Data + Community = Action

Erin Robinson,
Executive Director
ESIP

From space we can’t see barriers between people. Space reminds us of our common bonds – let’s never forget that.

UNOOSA @UNOOSA · Nov 29
From space we can’t see barriers between people. Space reminds us of our common bonds – let’s never forget that.

Helmholtz Open Science Webinars
Webinar 49 – 23 / 28 May 2019
Kansas Agricultural Smoke, April 12, 2003
Making Data Matter

Old Way

Pre-Science: Find data, Retrieve high volume data, Learn formats and develop readers, Extract parameters, Perform spatial and other subsetting, Identify quality and other flags and constraints, Perform filtering/masking, Develop analysis and visualization, Accept/discard/get more data (sat, model, ground-based).

Jan

Mar

[Open] Science User Barriers [to Open Data]

“The user cannot find the data; if he can find them, he cannot access them; if he can access them, he doesn't know how good they are; if he finds them good, he cannot merge them with other data.”

Information Technology and the Conduct of Research: The User’s View (1988)

Jun

Sept

Adapted from Leptoukh, 2012

DO SCIENCE: Exploration, Initial Analysis, Use the best data for the final analysis, Derive conclusions, Write the paper, Present @ AGU.

Dec
Making Data Matter

Old Way

Pre-Science
- Find data
- Retrieve high volume data
- Learn formats and develop readers
- Extract parameters
- Perform spatial and other subsetting
- Identify quality and other flags and constraints
- Perform filtering/masking
- Develop analysis and visualization
- Accept/discard/get more data (sat, model, ground-based)

DO SCIENCE
- Exploration
- Initial Analysis
- Use the best data for the final analysis
- Derive conclusions
- Write the paper
- Present @ AGU

Jan
Mar
Jun
Sept
Dec

Adapted from Leptoukh, 2012
FAIR Guiding Principles

FAIR is...
Findable
Accessible
Interoperable
Reusable

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data.

- From The FAIR Guiding Principles for scientific data management and stewardship
There is an urgent need to improve the [Global Collaborative] infrastructure supporting the (re)use of scholarly data.

- (Modified, Erin Robinson) From The FAIR Guiding Principles for scientific data management and stewardship
ESIP Vision

Leaders in promoting
the **collection, stewardship and (re)use**
Of Earth science data, information and knowledge
that is responsive to societal needs.
Earth Science Information Partners (ESIP)

By the numbers:

120+ Member Orgs
1000 Active Participants
30+ Working Groups

Supported by:
Types of ESIP Members

Type I: Data Centers

Type II: Researchers and Tool Developers

Type III: Application Developers

Type IV: Strategic Partners

MOU Collaborators

And many more! ESIP has 110+ Members
What ESIP community says...

The best network for those who want to work together across science domains and the data life cycle.

Magnet for experts and leaders – draws top-notch data, tech and science professionals together.

Highly skilled people who are a good sounding board for ideas.

Clear professional ‘home’ for earth science data people.

People participate across data life cycle – producers, researchers, to resolve common issues & create common pathways.
ESIP does:
Generate influential recommendations and work products
Have a lasting impact in the recommendation of standards.

ESIP does not:
• Provide data
• Sustain cyberinfrastructure
• Compete with our members
• Develop standards
DATA CITATION GUIDELINES

- ESIP Data Stewardship created
- ESIP Assembly endorsed in 2012 (Way ahead of it’s time)
- Served as a model for NASA, NOAA, NSF, Group on Earth Observations, ...
- ESIP has been influential in Force11 and RDA, influencing directions based on this work

http://commons.esipfed.org/node/308
THE DATA MANAGEMENT PUZZLE

Making research data: Shareable, Findable, Reproducible, Available for the long term

Who cares?

What is it?

Quick: 7 – 15 minutes each
courses complement each other

Download as slides or videos

Created by experienced scientists

Peer reviewed

Free of Charge

Benefits of short course

33 Modules
Created in late 2012-2013

http://commons.esipfed.org/datamanagementshortcourse
How do I learn / teach RDM?

Who else is doing?

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

Web Portal / Metadata Content Repository/ Clearinghouse

Search Browse Submit

Lab / Classroom / 1 on 1

Self – taught Researcher

Lead: Nancy Hoebelheinrich, Knowledge Motifs
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

http://dmtclearinghouse.esipfed.org
All Learning Resources

DataONE Data Management Module 02: Data Sharing
- Analyze
- Assure
- Collect
- Describe
- Discover
- Integrate
- Plan
- Preserve
May 2012

DataONE Data Management Module 03: Data Management Planning
- Analyze
- Plan
May 2012

Creating Documentation and Metadata: Creating a Citation for Your Data
- Analyze
- Plan
ESIP Data Management for Scientists Short Course
Local Data Management
September 2012

Local Data Management - Data Formats: Using Self-describing Data Formats
- Analyze
- Plan
ESIP Data Management for Scientists Short Course
Local Data Management
January 2013

Dmtclearinghouse.esipfed.org
2017 AGU received a Grant from Laura and John Arnold Foundations (LJAF)

Align publishers and repositories in following best practices to enable FAIR and open data and to create workflows so that researchers will have a simplified, common experience when submitting their paper to any leading Earth and space science journal.
Community-Driven Project – Partnership Includes:

