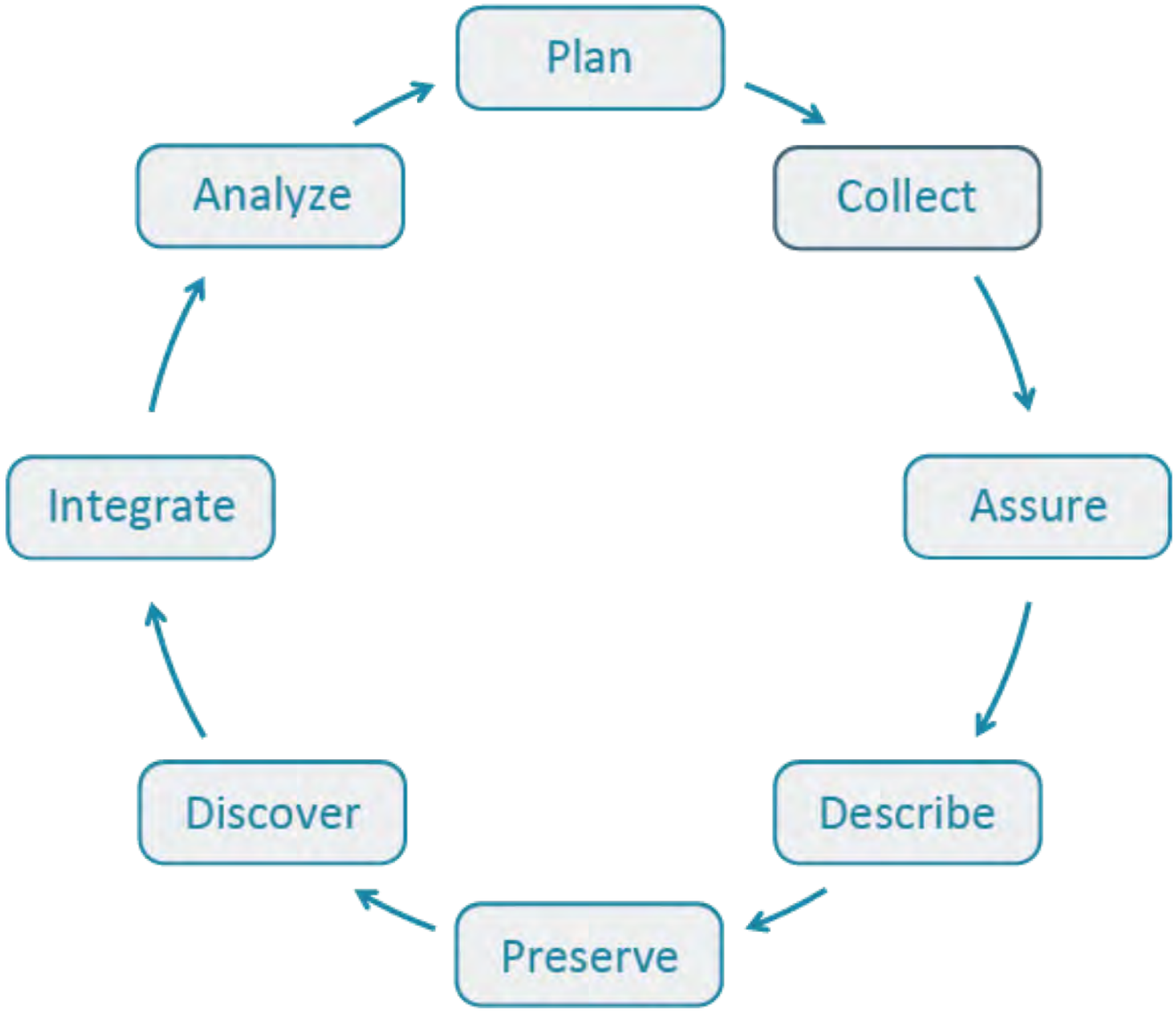
Using This Resource

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

Data life-cycle stages



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 BEST PRACTICE VIDEO

» Filter by stage of the Data Life Cycle

All
Plan
Collect
Assure
✓ Describe
Preserve
Discover
Integrate
Analyze

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.

Data Management Skillbuilding Hub

DataONE

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 on Oct 04, 2018

DataONE

< Home >

Version date: Apr 06, 2017

The Data Lifecycle

Several stages require critical attention to ensure effective data sharing

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

DataONE

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

CC image by Jessica Lucia on Flickr

DataONE

+ -

0:00:33

NOTES FOR CURRENT SLIDE

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.

NOTES FOR NEXT SLIDE

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.

Data Management Skillbuilding Hub

DataONE

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 on Oct 04, 2018

DataONE

< Home >

Version date: Apr 06, 2017

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

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 Practices by Data Life Cycle

All
Plan
Collect
Assure
Describe
Preserve
Discover
Integrate
Analyze

Learn more:
BP 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 da... (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

Best Practices by Data Life Cycle

All
Plan
Collect
Assure
Describe
Preserve
Discover
Integrate
Analyze

Learn more:

BP Primer

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.

Best Practices by Data Life Cycle

All
Plan
Collect
Assure
Describe
Preserve
Discover
Integrate
Analyze

Learn more:

BP Primer

Description Rationale

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

Additional Information

Hook, L.A., Beaty, T.W., Santhana-Vannan, S., Baskaran, L. and Cook, R.B. 2007. Best practices for preparing environmental data sets to share and archive. Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A. (daac.ornl.gov/PI/bestprac.html)

Sheldon, W., Henshaw, D. and Ramsey, K. 2007. Final Report: Workshop to define quality management standards for data completeness in derived data products. Long Term Ecological Research Network Document Archive, University of New Mexico, Albuquerque, NM.

Additional Information (Biblio)

Best Practices for Preparing Ecological and Ground-Based Data Sets to Share and Archive

Data Management Skillbuilding Hub

DataONE

Home Contribute FAQ GitHub

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.

Data Management Skillbuilding Hub

DataONE

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 on Oct 04, 2018

DataONE

< Home >

Version date: Apr 06, 2017

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

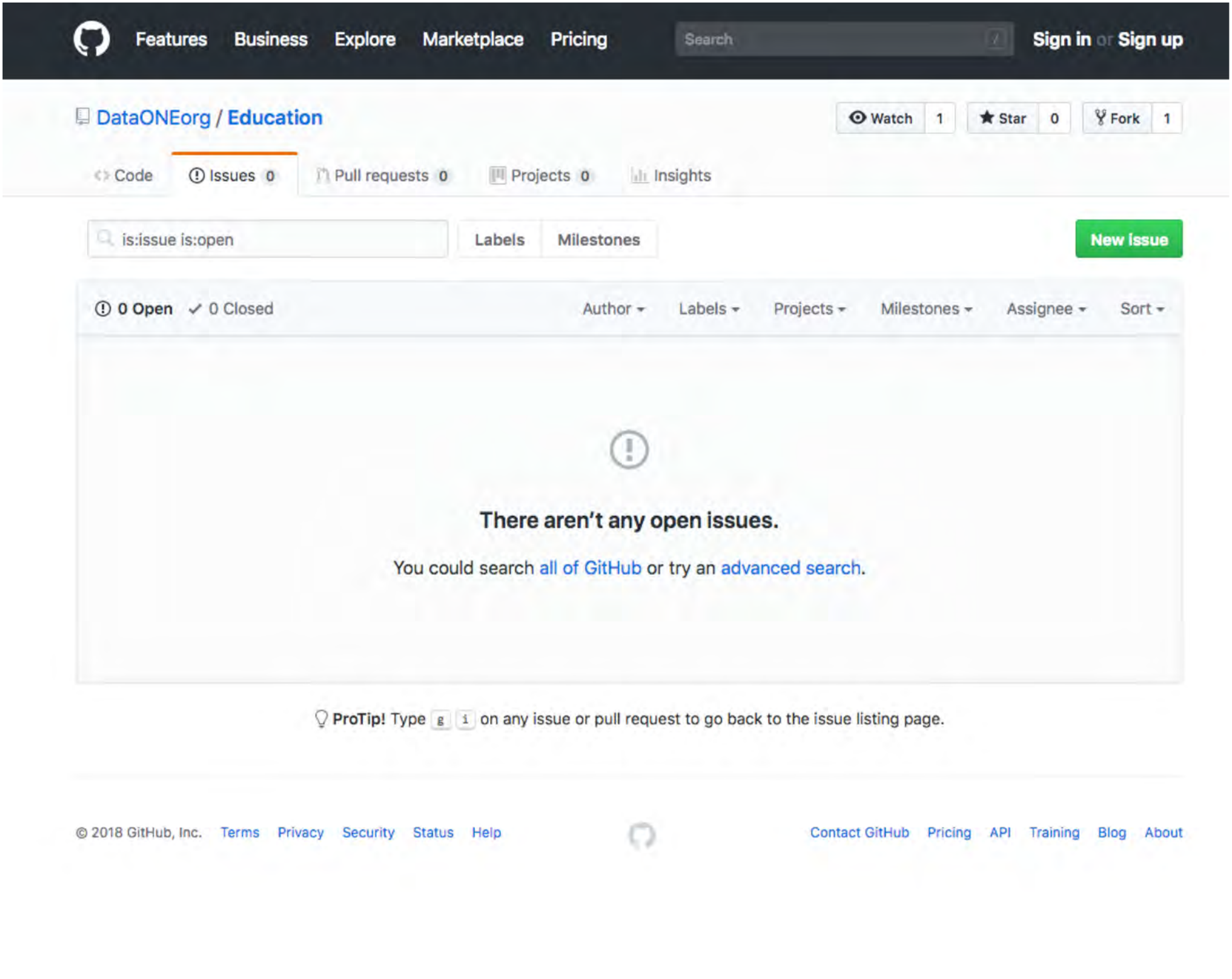
1. Create a fork of the lessons or best practices repository into your github account, depending on which content you wish to edit.
2. Modify the files that you want to change (See "Structure" below for tips on making changes).
3. Submit a pull-request against the `master` branch of this repository.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


Features
Business
Explore
Marketplace
Pricing
Search
Sign in or Sign up
DataONEorg / Education
Watch 1
Star 0
Fork 1
Code
Issues 0
Pull requests 0
Projects 0
Insights
is:issue is:open
Labels
Milestones
New issue
0 Open
0 Closed
Author
Labels
Projects
Milestones
Assignee
Sort
There aren't any open issues.
You could search all of GitHub or try an advanced search.
ProTip! Type g i on any issue or pull request to go back to the issue listing page.
© 2018 GitHub, Inc.
Terms
Privacy
Security
Status
Help
Contact GitHub
Pricing
API
Training
Blog
About

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)

