

Do I have to make my models FAIR?

Current practices in making models and data

Findable, Accessible, Interoperable, and Reusable

Leslie Hsu, lhsu@usgs.gov

CSDMS Webinar

April 24, 2019



Intended outcome of today's presentation

- You are able to explain to someone else what FAIR means for CSDMS-related data
- You know where to go to find out more about FAIR
- You have some strategies for making your proposals and publications FAIR-savvy



Poll 1

On a scale of 1-5, how comfortable are you with explaining what FAIR means for your models and data? (5 is most comfortable)

Answer at <https://app2.sli.do/event/twpq1kcu>

Or, use your mobile device browser and go to [slido.com](https://www.slido.com), Event Code #3580

What do we mean by FAIR?



This article describes four foundational principles—Findability, Accessibility, Interoperability, and Reusability—that serve to guide data producers and publishers as they navigate around these

come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to

Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3 (2016). <https://doi.org/10.1038/sdata.2016.18>

What do we mean by FAIR?



scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first

digital publishing. Importantly, it is our intent that the principles apply not only to 'data' in the conventional sense, but also to the algorithms, tools, and workflows that led to that data. All

Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3 (2016). <https://doi.org/10.1038/sdata.2016.18>

What do we mean by FAIR?

Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3 (2016). <https://doi.org/10.1038/sdata.2016.18>

What do we mean by FAIR?

1. Foundational principles to guide data producers and publishers
2. Concise and measurable
3. Emphasis on machines' ability to find and use data
4. "Data" also means algorithms, tools, workflows

Why are we talking about FAIR?



Signatories to the commitment statement include AGU, UNAVCO, Open Topography, USGS, Wiley, Elsevier, others...

Author guidelines strive to get publishers to address the same set of topics related to publication of supporting data.

Why are we talking about FAIR?



National Science Foundation
WHERE DISCOVERIES BEGIN



Award Abstract #1835822

**Framework: Data: HDR: Extensible Geospatial Data Framework
towards FAIR (Findable, Accessible, Interoperable, Reusable) Science**

NSF Org: [OAC](#)
[Office of Advanced Cyberinfrastructure \(OAC\)](#)



Award Abstract #1839030

**Findable Accessible Interoperable and Reusable (FAIR) Hackathon
Workshop for Mathematical and Physical Sciences (MPS) Research
Communities**

NSF Org: [OAC](#)
[Office of Advanced Cyberinfrastructure \(OAC\)](#)



Award Abstract #1838958

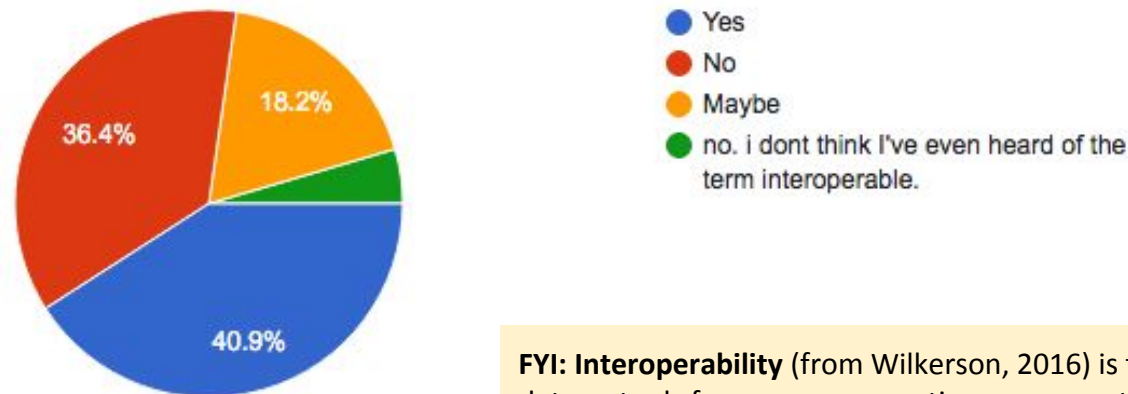
**FAIR Publishing Guidelines for Spectral Data and Chemical Structures
in Support of Chemistry and Related Disciplinary Communities**

NSF Org: [OAC](#)
[Office of Advanced Cyberinfrastructure \(OAC\)](#)

What you said about FAIR

Have you ever heard of the acronym FAIR (findable, accessible, interoperable, reusable)?

