

Data Sharing

**DataONE Community Engagement & Outreach Working
Group**

Data Sharing



Public domain image from Unsplash

Lesson Topics

- Data sharing within the **data lifecycle**
- **Value** of data sharing
- **Concerns** about data sharing
- **Methods** for making data sharable



Public domain image from Unsplash

Learning Objectives

After completing this lesson, the participant will be able to:

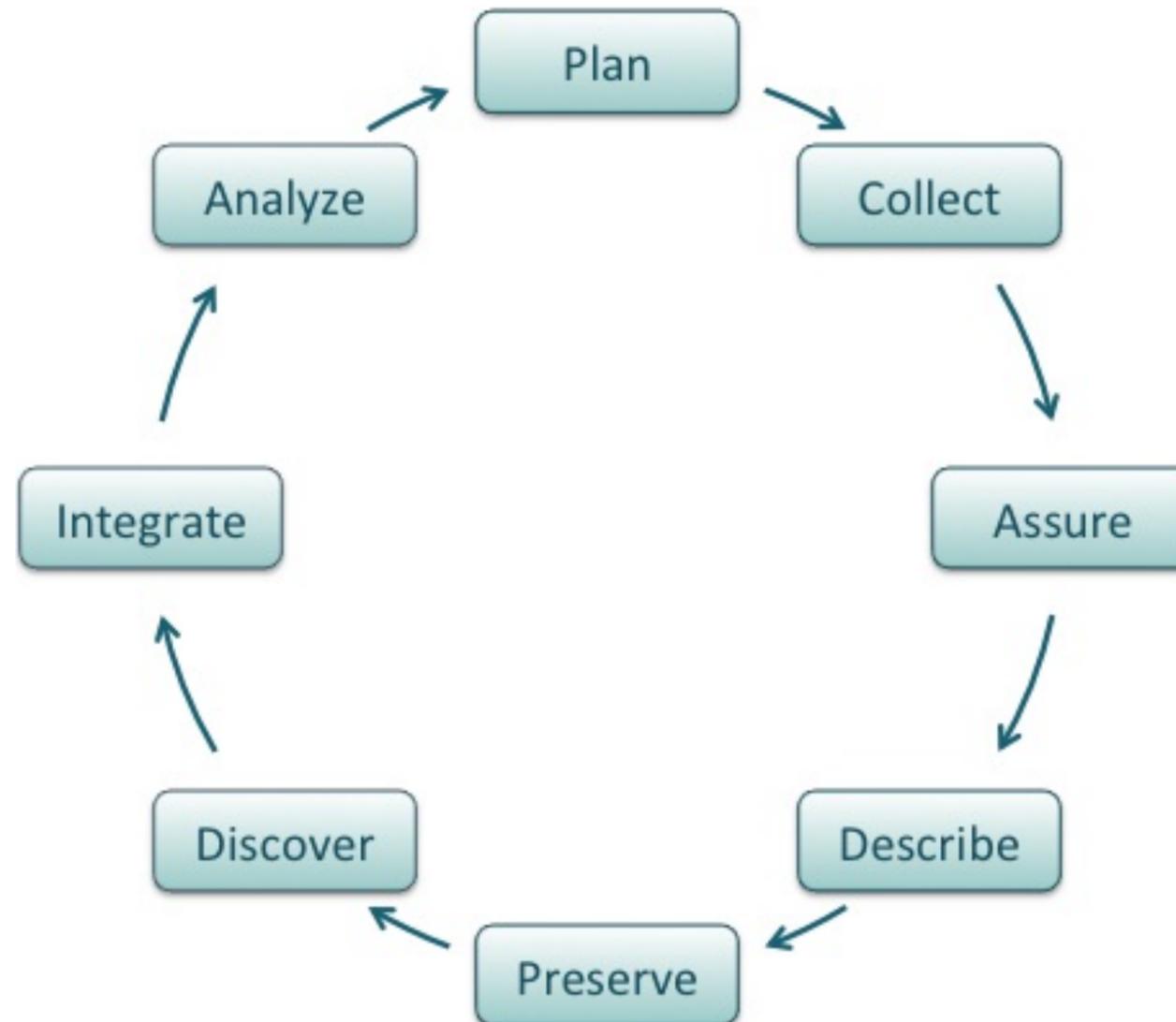
1. Discuss data sharing considerations within the data lifecycle
2. Recognize the benefits of sharing scientific data
3. Address concerns about sharing data
4. Outline a process for making data sharable
5. Identify mechanisms for sharing data



pictureYouth

The Data Lifecycle

Data sharing should be addressed throughout the data lifecycle.