Data Management Skillbuilding Hub

DataONE

Home Contribute FAQ GitHub

Edit or fork content for your own use

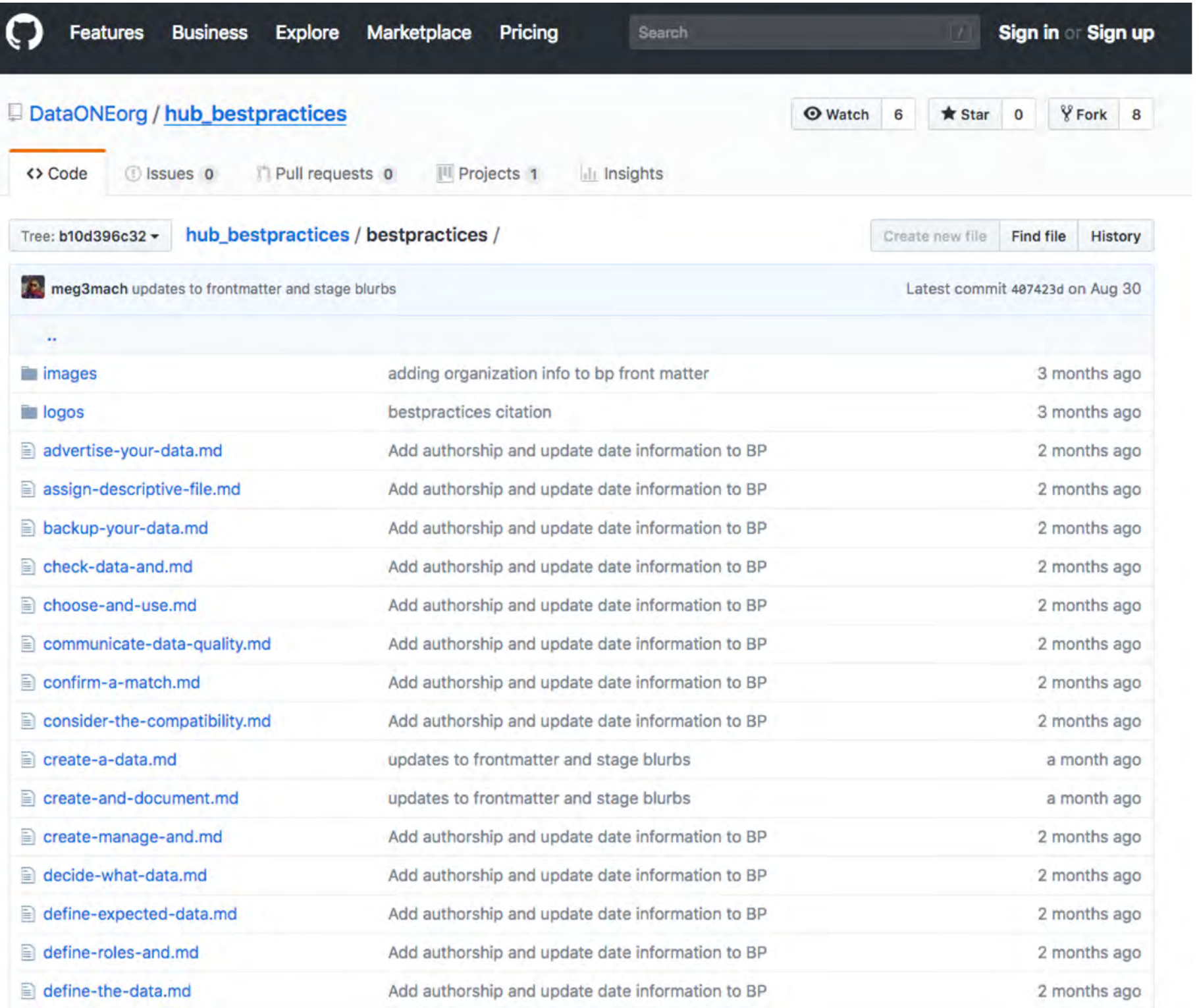
Select one of the buttons below to open the appropriate GitHub repository:

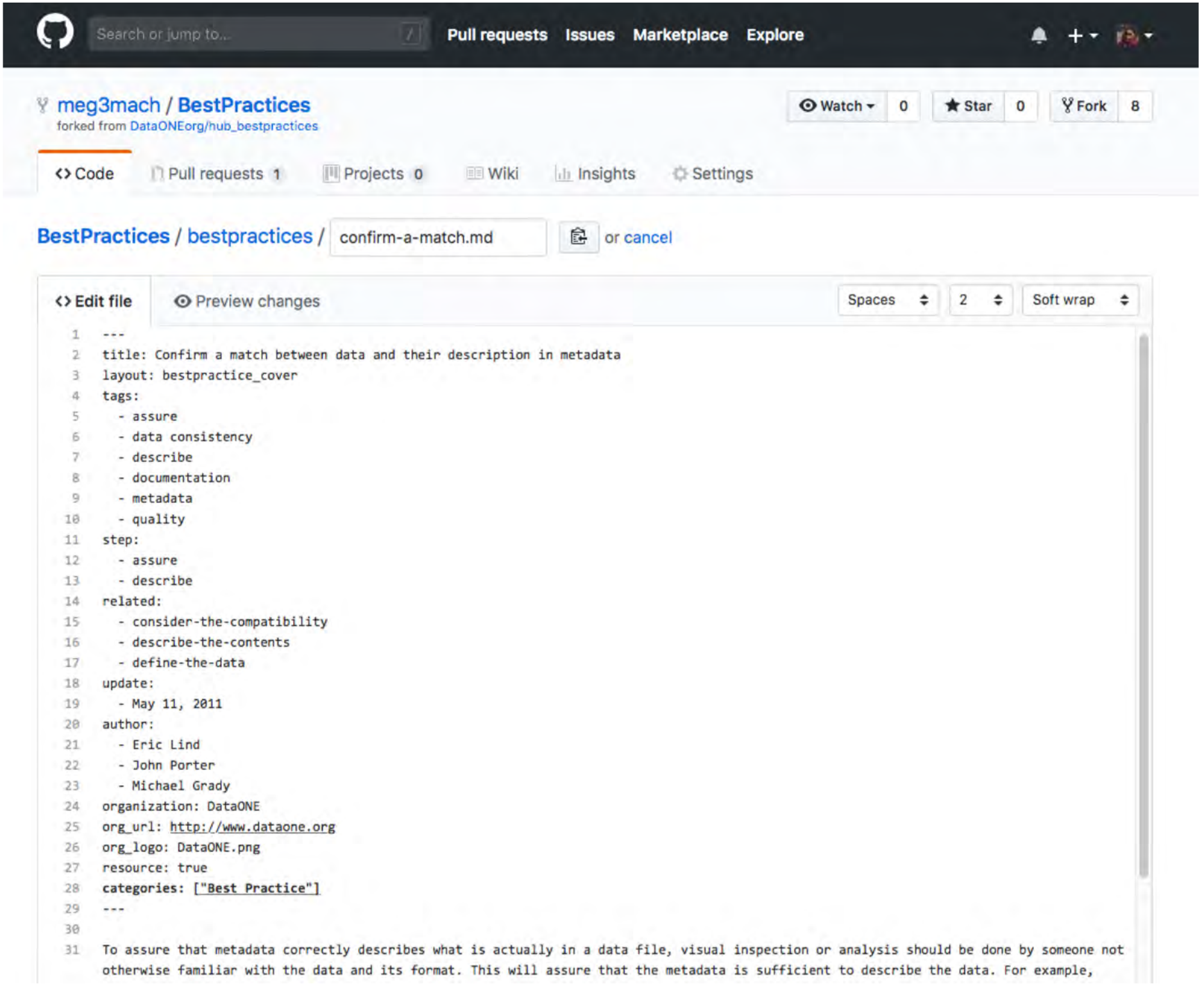
Teaching Module Best Practice

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.


Features Business Explore Marketplace Pricing
Search
Sign in or Sign up
DataONEorg / hub_bestpractices
Watch 6
Star 0
Fork 8
Code
Issues 0
Pull requests 0
Projects 1
Insights
Tree: b10d396c32
hub_bestpractices / bestpractices /
Create new file
Find file
History
meg3mach updates to frontmatter and stage blurbs
Latest commit 407423d on Aug 30
..
images adding organization info to bp front matter 3 months ago
logos bestpractices citation 3 months ago
advertise-your-data.md Add authorship and update date information to BP 2 months ago
assign-descriptive-file.md Add authorship and update date information to BP 2 months ago
backup-your-data.md Add authorship and update date information to BP 2 months ago
check-data-and.md Add authorship and update date information to BP 2 months ago
choose-and-use.md Add authorship and update date information to BP 2 months ago
communicate-data-quality.md Add authorship and update date information to BP 2 months ago
confirm-a-match.md Add authorship and update date information to BP 2 months ago
consider-the-compatibility.md Add authorship and update date information to BP 2 months ago
create-a-data.md updates to frontmatter and stage blurbs a month ago
create-and-document.md updates to frontmatter and stage blurbs a month ago
create-manage-and.md Add authorship and update date information to BP 2 months ago
decide-what-data.md Add authorship and update date information to BP 2 months ago
define-expected-data.md Add authorship and update date information to BP 2 months ago
define-roles-and.md Add authorship and update date information to BP 2 months ago
define-the-data.md Add authorship and update date information to BP 2 months ago


Search or jump to...
Pull requests
Issues
Marketplace
Explore
meg3mach / BestPractices
forked from DataONEorg/hub_bestpractices
Watch 0
Star 0
Fork 8
Code
Pull requests 1
Projects 0
Wiki
Insights
Settings
BestPractices / bestpractices / confirm-a-match.md
or cancel
Edit file
Preview changes
Spaces
2
Soft wrap
1 ---
2 title: Confirm a match between data and their description in metadata
3 layout: bestpractice_cover
4 tags:
5 - assure
6 - data consistency
7 - describe
8 - documentation
9 - metadata
10 - quality
11 step:
12 - assure
13 - describe
14 related:
15 - consider-the-compatibility
16 - describe-the-contents
17 - define-the-data
18 update:
19 - May 11, 2011
20 author:
21 - Eric Lind
22 - John Porter
23 - Michael Grady
24 organization: DataONE
25 org_url: http://www.dataone.org
26 org_logo: DataONE.png
27 resource: true
28 categories: ["Best Practice"]
29 ---
30
31 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**

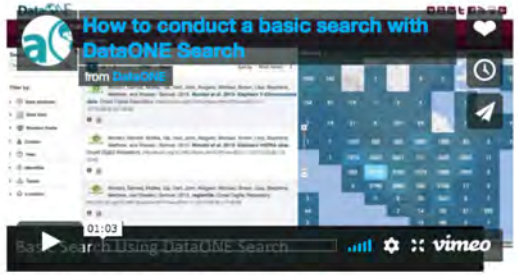
Data Management Skillbuilding Hub

- 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:

Conducting a Basic Search



DataONE Search: Basic Search from DataONE on Vimeo.

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 a 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

Read More

DataONE Life Cycle - https://www.dataone.org/data-life-cycle

Search

Find learning resources by keyword, name, date, license and cost

DataONE Search

Browse

See a list of learning resources by educational framework

Browse

Submit

Submit your learning resources to the Clearinghouse

Submit

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

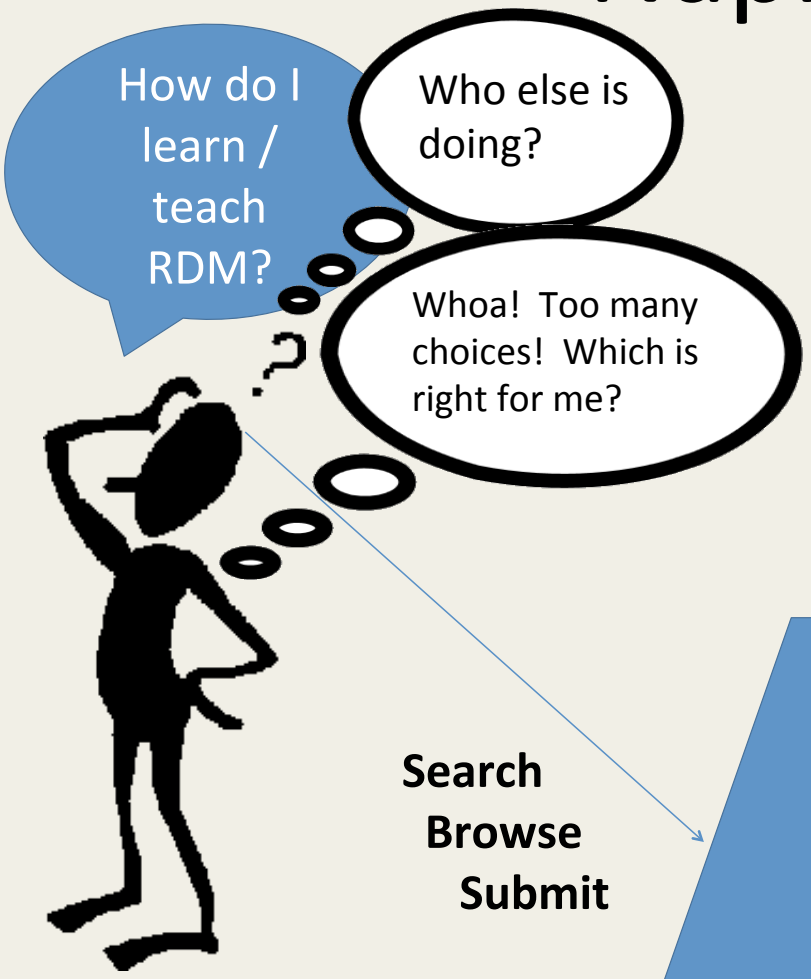
Knowledge Motifs LLC
Mapping sensible data relationships

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…?

but where to find training resources?