22 responses

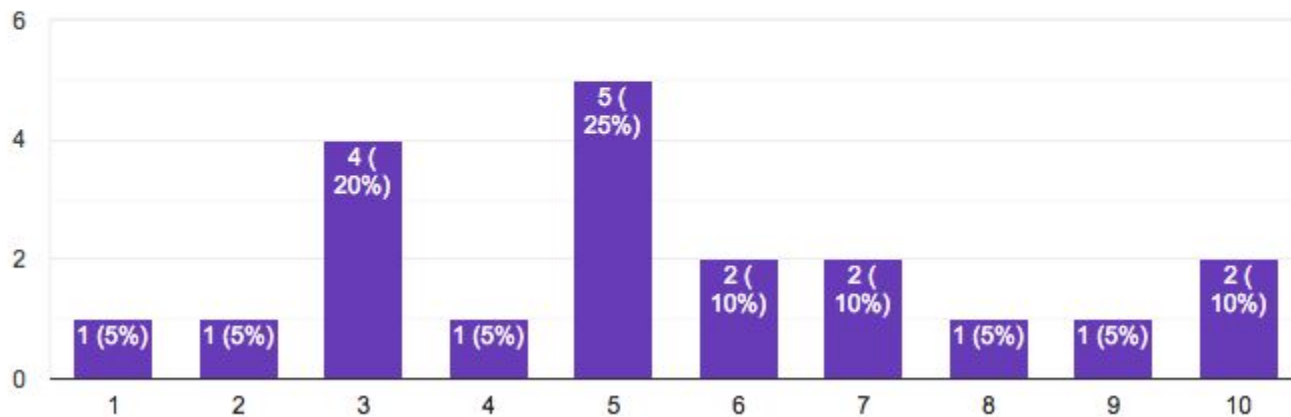


FYI: Interoperability (from Wilkerson, 2016) is the ability of data or tools from non-cooperating resources to integrate or work together with minimal effort.

What you said about FAIR

On a scale of 1-10, how well do you think you make your models FAIR?

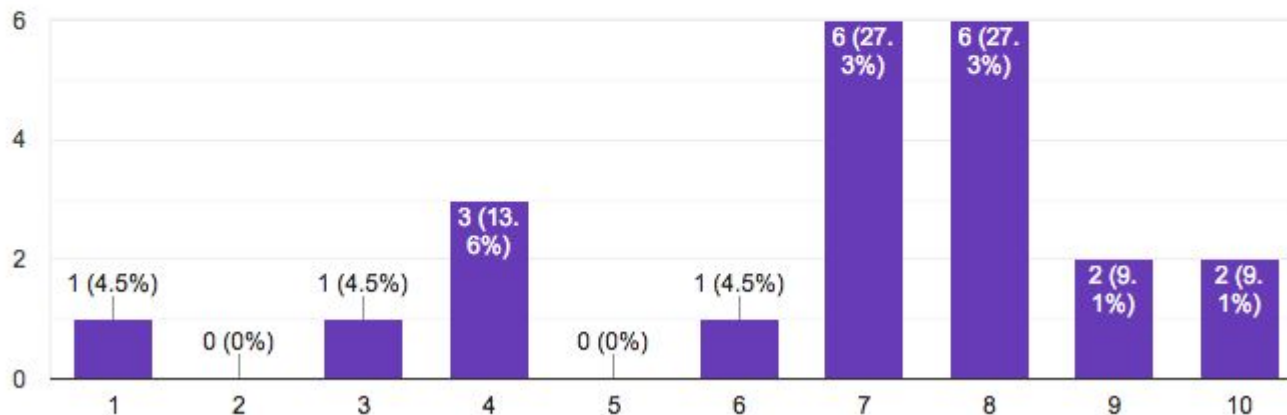
20 responses



What you said about FAIR

On a scale of 1-10, how well do you think you make your data FAIR?