- **Science Data Communities**
  - AGU and EGU
  - Earth Science Information Partners (ESIP)
  - Research Data Alliance (RDA)
  - EarthCube / Council for Data Facilities
  - FORCE11
- **Publishers**
  - AGU
  - Proceedings of the National Academy of Sciences (PNAS)
  - Nature
  - Science/AAAS
  - Elsevier
  - PLOS
  - Hindawi
  - Copernicus
  - Wiley
- **International Repositories**
  - National Computational Infrastructure (NCI)
  - AuScope
  - Australian Research Data Commons (ARDC)
  - Center for Open Science
  - DataCite / re3data
- **And Growing!!**
  - ORCID
  - CrossRef
  - CHORUS
  - Scholix
  - OSGeo
  - Pangaea
  - DataONE
Enabling FAIR Data Project - Objectives

• FAIR-aligned data repositories add value to research data, provide metadata and landing pages for discoverability, and support researchers with documentation guidance, citation support, and curation.

• FAIR-aligned Earth, space, and environmental science publishers align their policies to establish a similar experience for researchers. Data, software, technology will be available through citations that resolve to repository landing pages. Availability statements are provided.

Data are not placed in journal supplements.
Commitment Statement

Link to the Commitment Statement: http://bit.ly/FAIRCommitment (case sensitive)

Final Version
FAIR-Aligned: Researcher Commitment

• Locating trustworthy, community-accepted, FAIR-aligned repositories that support:
  – Documenting data and software (and other research outputs as is possible) to agreed community standards that describe provenance and enable discovery, assessment of reliability, and reuse
  – Persistent identifiers for data and software (and other research outputs as is possible)
  – Licenses for data and software (and other research outputs as is possible) that is as open as possible to enable the widest potential reuse.

• Citing data, software, physical samples, and other research products
• Developing data availability statements
• Preparing and managing data management plans. Make them living documents.
Resource Object Citation Cluster

Title: Software and Services Citation Guidelines and Examples
Collaboration Area:
Software and Service Citations Cluster
DOI:
10.6084/m9.figshare.7640426
Technical Reports:
Version: 1
Recommended Citation:
Figshare link:
https://esip.figshare.com/articles/Software_and_Services_Citation_Guidelines_and_Examples/7640426

Document Status
Approved by ESIP Assembly Meeting 16 January 2019

Data Citation Guidelines for Earth Science Data
Version 2

Suggested Citation:

Table of Contents
Document Status 2
Related ESIP Documents 2
Introduction 2
Citation Content 3
Overview 3
Details on Core Concepts 4
Author or Creator 4
Public Release Date 5
Title 6
Version ID 6
Repository 7
Resolvable Persistent Identifier (PID) 7
Access Data and Time 8
Additional Considerations 8
Resource type 9
Editor, Compiler, or other important roles 9
Data Within a Larger Work 9
Dynamic and Micro-citation 10
Versioning 10
Submit Used 11
Resolving Citations 12
Note on Locators vs. Identifiers 12
Landing Pages 13
Content 14
Actionability 15
Acknowledgements 16
Bibliography 16
Appendix: Mapping of Core Concepts to Common Metadata Dialects 18
Resource Object Citation Cluster

• New cluster spun of the Data Stewardship Committee and in combination with the Software and Services Citation Cluster

• First task: update data citation guidelines.
  • Refinement core concepts and issues and mapped concepts to more metadata dialects
  • New guidance on “Dynamic data citation” notably to use the RDA Recommendation
  • New section on resolving citations: especially how to construct landing pages and make them machine actionable.

• Now examining all the “concerns” and research objects that citation can or should address.
Resource Object Citation Cluster

• New cluster spun from the Data Stewardship Committee and in combination with the Software and Services Citation Cluster

• First task: update **data citation guidelines**.
  
  • Refinement core concepts and issues and mapped core metadata dialects

  • New guidance on “Dynamic data citation” notably the RDA Recommendation

  • New section on resolving citations: especially how to construct landing pages and make them machine actionable.

• Now examining all the “concerns” and research objects that citation can or should address.
Repository Finder, a pilot project of the Enabling FAIR Data Project led by the American Geophysical Union (AGU) in partnership with DataCite and the Earth, space and environment sciences community, can help you find an appropriate repository to deposit your research data. The tool is hosted by DataCite and queries the re3data registry of research data repositories.

Search re3data for a repository to upload your data

Search

See the repositories in re3data that meet the criteria of the Enabling FAIR Data Project.
FAIR Data Principles

Overview of Interdisciplinary Earth Data Alliance (IEDA) Data Management Resources

Introduction to Scientific Visualization

Simplifying the Reuse and Interoperability of Hydrologic Data Sets and Models with Semantic Metadata that is Human-Readable & Machine-Actionable

EarthChem Library: How to Complete a Data Submission Template

iData Tutorial
The possibility of being able to implement things that you thought about 20 years ago because the computational capability is available now is quite exciting.
– Rama Ramapriyan, NASA/SSAI
Data Citation

DMT Clearinghouse

Research Object Citation

FAIR Labeled resources

Enabling FAIR Researchers

Data Management Training
“You can’t connect the dots looking forward.”
- Steve Jobs
Thank you!

Acknowledgements:
ESIP is funded by NASA, NOAA & USGS
Work presented was done by ESIP’s Data Stewardship Committee
Nancy Hoebelheinrich & Karl Benedict are DMT Clearinghouse Leads