Napkin Drawing



Lab / Classroom / 1 on 1

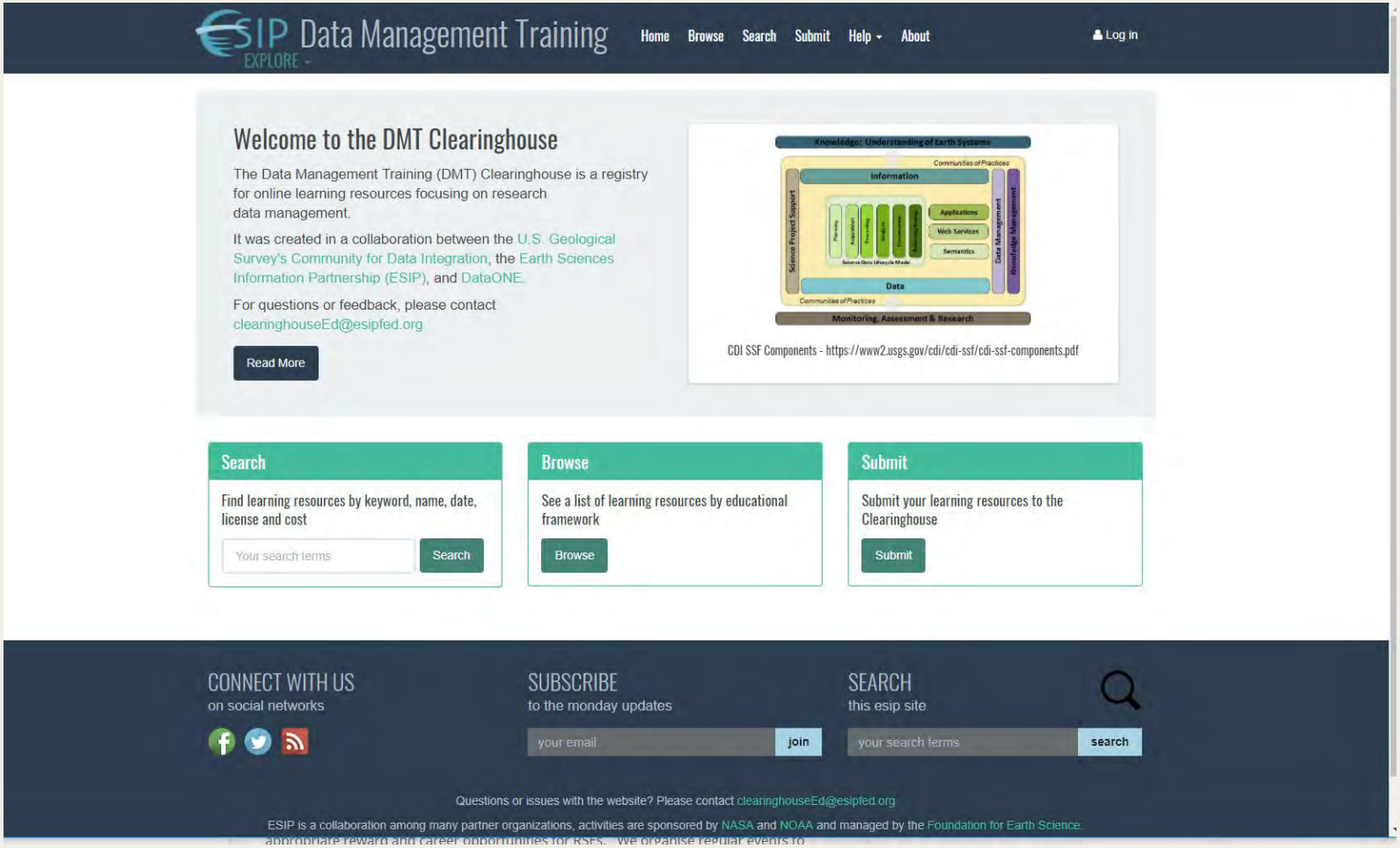
Search
Browse
Submit

Web Portal / Metadata Content Repository/ Clearinghouse

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??

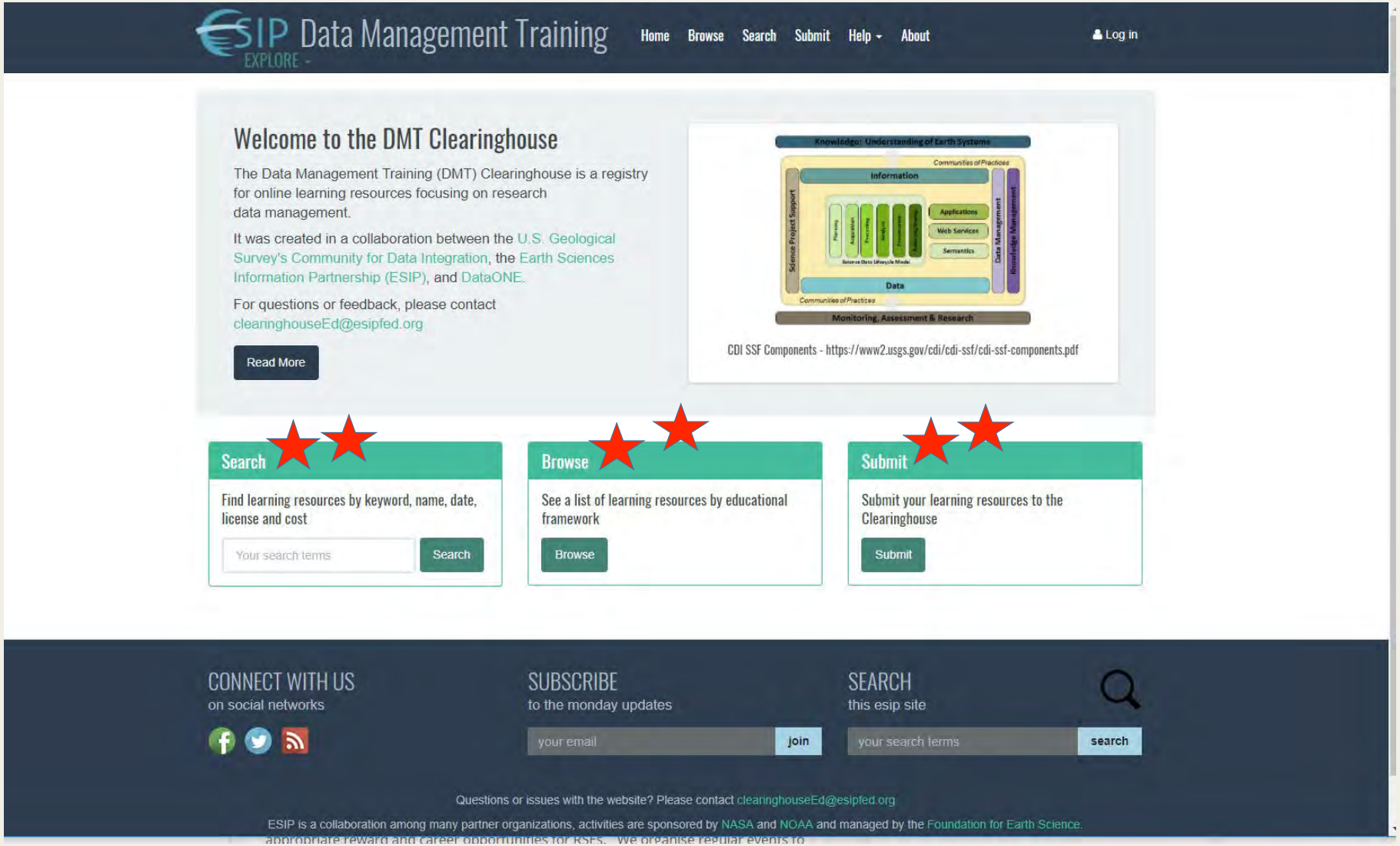
Currently, hosted & maintained

- 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

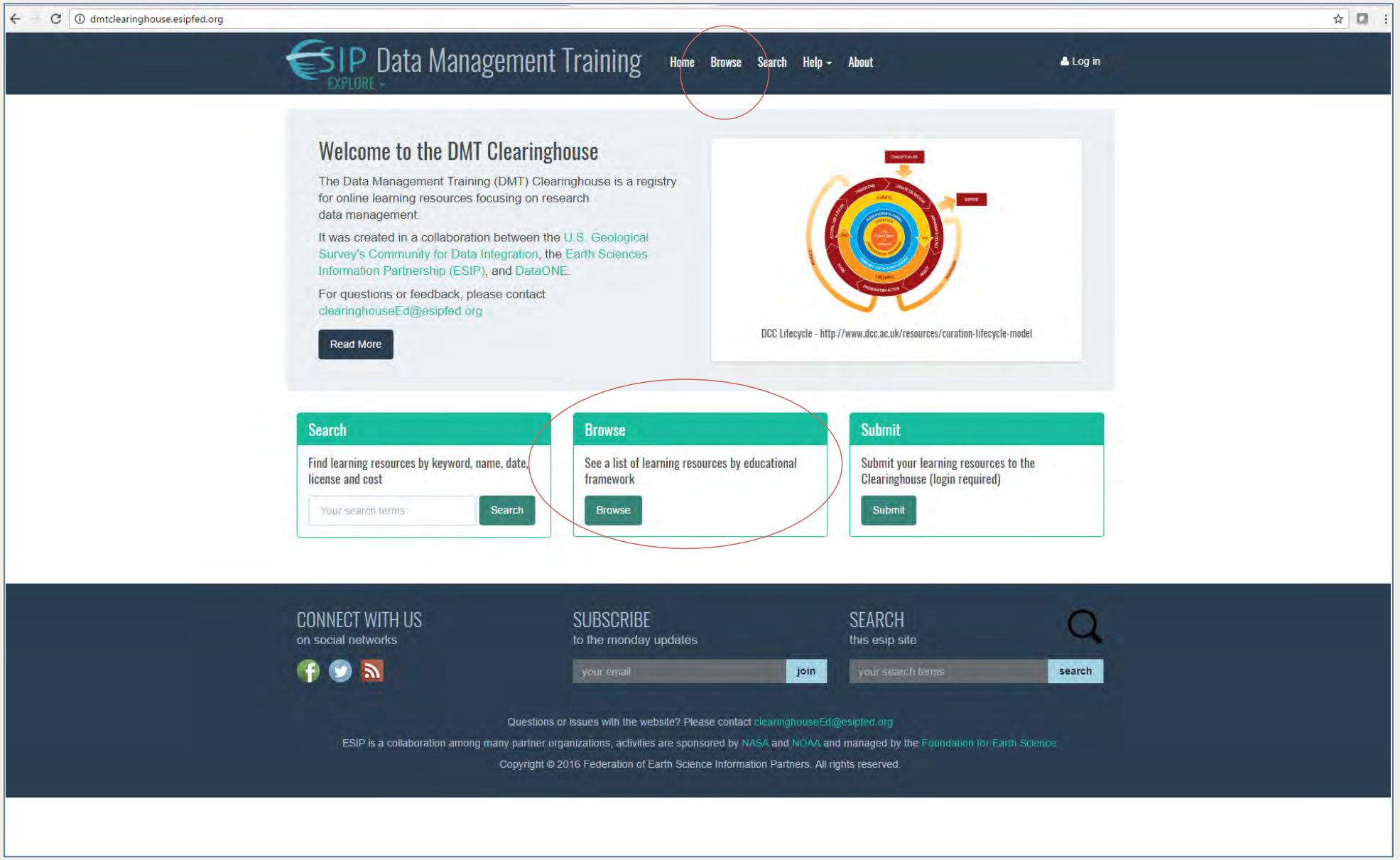
- Crowd-sourced submissions
- Domain-knowledgeable reviewers & editors to maintain quality & currency of resources
- Always seeking user interface & functionality feedback
- After initial seed $$, have been funded by IMLS for 3 year National Leadership Grant
- Exploring options for longer term Sustainability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

Let's take a look!