22 responses



What you said about FAIR

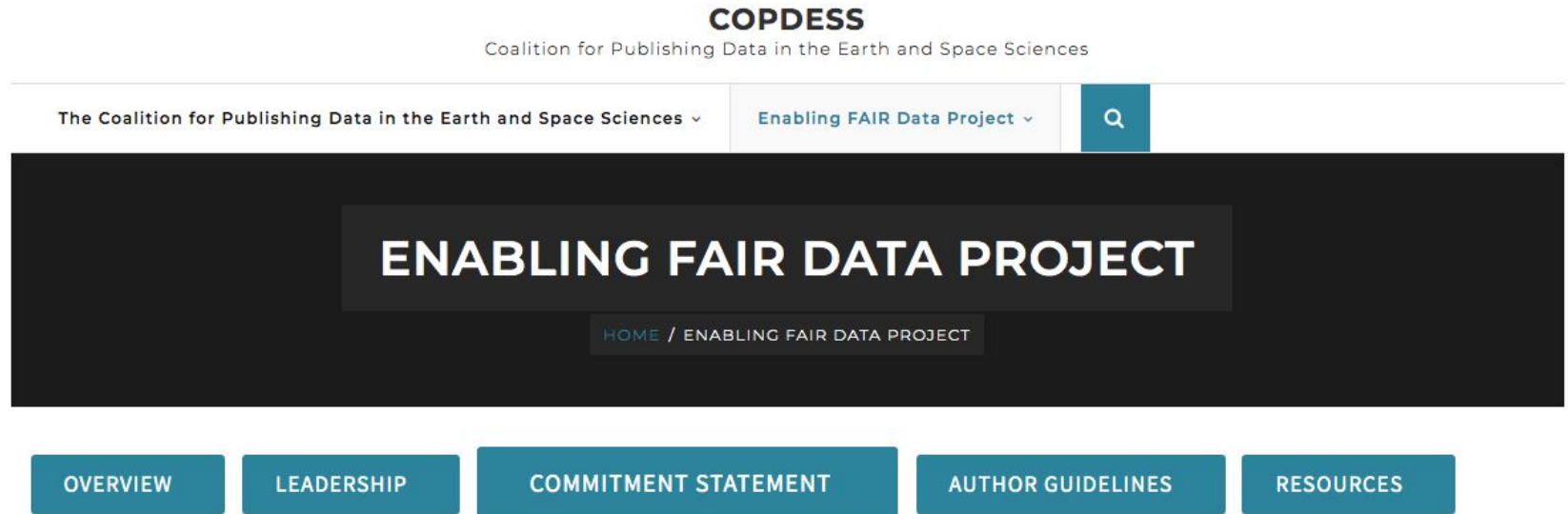
FAIR seems like a moving target... there are multiple unknowns

Effort makes it a barrier, need more carrots and sticks

FAIR for models and data may be different

... now it's just intimidating

Resources for understanding FAIR



<http://www.copdess.org/enabling-fair-data-project/enabling-fair-project-overview/>

Resources for understanding FAIR



A bottom-up international approach

for the practical implementation of the European
Open Science Cloud (EOSC) as part of a global
Internet of FAIR Data & Services

<https://www.go-fair.org/>
<https://www.go-fair.org/fair-principles/>

Where to stay up to date on FAIR



Three decisions you can make to increase the FAIRness of your models/data in proposals and publications

Where to share
your code

What license
to use

What information
would you need for
your code to be
reusable to yourself?

Where to share your code

- A place where your colleagues go and therefore will be more likely to come across it.
- A place that is set up to be comprehensively indexed by different searches.
- A place that allows versioning and persistent identifiers like DOIs.
- A place that subscribes to the FAIR principles/commitment statement.