The Data Lifecycle

Several stages require critical attention to ensure effective data sharing

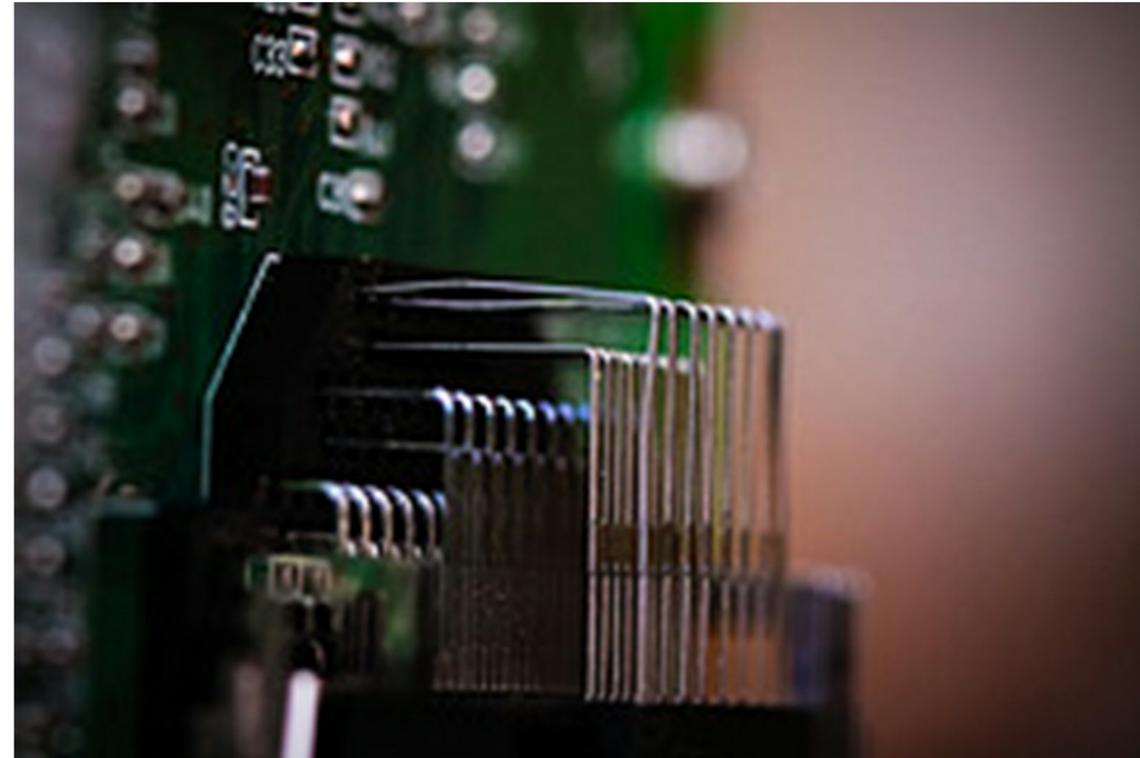
Step	Action
Describe	document the data content, character and process
Deposit	store the data in a location from which it can be accessed
Preserve	select storage formats and media with long term use in mind
Discover	publish information about the data so that others can find it

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

Value of Data Sharing: To the Public

A better informed public yields better decision making with regard to:

- Environmental and economic **planning**
- Federal, state, and local **policies**
- **social choices** such as use of tax dollars and education options
- personal **lifestyle and health** such as nutrition and recreation



CC image by falonyates on Flickr

Value of Data Sharing: To Research Sponsor

- Organizations that sponsor research must maximize the value of research dollars
- Data sharing enhances the value of research investments by enabling:
 - verification of performance **metrics and outcomes**
 - new research and increased **return on investment**
 - advancement of the **science**
 - **reduced** data duplication **expenditures**
 - enhancing and **extending the record** of science

Value of Data Sharing: To Scientific Community

Access to related research enables community members to:

- build upon the work of others and further, rather than repeat, the science
- perform meta analyses that cannot be performed with individual datasets or laboratories
- share resources and perspectives so that comprehension is expanded and enhanced



CC image by Lawrence Berkeley National Laboratory on Flickr

Value of Data Sharing: To Scientific Community

Access to related research enables community members to (cont'd):

- increase transparency, reproducibility and comparability of results
- expand methodology assessment, recommendations and improvement
- educate new researchers as to the most current and significant findings



Value of Data Sharing: To the Scientist

Scientists that share data gain the benefit of:

- research sponsor recognition as an authoritative source and wise investment
- improved data quality due to expanded use, field checks, and feedback
- greater opportunity for data exchange
- improved connections to scientific network, peers, and potential collaborators



CC image by SLU Madrid Campus on Flickr

Concerns About Data Sharing

Even if the value of data sharing is recognized, concerns remain as to the impacts of increased data exposure.