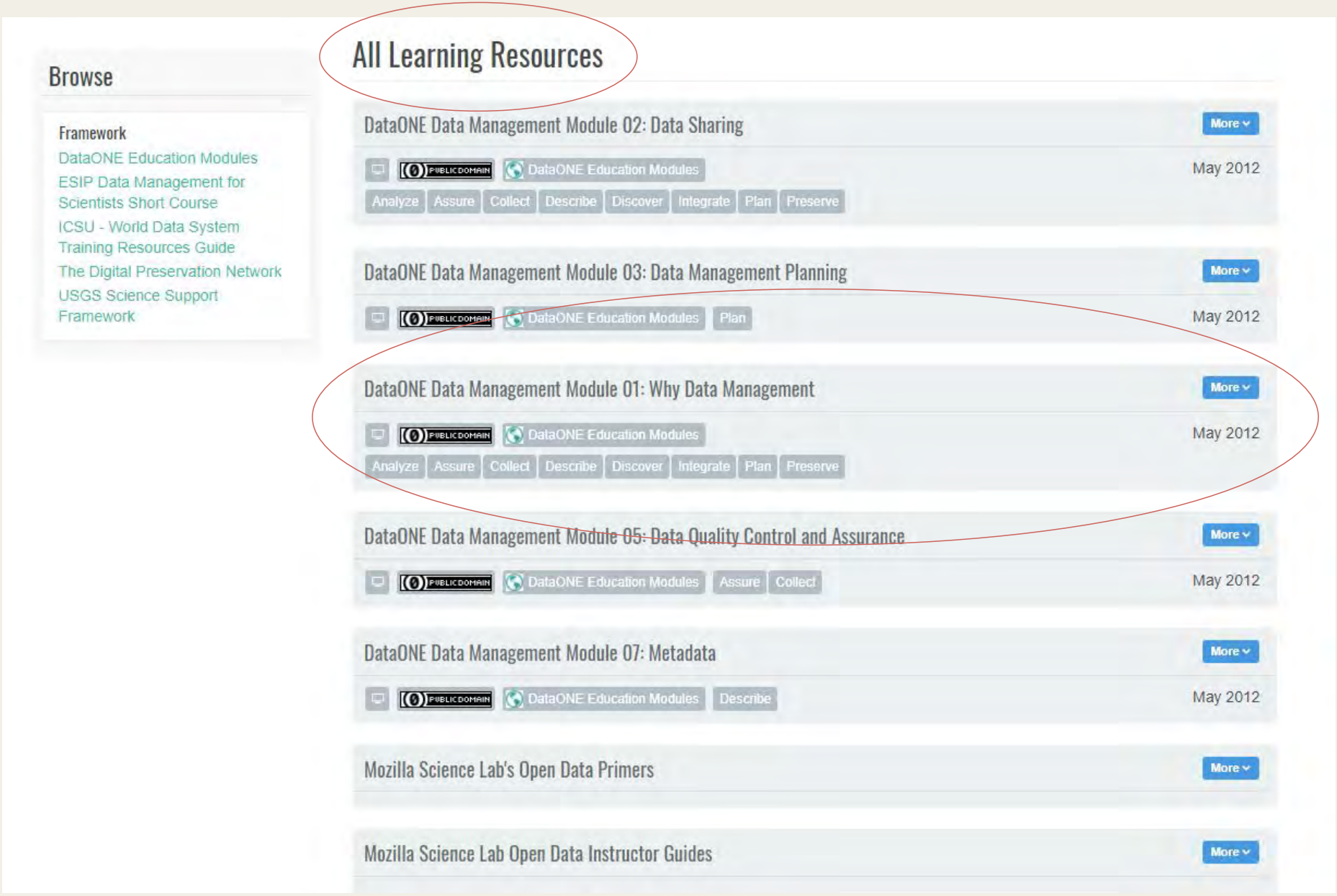


http://dmtclearinghouse.esipfed.org/

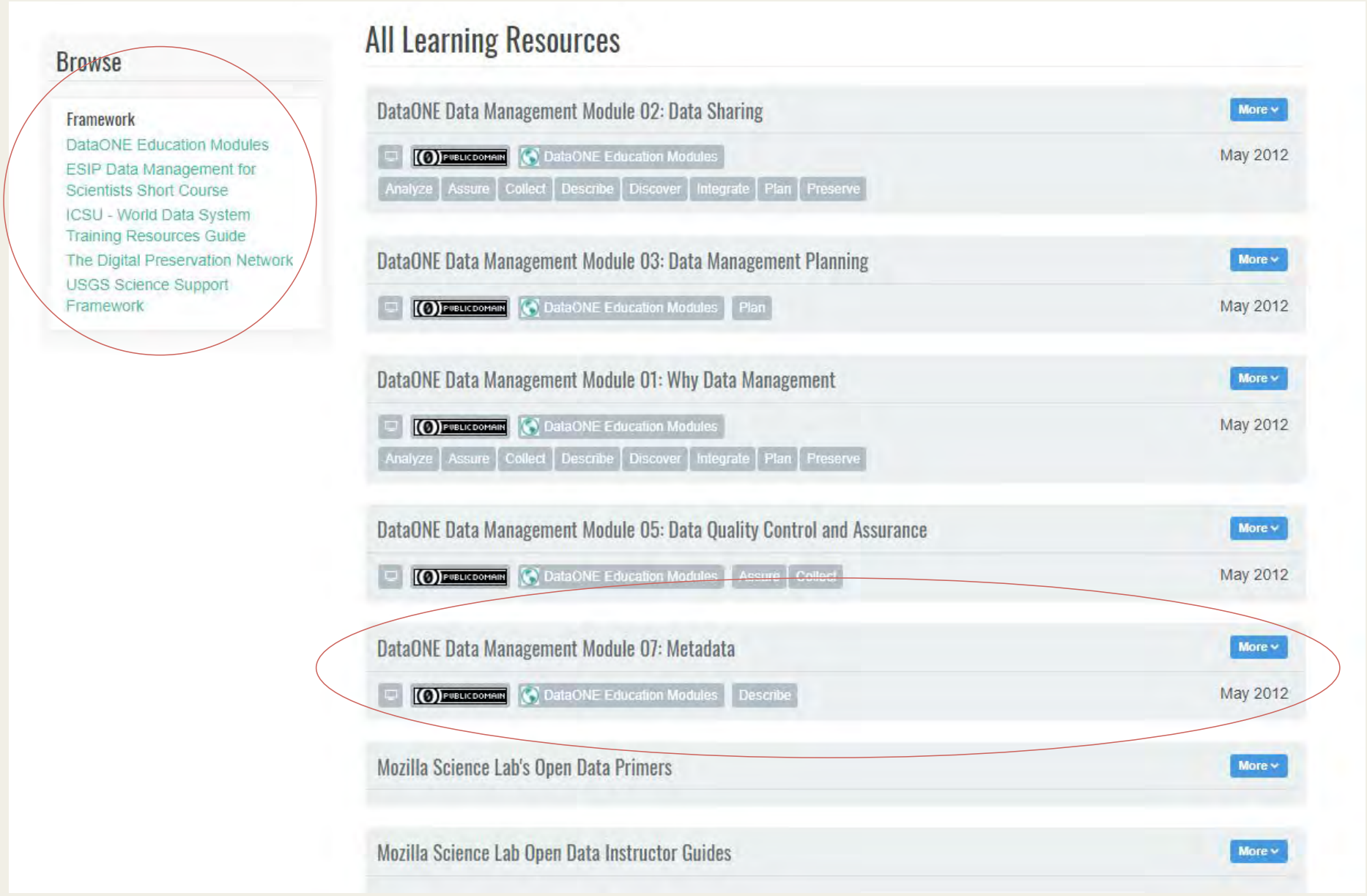
Browse Function



Browse Function



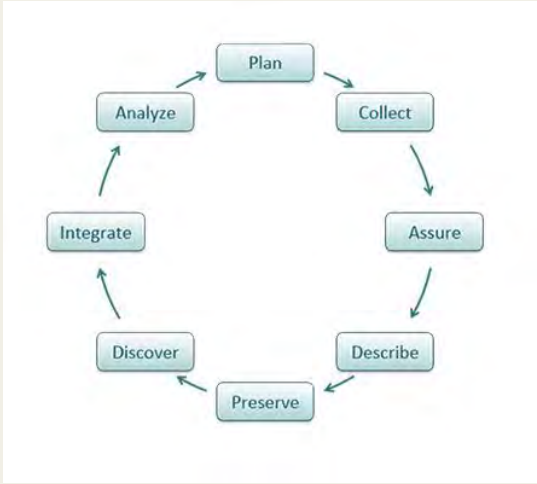
Browse Function



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".*

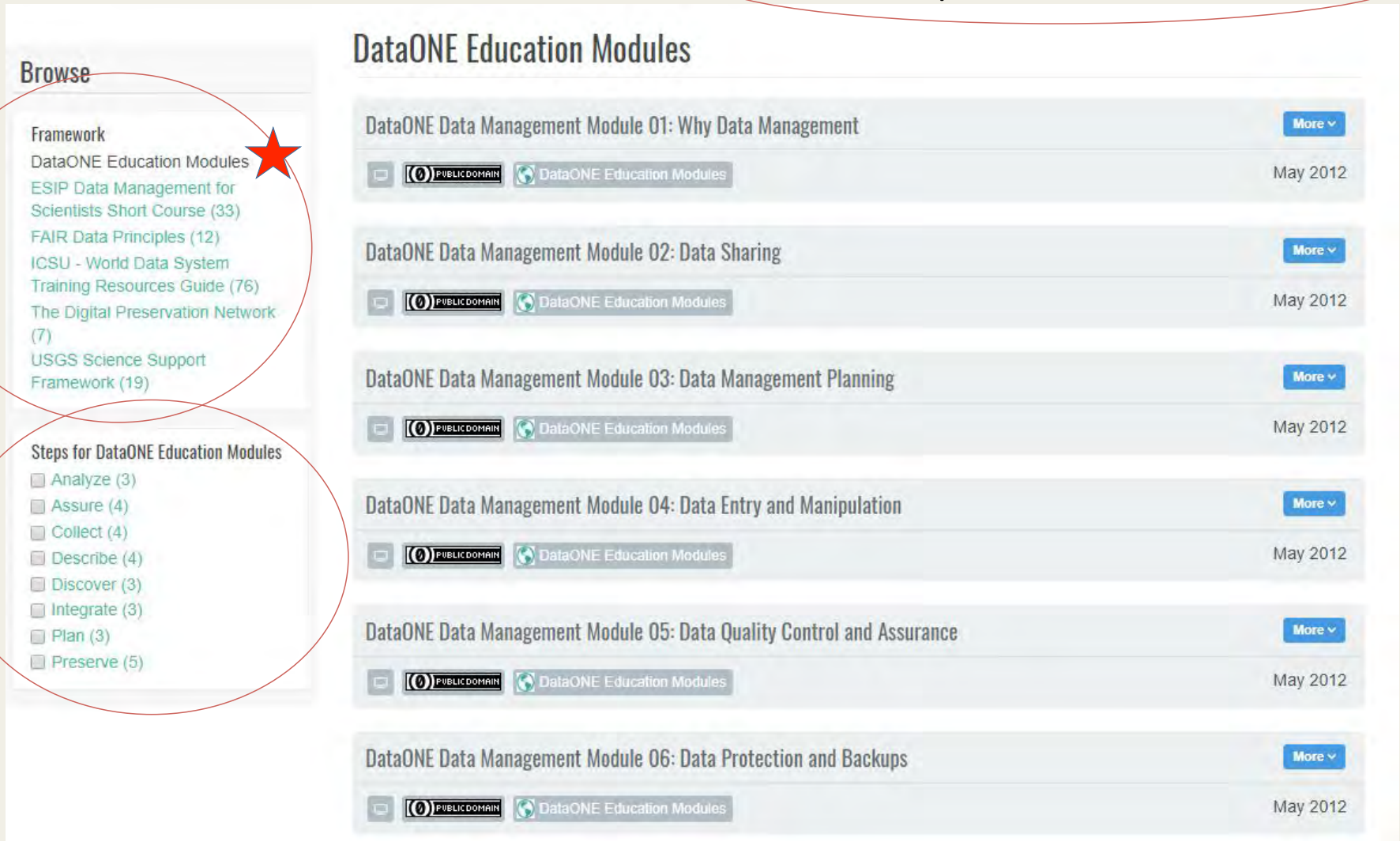


https://www.dataone.org/data-life-cycle

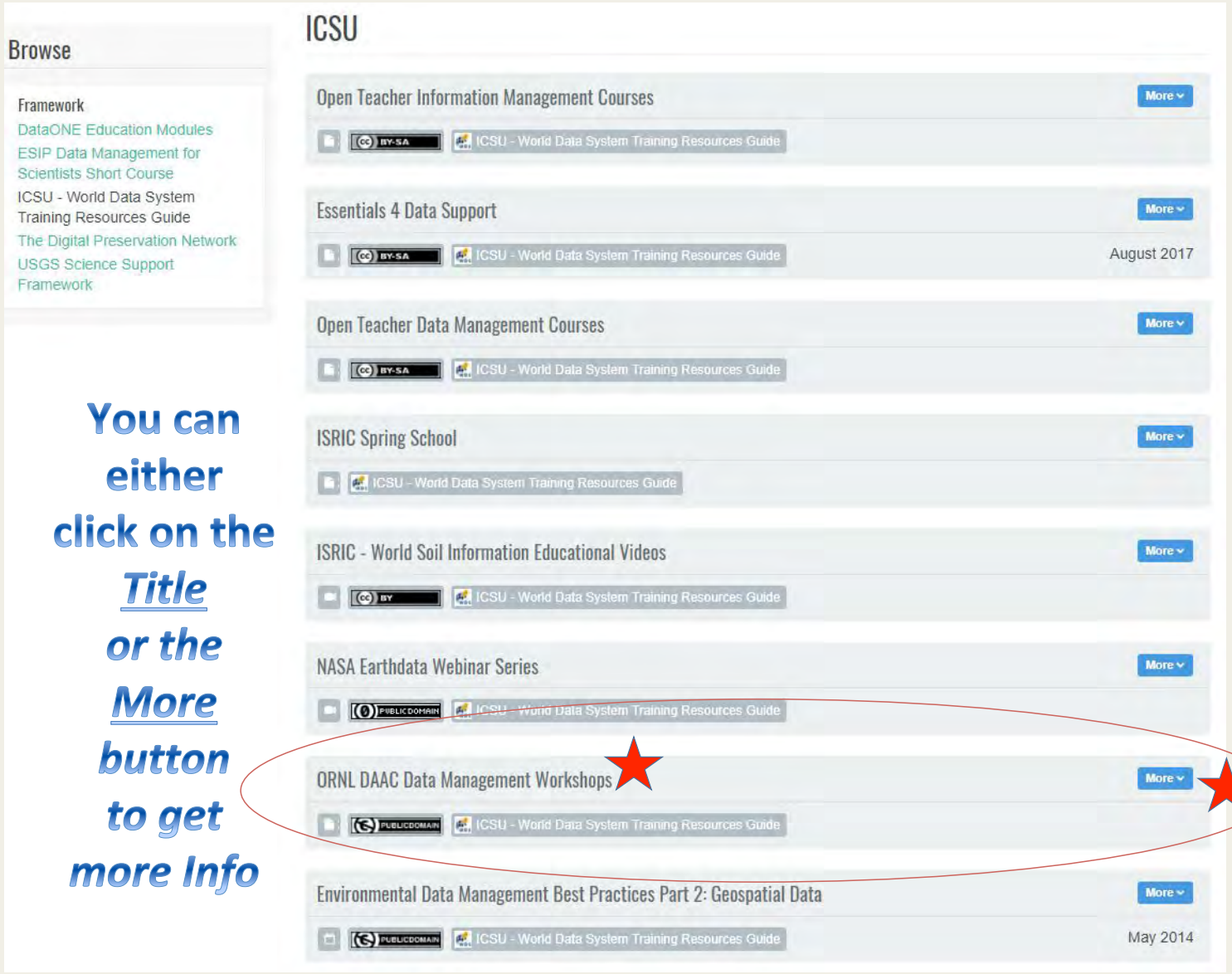
The DataONE data life cycle was developed …in collaboration with the broader DataONE community … [and] serves as an underlying framework for the development of tools, services and education materials by DataONE.

Browse Function

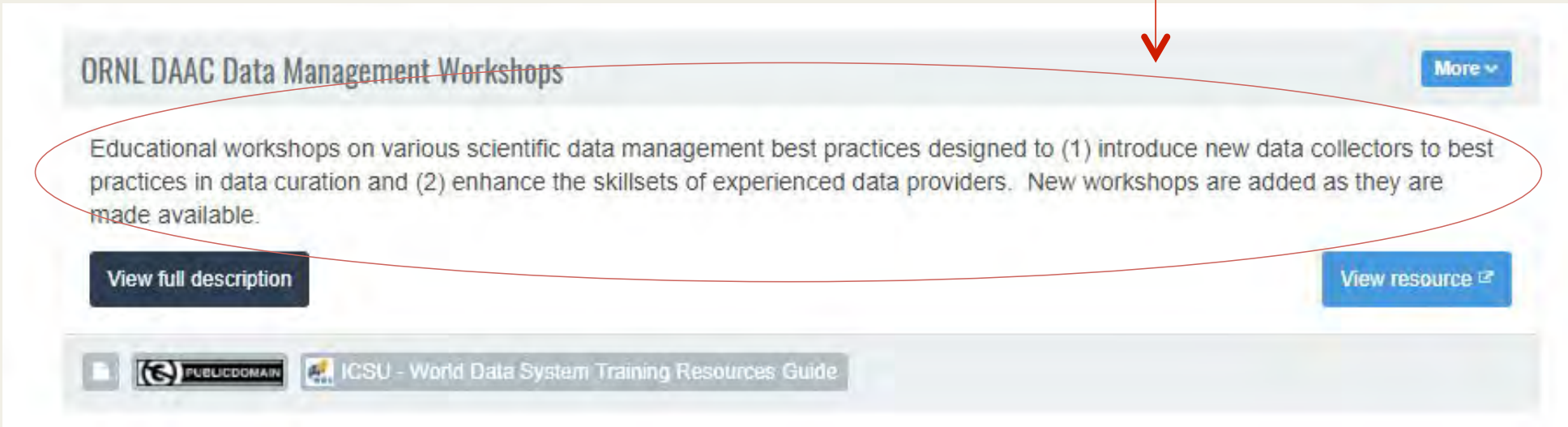
From ~250 'published' resources ➔ ~10



Browse Function -- you found one that looks useful! What next?



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.

Search Function

dmtclearinghouse.esipfed.org

ESIP Data Management Training EXPLORE