What license to use

PEOPLE GET VERY
_____ ABOUT
LICENSING



<https://softwaresaved.github.io/software-licensing-workshop/#/>

What license to use

Considerations in choosing a license

- What rights do you want to retain or grant?
 - Who can use the program? (proprietary vs open)
 - Can users see the source code? (proprietary vs open)
 - Can users modify the source code? (proprietary vs open)
 - Can the users redistribute original or modified code? (proprietary vs open)
 - Can modified code be relicensed? (permissive vs copyleft)
- Compatibility with software under other licenses
 - Permissive licenses have fewer issues
 - <http://www.fsf.org/licensing/>
- Labeling of derived works
 - Derived works must be identified differently than original work
- Patent grant/retaliation
- Expectations of the community you want to engage?

***Use an existing
free/open source
license rather than
inventing a new one!***

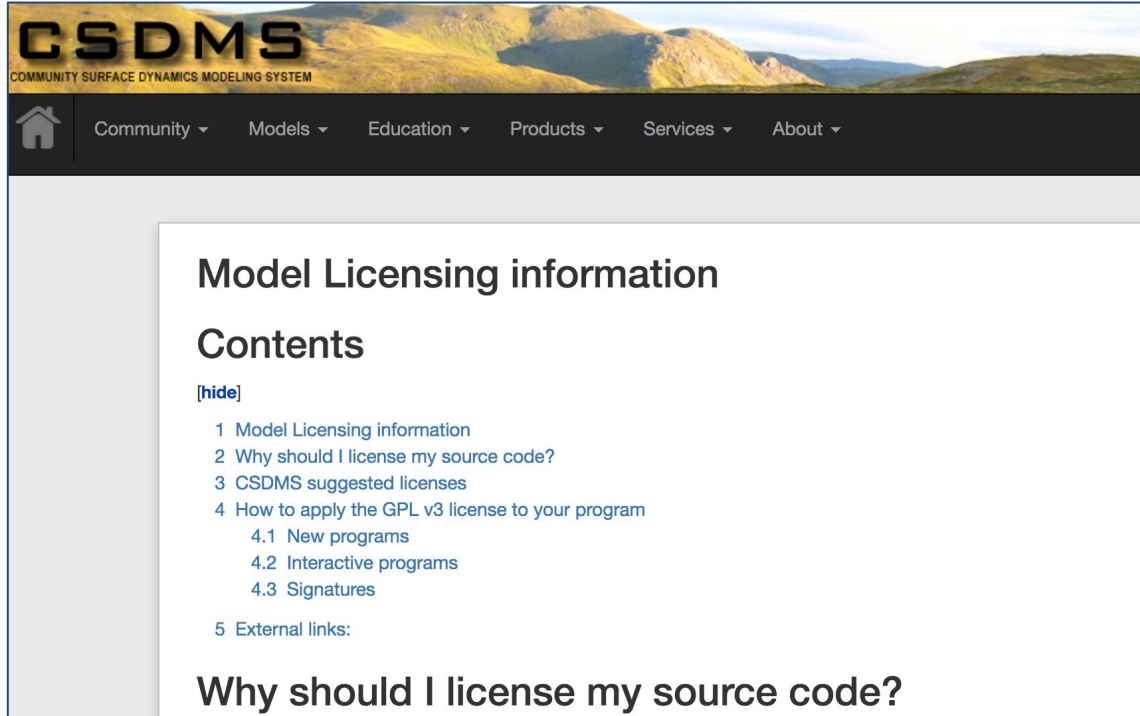
*FSF and OSI certify
many existing licenses
(~80) as meeting their
criteria*

14



https://figshare.com/articles/An_Introduction_to_Software_Licensing/7409573

What license to use



The screenshot shows the CSDMS (Community Surface Dynamics Modeling System) website. The header features the CSDMS logo and a navigation bar with links to Community, Models, Education, Products, Services, and About. The main content area is titled 'Model Licensing information' and includes a 'Contents' section with a list of links. The first link is '1 Model Licensing information', which is the current page. The second link is '2 Why should I license my source code?', which is the section highlighted in the image. The third link is '3 CSDMS suggested licenses'. The fourth link is '4 How to apply the GPL v3 license to your program', which has sub-links for '4.1 New programs', '4.2 Interactive programs', and '4.3 Signatures'. The fifth link is '5 External links:'.