Public domain image by succo on pixabay

Concerns About Data Sharing

Problem	Solution
inappropriate use due to misunderstanding of research purpose or parameters	?
security and confidentiality of sensitive data	?
lack of acknowledgement / credit	?
loss of advantage when competing for research dollars	?

Concerns About Data Sharing

Problem	Solution
inappropriate use due to misunderstanding of research purpose or parameters	metadata
security and confidentiality of sensitive data	metadata
lack of acknowledgement / credit	metadata
loss of advantage when competing for research dollars	metadata

Concerns About Data Sharing

Problem	Solution
inappropriate use due to misunderstanding of research purpose or parameters	provide rich Abstract, Purpose, Use Constraints and Supplemental Information where needed
security and confidentiality of sensitive data	the metadata does NOT contain the data Use Constraints specify who may access the data and how

Concerns About Data Sharing Cont'd

Problem	Solution
lack of acknowledgement / credit	specify a required data citation within the Use Constraints
loss of advantage when competing for research dollars	create second, public version with generalized Data Processing Description

Making Data Sharable

Step One:

- Create robust metadata that is discoverable
- specify geography and time periods
- use discipline specific theme, place and temporal keywords, thesauri, and ontologies
- describe attributes
- include links to associated data catalogues, data downloads, project websites, etc.

Making Data Sharable

Step Two: Include archival and reference information

- properly formatted data citations for the data and all sources
- Universally Unique Identifiers (UUID) that uniquely identify your data and help to link the data with the metadata See the DataONE unique identifier guidance at: <http://mule1.dataone.org/ArchitectureDocs-current/design/PIDs.html>

Data Citation Example: Sidlauskas, B. 2007. Data from: Testing for unequal rates of morphological diversification in the absence of a detailed phylogeny: a case study from characiform fishes. Dryad Digital Repository. doi:10.5061/dryad.20

Making Data Sharable

Step Three: Have data contributors review your metadata to ensure validity and organizational ‘correctness’?

- are the processes described accurately?
- are all contributions adequately identified?
- has management reviewed the product and documentation?
- is the funding organization properly recognized?

Making Data Sharable

Step Four: Publish your metadata and/or data via:

- **Federal Data Catalogs**
 - data.gov
- **Data Repositories**
 - Knowledge Network for Biodiversity (KNB) Data Portal
 - Long Term Ecological Research (LTER) Network Data Portal
 - Institutional data repositories
- **Other Online Resources**
 - Project and/or Program websites
 - Web-accessible folders (WAF)
 - Community or Public Cloud
- Searchable directory of repositories for publishing your data
 - <http://service.re3data.org/search>

Making Data Sharable

Step Four: Publish your metadata and/or data via: (cont'd)

- **Other Online Resources**

- Project and/or Program websites
- Links within online lessons and outreach products
- Web-accessible folders (WAF)
- Community or Public Cloud

Best Practices

- Document and publish data using standards
- Promote data use via presentations and meetings
- Solicit feedback from data users and address identified issues
- Monitor publications and websites for data use and address misapplications

Data in Real Life

- In 2003, a group of scientists from the National Institutes of Health, the Food and Drug Administration, drug and medical imaging industries, universities, and nonprofit groups joined in a collaborative effort to find the biological markers that show the progression of Alzheimer's disease in the human brain.
- The goal of this project was to do research on a massive scale that would involve sharing and making accessible all the data uncovered to anyone in the world with a computer.
- Dr. John Trojanowski an Alzheimer's researcher at the University of Pennsylvania stated, "It's not science the way most of us have practiced it in our careers. But we all realized that we would never get biomarkers unless all of us parked our egos and intellectual-property noses outside the door and agreed that all of our data would be made public immediately."

<http://www.nytimes.com/2010/08/13/health/research/13alzheimer.html>



Summary

- Data sharing adds value to the data
- It is the responsibility of the researcher to share their data
- Metadata supports data accountability, liability, and usability
- Sponsors expect, some require, data to be shared
- Data sharing is essential to the advancement of science

About

Participate in our GitHub repo: https://dataoneorg.github.io/dataone_lessons/

The full slide deck (in PowerPoint) may be downloaded from:

<http://www.dataone.org/education-modules>

Suggested citation: DataONE Education Module: Data Management. DataONE.

Retrieved November 12, 2016. From

http://www.dataone.org/sites/all/documents/L01_DataManagement.pptx

Copyright license information: No rights reserved; you may enhance and reuse for your own purposes. We do ask that you provide appropriate citation and attribution to DataONE.