Home Browse Search Help About Log in

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 a 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

Read More

DCC Lifecycle - http://www.dcc.ac.uk/resources/curation-lifecycle-model

Search

Find learning resources by keyword, name, date, license and cost

Your search terms Search

Browse

See a list of learning resources by educational framework

Browse

Submit

Submit your learning resources to the Clearinghouse (login required)

Submit

CONNECT WITH US
on social networks

SUBSCRIBE
to the monday updates

your email join

SEARCH
this esip site

your search terms search

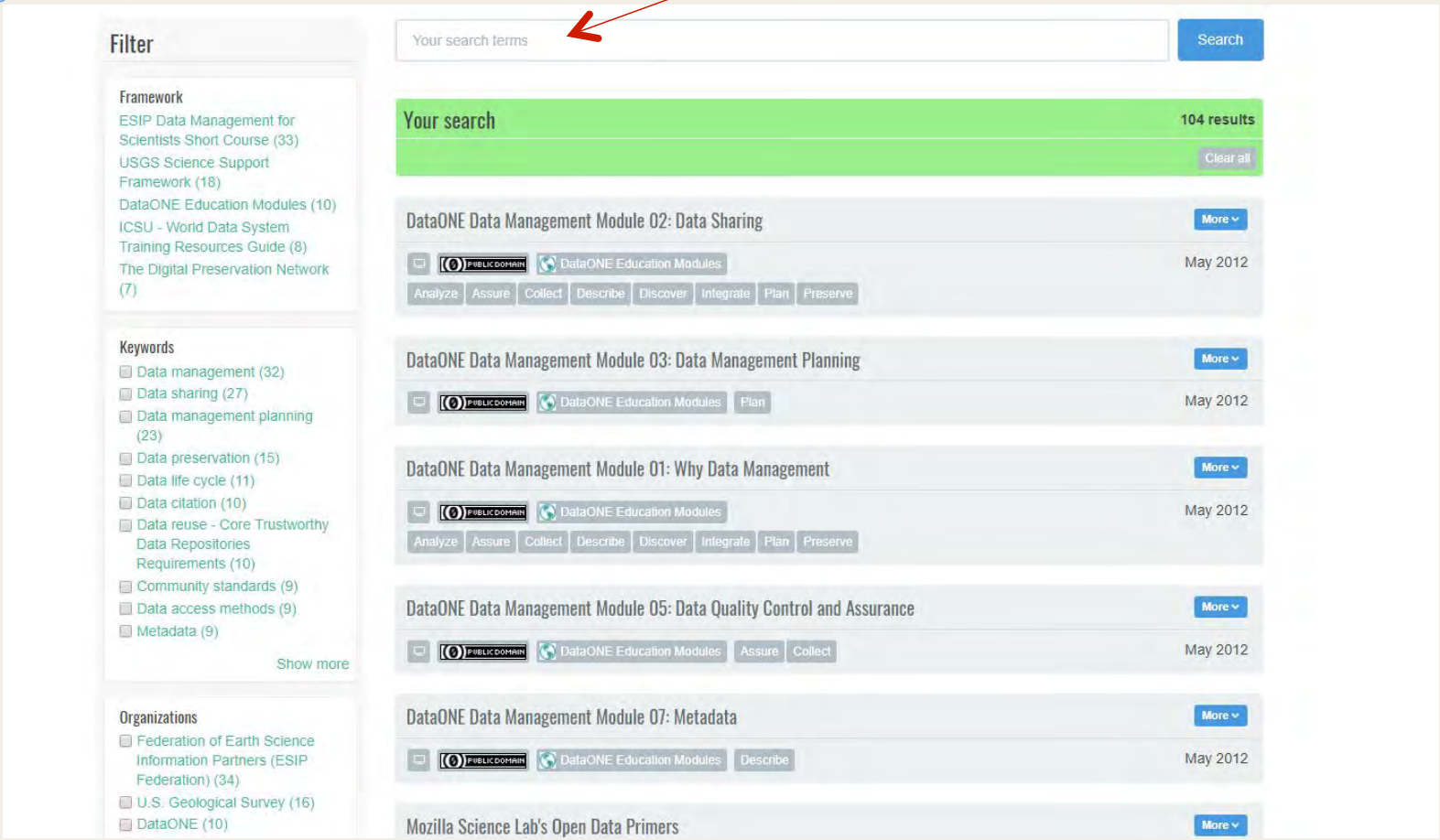
Questions or issues with the website? Please contact clearinghouseEd@esipfed.org

ESIP is a collaboration among many partner organizations, activities are sponsored by NASA and NOAA and managed by the Foundation for Earth Science.