CSDMS
COMMUNITY SURFACE DYNAMICS MODELING SYSTEM

Community ▾ Models ▾ Education ▾ Products ▾ Services ▾ About ▾

Model Licensing information

Contents

[\[hide\]](#)

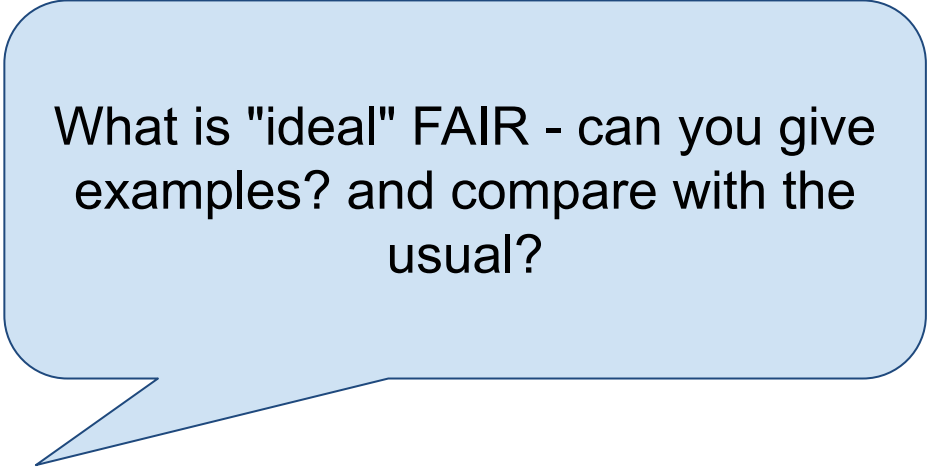
- 1 Model Licensing information
- 2 Why should I license my source code?
- 3 CSDMS suggested licenses
- 4 How to apply the GPL v3 license to your program
 - 4.1 New programs
 - 4.2 Interactive programs
 - 4.3 Signatures
- 5 External links:

Why should I license my source code?

<https://csdms.colorado.edu/wiki/License>

What information would you need for the code to be reusable to yourself?

What you said about FAIR



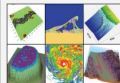
What is "ideal" FAIR - can you give examples? and compare with the usual?

The CSDMS role in making models FAIR

CSDMS Model repository

- [Watch this movie](#) to learn about the full functionality of the model repository.
- And [here are some options](#) if you would like to publish your code.

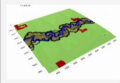
All models



Number of models: **226**, Number of tools: **88**, WMT compliant: **35**, PyMT compliant: **7**

Type: All of the models mentioned below

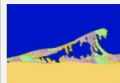
Terrestrial models



Number of models: **103**, Number of tools: **78**, WMT compliant: **10**, PyMT compliant: **4**

Type: Landscape evolution models, avulsion models, sediment transport models, advection diffusion models, ice sheet evolution models, lithospheric flexure models, groundwater models, surface water-quality models, water balance models, etc.

Coastal models



Number of models: **70**, Number of tools: **8**, WMT compliant: **10**, PyMT compliant: **2**

Type: Coastline evolution models, delta sedimentation models, tidal flat models, storm surge models, plume models, turbidity current models, stratigraphic models, wave refraction models, etc.

The CSDMS role in making models FAIR

Metadata

Summary

Contact

Technical specs

In/Output

Process

Testing

Other

Component info

Also known as

Model type

Modular

Model part of larger framework

Incorporated models or components:

Spatial dimensions

3D

Spatial extent

Model domain

Terrestrial

One-line model description

Landscape Evolution Model

Extended model description

CHILD computes the time evolution of a topographic surface $z(x,y,t)$ by fluvial and hillslope erosion and sediment transport.