Search Function: 3 approaches

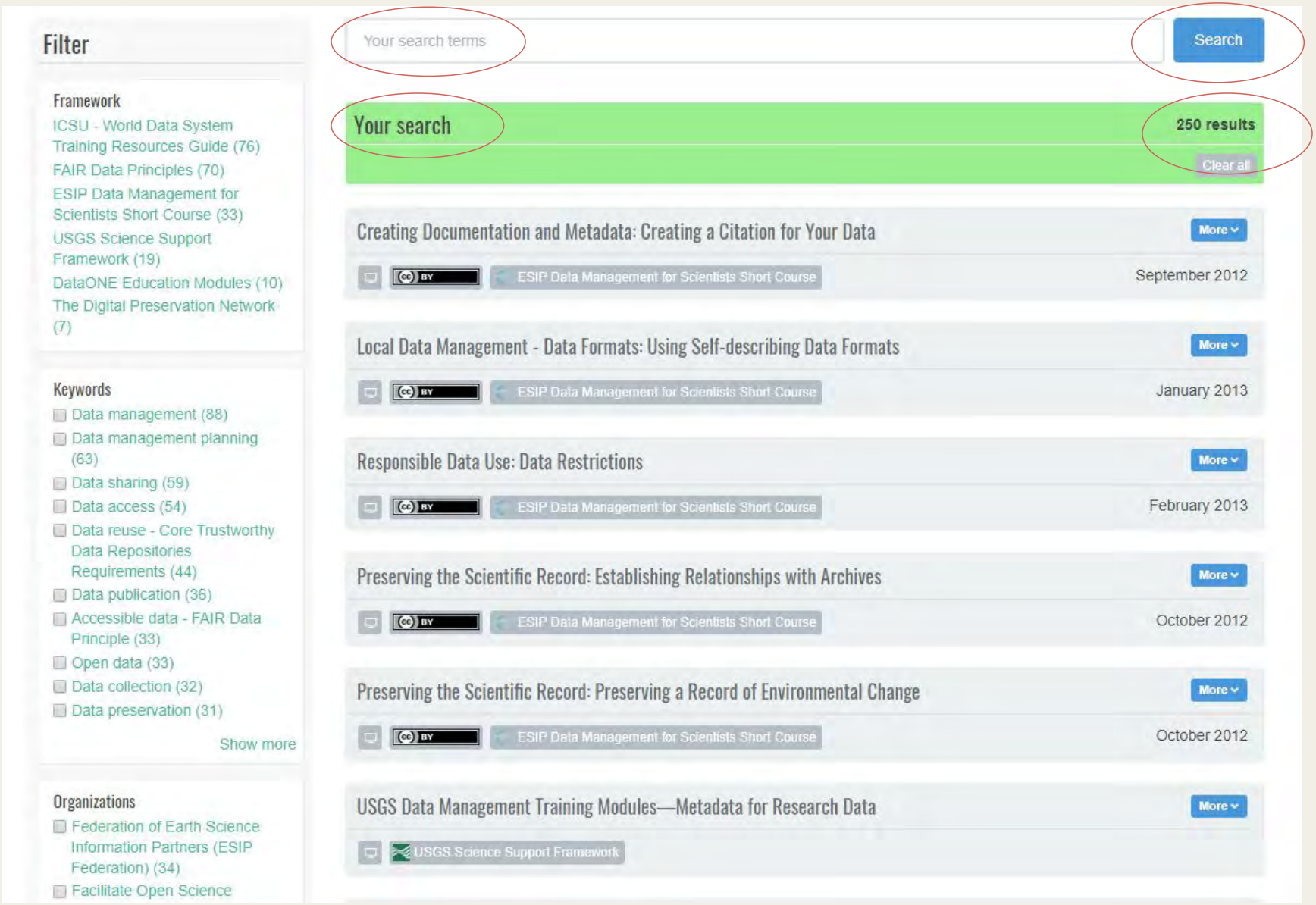
You filter by pre-set categories ("filters")

You enter your own search terms

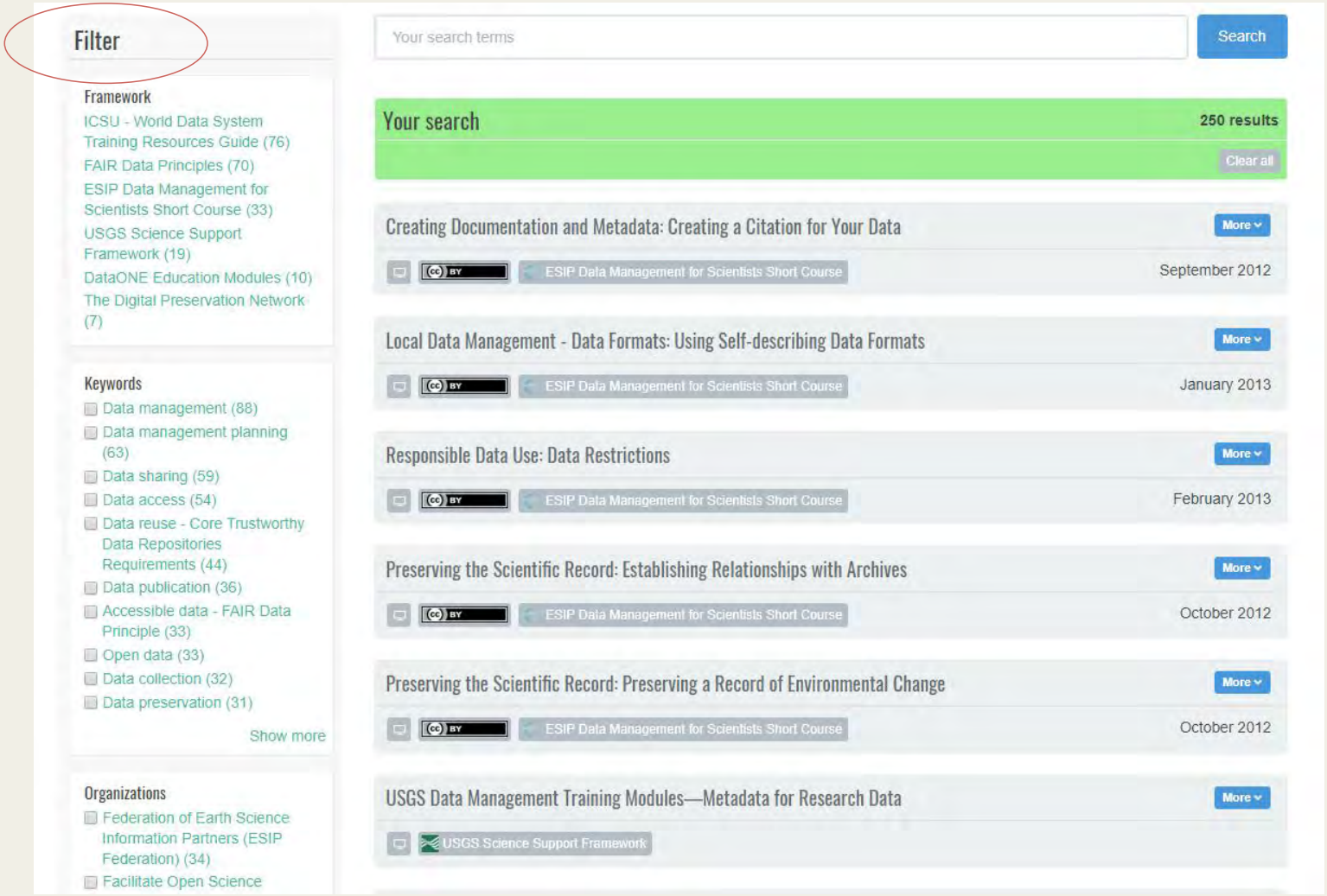


Search Function: Approach 3: Combining Approach 1 & Approach 2 …

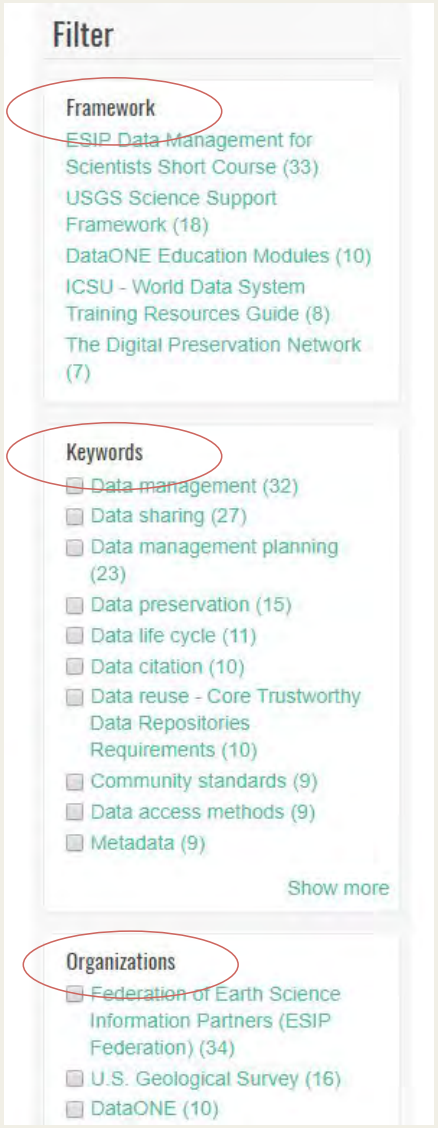
Search Function: Approach 1 = You search using your own terms

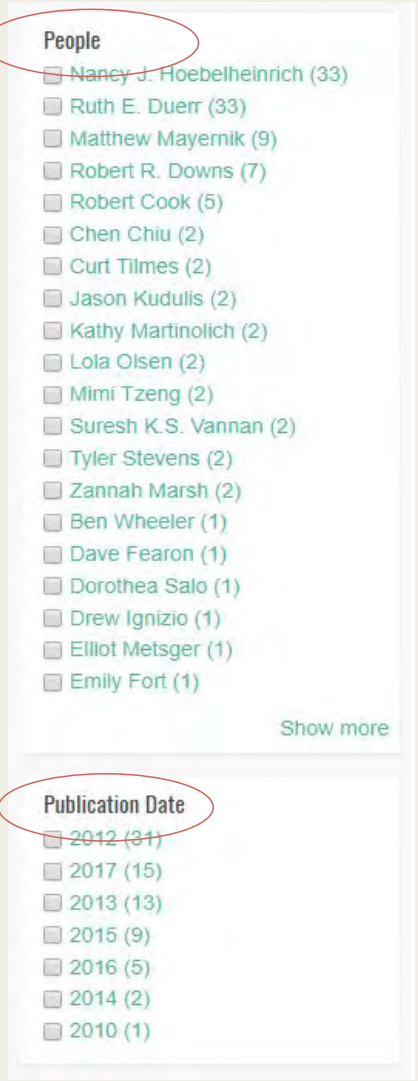


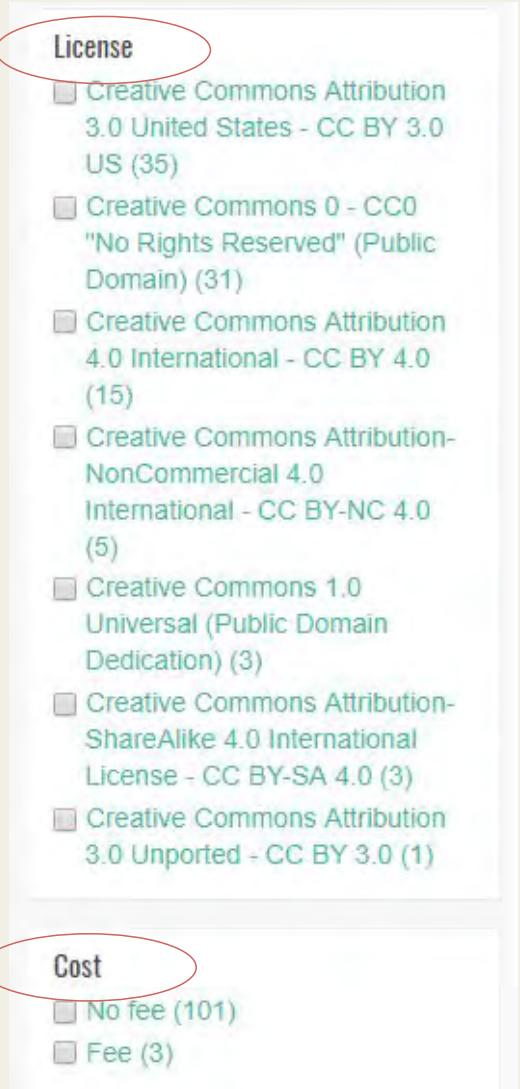
Search Function: Approach 2 = You start by checking boxes within the pre-set categories ("filters")



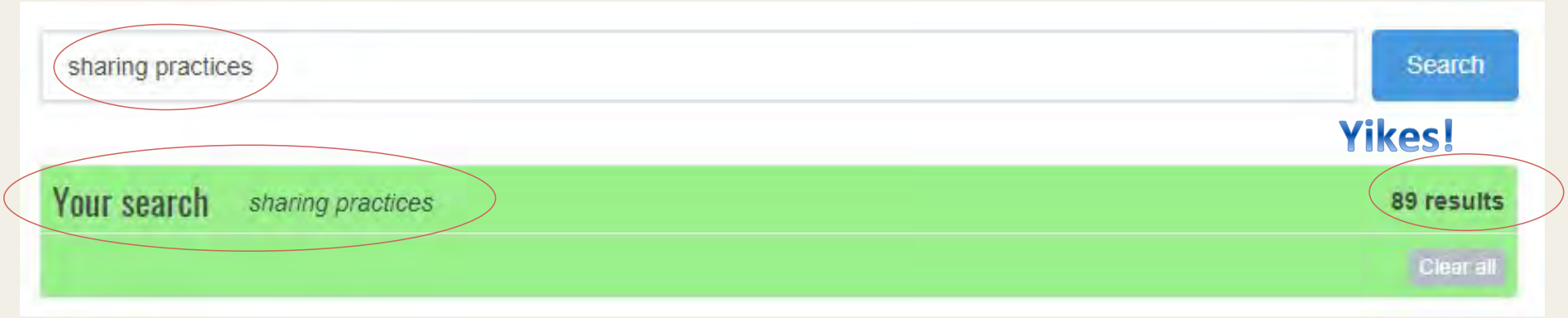
Built-in Search Filters



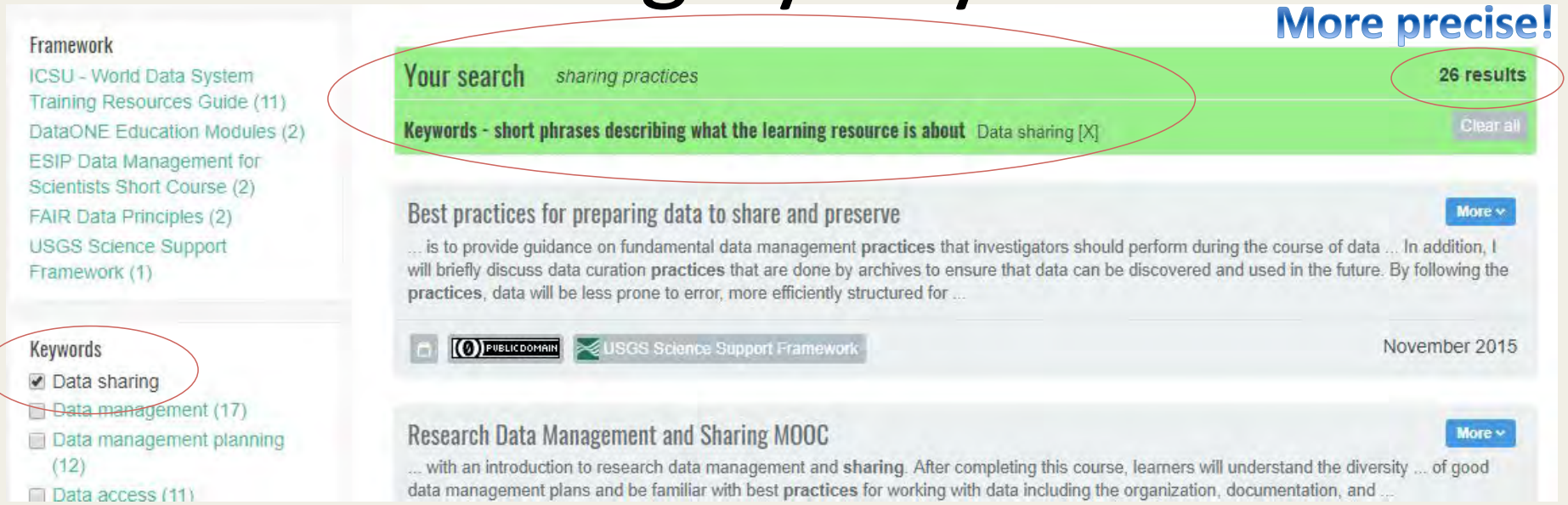




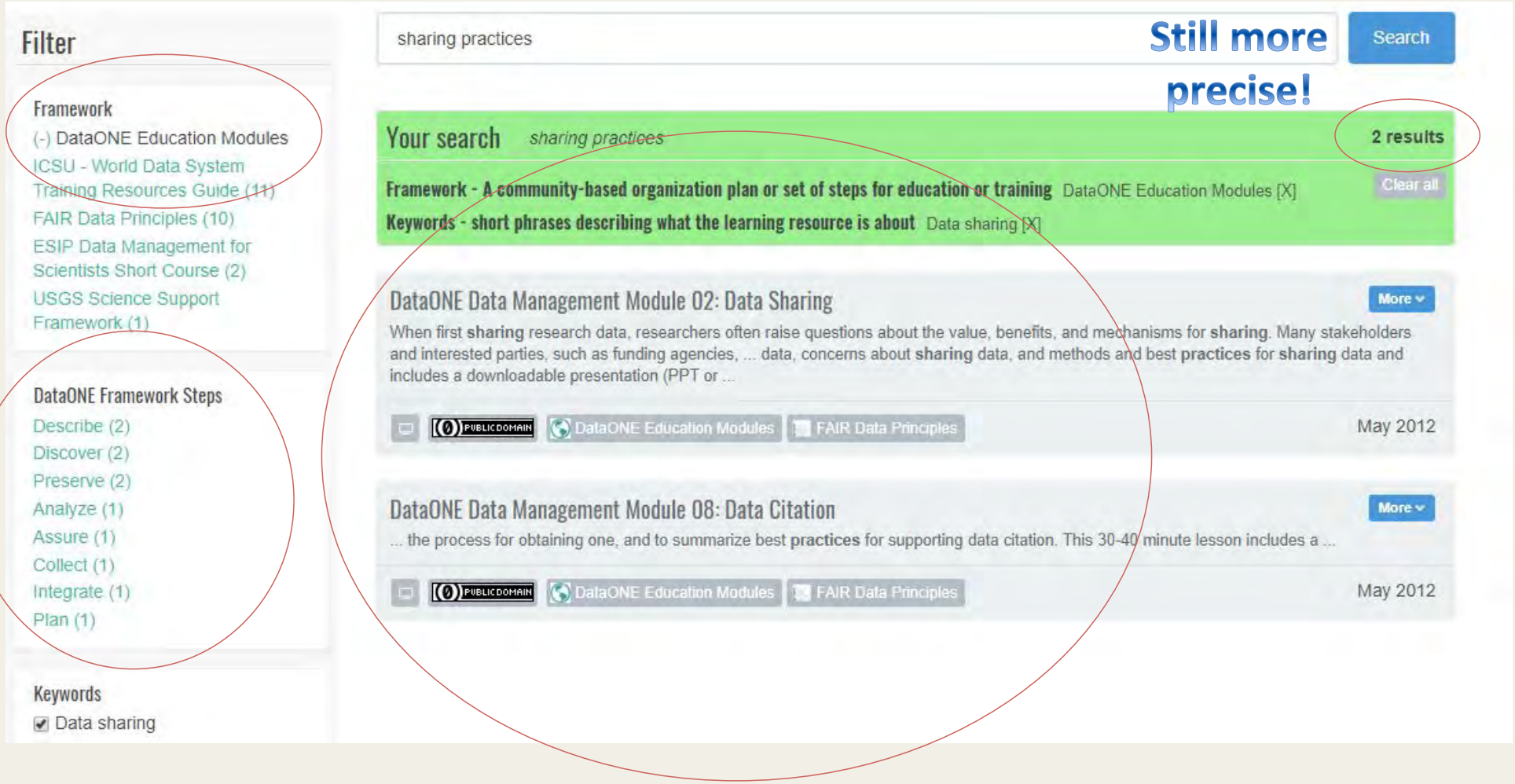
Demonstrating Approach 3: Start by entering your own search term…