Keywords:

landscape evolution, erosion, sediment transport, soil erosion, landform development, drainage basin,

The CSDMS role in making models FAIR

Model info

Authors [\[Collapse\]](#)

[Greg Tucker](#)

Source code [\[Collapse\]](#)

- [Go to external source code site](#)

DOI [\[Collapse\]](#)

- Download CHILD [version: 2010.07.06](#)
Doi: 10.1594/IEDA/100102

Model citations [\[Collapse\]](#)

Citation indices **CHILD**

Citations:	2047
h-index:	21

QR-code [\[Collapse\]](#)



Link to this page

Other models by this author [\[Collapse\]](#)

- [Bedrock Fault Scarp](#)
- [CHILD](#)
- [GOLEM](#)
- [Landlab](#)
- [ModelParameterDictionary](#)

The CSDMS role in making models FAIR

pymt

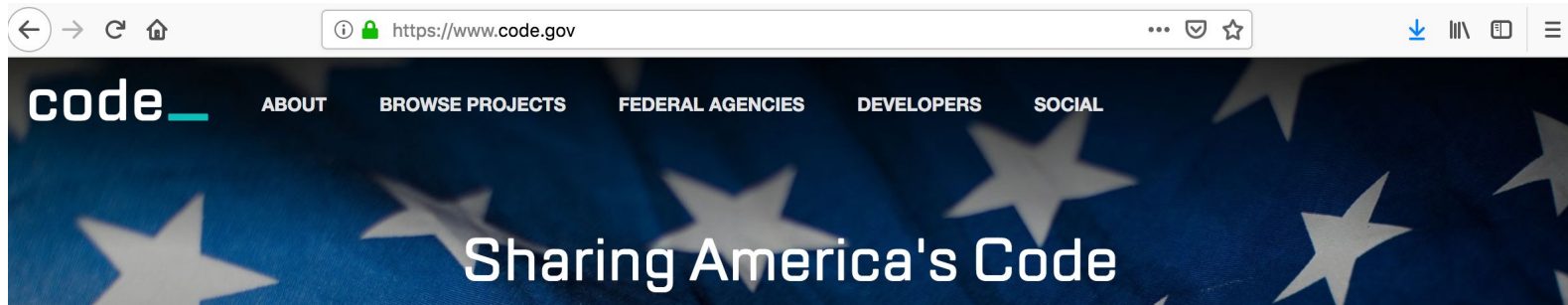
pymt is the *Python Modeling Toolkit*. It is an Open Source Python package, developed by the Community Surface Dynamics Modeling System (CSDMS), that provides the tools needed for coupling models that expose the Basic Model Interface (BMI).

pymt in three points:

- Tools for coupling models of disparate time and space scales
- A collection of Earth-surface models
- Extensible plug-in framework for adding new models

A note about working with Federal partners

If you have collaborators that are employed by Federal agencies, they will likely have stricter requirements for data and software sharing and documentation.



About Us

Code.gov is on track to become the nation's primary platform for sharing America's code. Find out how we'll get there.



Policy Info

The Federal Source Code Policy helps government agencies reduce duplicative software development and save taxpayer dollars.



Progress Report

Government agencies are working hard to unlock the potential of code sharing and collaboration. Here's a quick snapshot of their progress.

A note about working with Federal partners

If you have collaborators that are employed by Federal agencies, they will likely have stricter requirements for data and software sharing and documentation.

Fundamental Science Practices SM Chapters

[SM 502.1](#) - Fundamental Science Practices: Foundation Policy

[SM 502.2](#) - Fundamental Science Practices: Planning and Conducting Data Collection and Research

[SM 502.3](#) - Fundamental Science Practices: Peer Review

[SM 502.4](#) - Fundamental Science Practices: Review, Approval, and Release of Information Products

[SM 502.5](#) - Fundamental Science Practices: Safeguarding Unpublished U.S. Geological Survey Data, Information, and Associated Scientific Materials

[SM 502.6](#) - Fundamental Science Practices: Scientific Data Management

[SM 502.7](#) - Fundamental Science Practices: Metadata for USGS Scientific Information Products Including Data