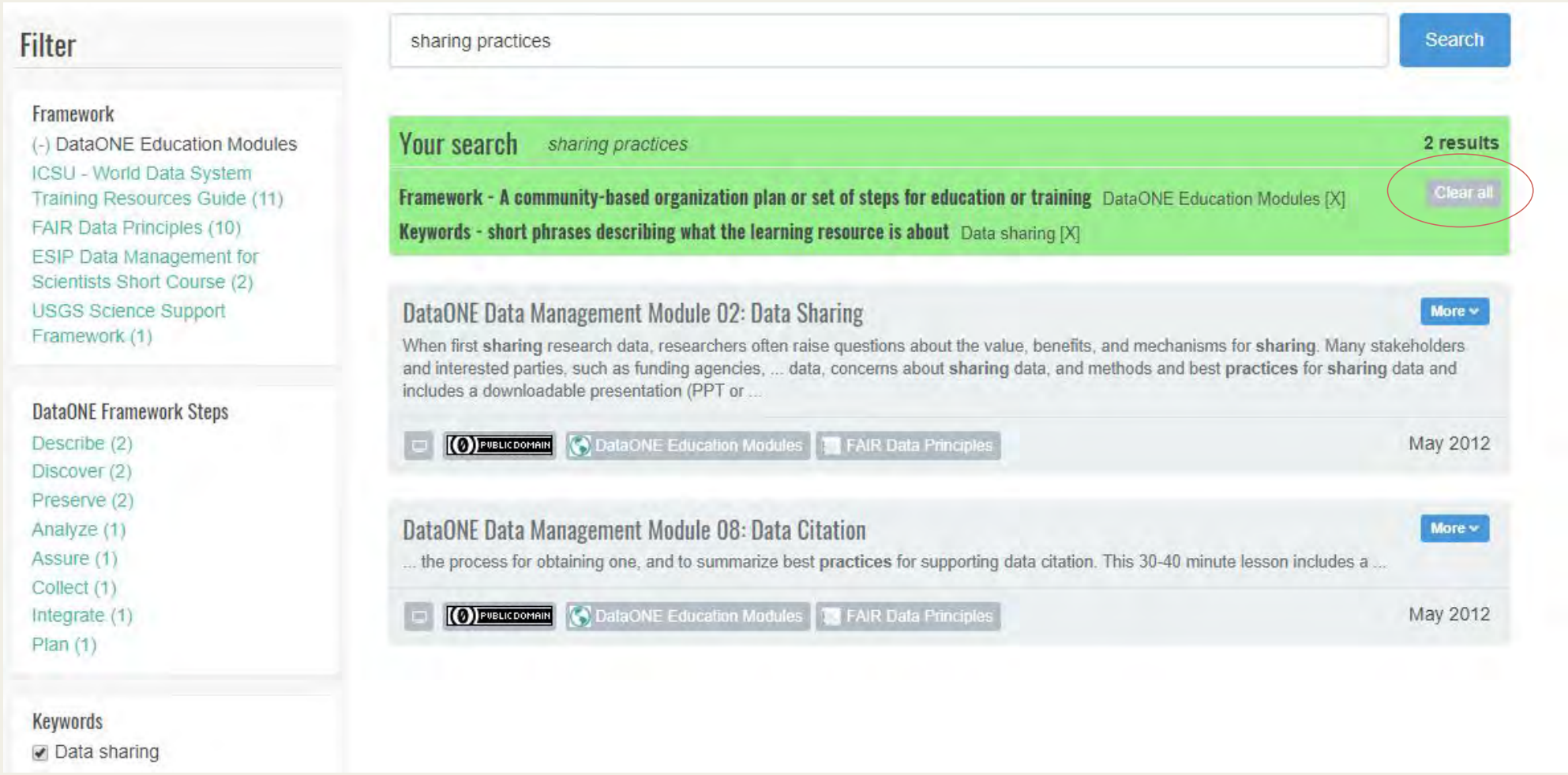
… then limiting by Keyword filter…



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



Full Description – 1 of 2

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

Full Description – 2 of 2

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

ESIP Data Management Training EXPLORE

Home Browse Search Submit Help About Log in

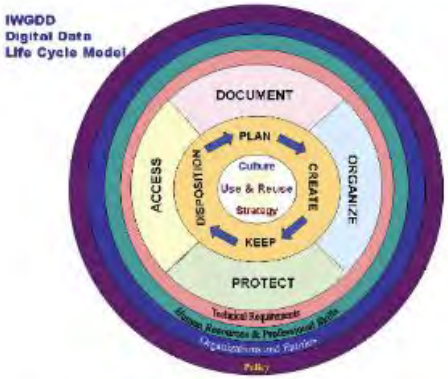
Welcome to the DMT Clearinghouse

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

It was created in a 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

Read More



IWGDD Lifecycle - https://www.nitrd.gov/About/Harnessing_Power_Web.pdf

Search

Find learning resources by keyword, name, date, license and cost

sharing practices Search

Browse

See a list of learning resources by educational framework

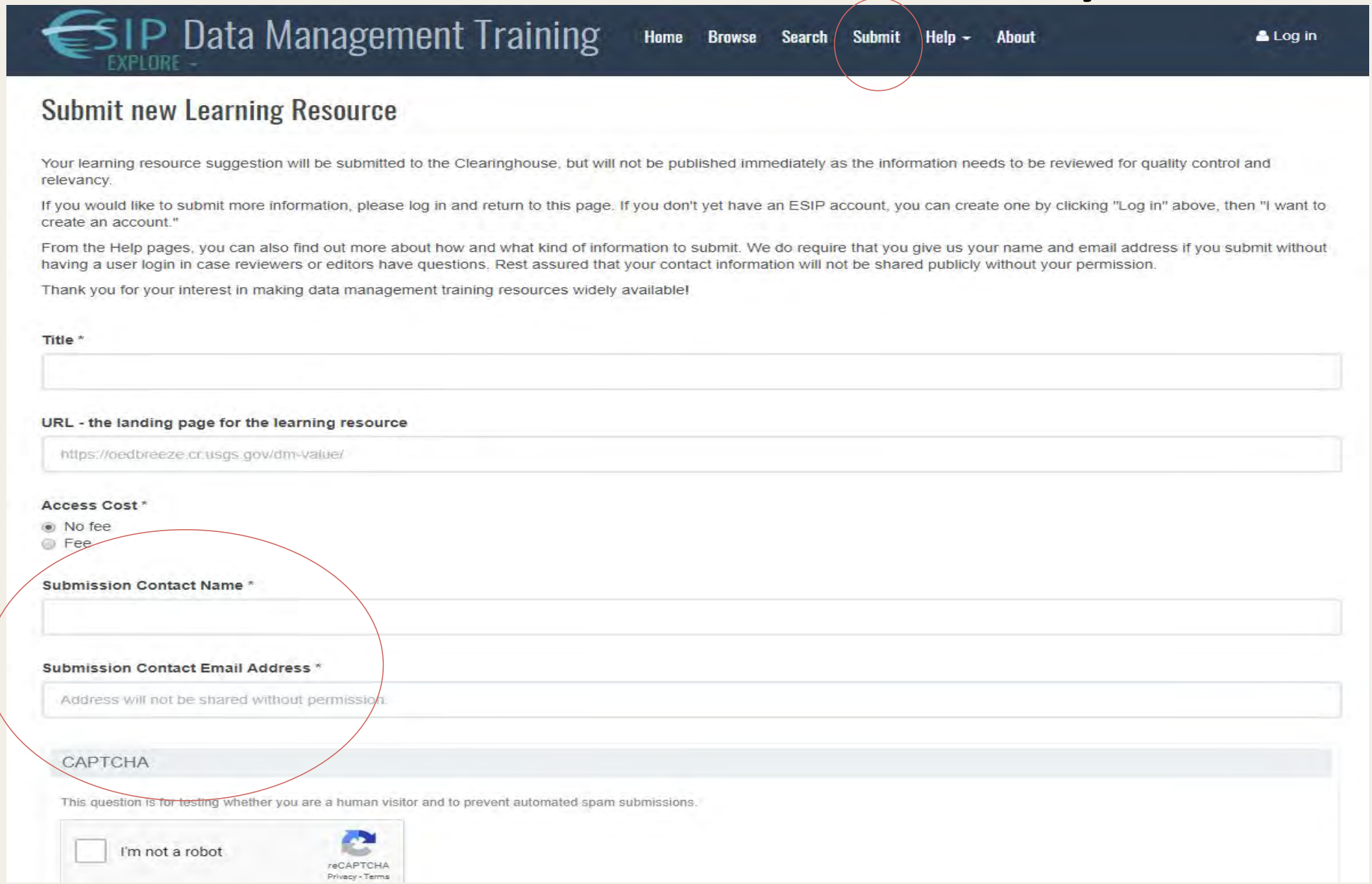
Browse

Submit

Submit your learning resources to the Clearinghouse

Submit

Submit Function, briefly



As a community supported resource we'd love to have you to join us by…

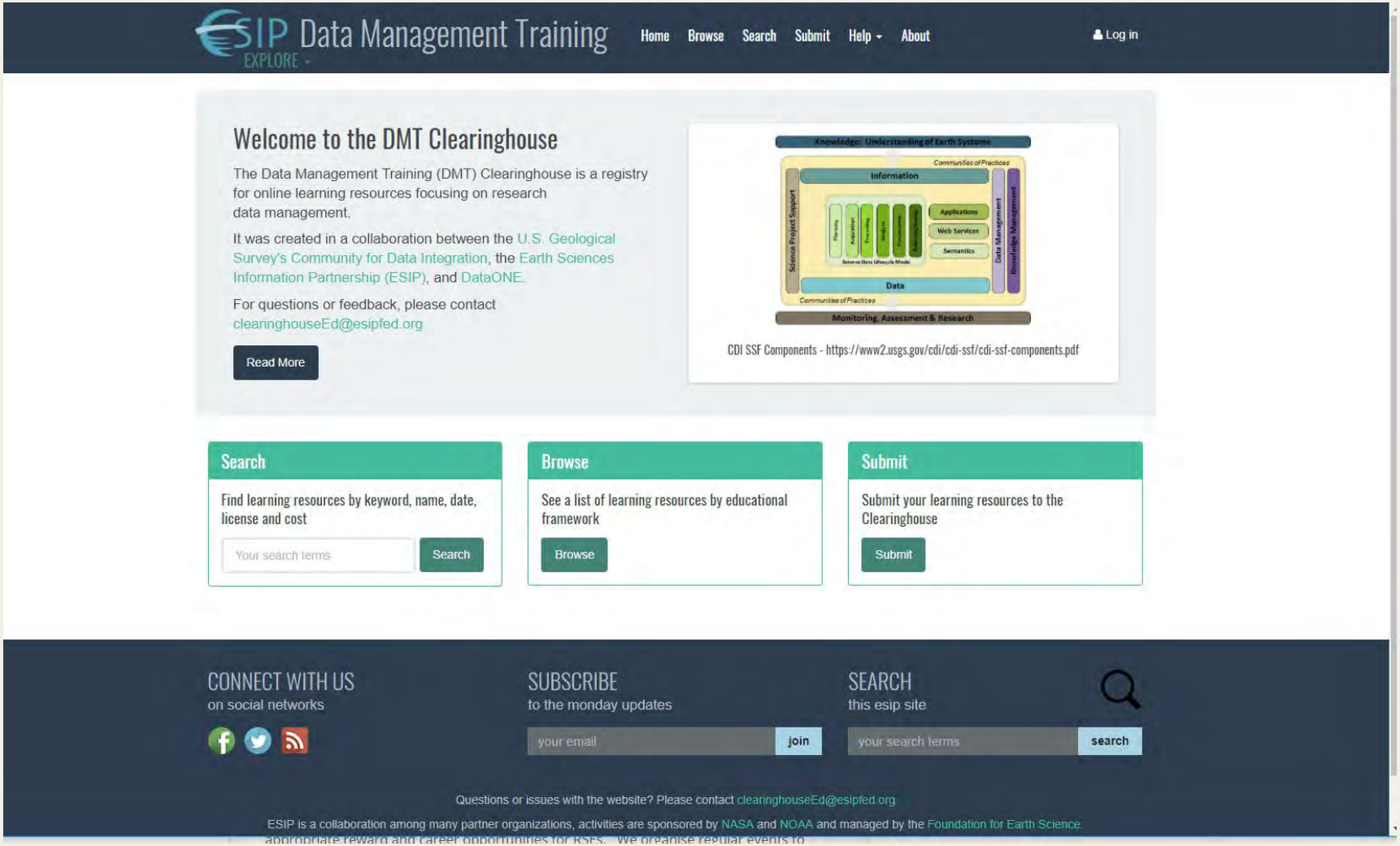
- Submitting your learning resources
- Joining our Working Groups on
  - Assessment Framework
  - Metadata Enhancement
  - Content Diversification
- Editorial assistance
- Usability testing
- Spreading the word
- Jumping in on our crowdsourcing events

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

http://dataoneorg.github.io/Education

DMT Clearinghouse

- Index here
- Metadata
- This is a registry

http://dmtclearinghouse.esipfed.org

AGU 100 ADVANCING EARTH AND SPACE SCIENCE

FALL MEETING

Washington, D.C. | 10–14 Dec 2018

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

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

DataONE

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)