[SM 502.8](#) - Fundamental Science Practices: Review and Approval of Scientific Data for Release

[SM 502.9](#) - Fundamental Science Practices: Preservation Requirements for Digital Scientific Data

Office of Science Quality and Integrity (OSQI) Instructional Memoranda (IM)

[IM OSQI 2016-01](#) - Review and Approval of Software for Release

[IM OSQI 2018-01](#) – Quality Management System for USGS Laboratories

Review and take-aways

FAIR is not a standard or specification, they are a **guide to assist discovery and reuse by third-parties.**

FAIR as originally conceived put specific emphasis on **machine-actionable** characteristics through open technologies.

FAIR can apply to **data, code, algorithms, workflows.**

By referring to a **few considerations of FAIR in your proposals and procedures**, you can do better than the *status quo*.

Intended outcome of today's presentation

- You are able to explain to someone else what FAIR means for CSDMS-related data
- You know where to go to find out more about FAIR
- You have some strategies for making your proposals and publications FAIR-savvy



Poll 2

If you were to attend a 1.5 hour clinic on FAIR models and data, how would you like to spend your time?

Answer at <https://app2.sli.do/event/twpq1kcu>

Or, use your mobile device browser and go to [slido.com](https://www.slido.com), Event Code #3580

Selected References

Stall, S., E. Robinson, L. Wyborn, L. R. Yarmey, M. A. Parsons, K. Lehnert, J. Cutcher-Gershenfeld, B. Nosek, and B. Hanson (2017), **Enabling FAIR data across the Earth and space sciences**, *Eos*, 98, <https://doi.org/10.1029/2017EO088425>. Published on 08 December 2017.

Stall, S., et al. (2018), **Advancing FAIR data in Earth, space, and environmental science**, *Eos*, 99, <https://doi.org/10.1029/2018EO109301>. Published on 05 November 2018.

Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3 (2016). <https://doi.org/10.1038/sdata.2016.18>

Wilkinson, M. D. et al. **A design framework and exemplar metrics for FAIRness**. *Sci. Data* 5:180118 <http://doi.org/10.1038/sdata.2018.118> (2018).

Other links in this talk

1. <http://www.copdess.org/enabling-fair-data-project/enabling-fair-project-overview/>
2. <http://www.copdess.org/enabling-fair-data-project/enabling-fair-data-faqs/>
3. <http://www.copdess.org/enabling-fair-data-project/commitment-to-enabling-fair-data-in-the-earth-space-and-environmental-sciences/signatories/>
4. <https://www.go-fair.org/>
5. <https://www.go-fair.org/fair-principles/>
6. AGU Earth and Space Science Informatics, activities and sessions at the Fall meeting: <https://essi.agu.org/>
7. GSA Geoinformatics: <https://community.geosociety.org/geoinformaticsdivision/home>
8. Earth Science Information Partners: <http://www.esipfed.org>
9. USGS Community for Data Integration: <https://www.usgs.gov/centers/cdi>
10. https://figshare.com/articles/An_Introduction_to_Software_Licensing/7409573
11. <https://csdms.colorado.edu/wiki/License>
12. <https://softwaresaved.github.io/software-licensing-workshop/#/>
13. https://csdms.colorado.edu/wiki/Model_download_portal
14. <https://pymt.readthedocs.io/en/latest/>
15. <https://www.code.gov>
16. <https://www.usgs.gov/about/organization/science-support/science-quality-and-integrity/fsp-policy-directives>
17. <https://www.go-fair.org/faq/ask-question-difference-fair-data-open-data/>

Poll 1

On a scale of 1-5, how comfortable are you with explaining what FAIR means for your models and data? (5 is most comfortable)

Answer at <https://app2.sli.do/event/twpq1kcu>

Or, use your mobile device browser and go to [slido.com](https://www.slido.com), Event Code #3580

Extra Poll

True or False: FAIR models and open source models are basically the same thing.

Answer at <https://app2.sli.do/event/twpq1kcu>

Or, use your mobile device browser and go to slido.com, Event Code #3580