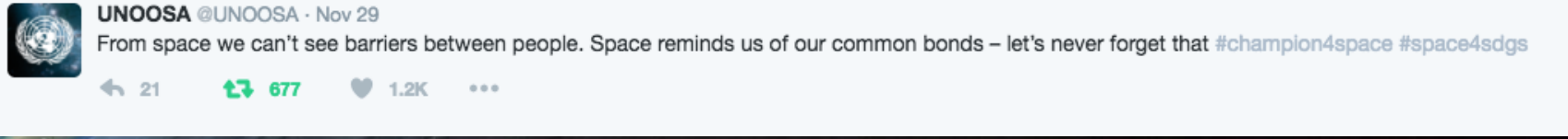
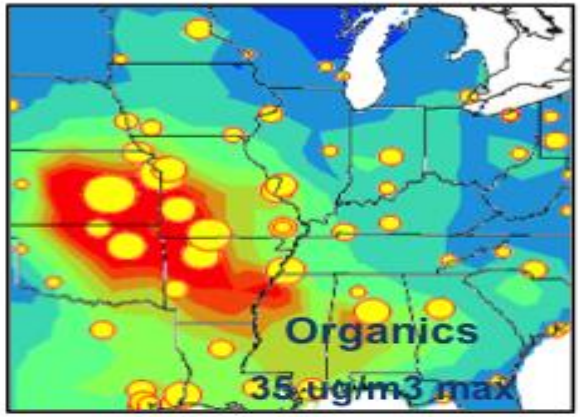
Data + Community = Action

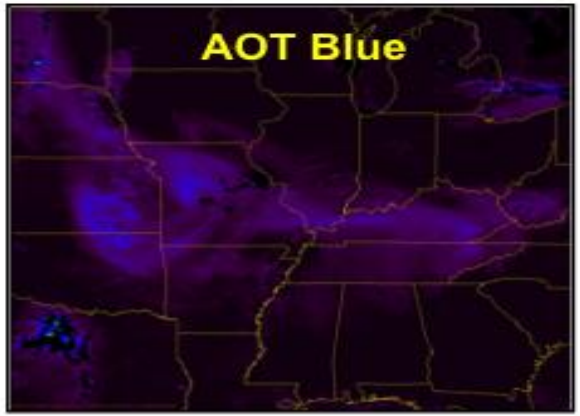
Erin Robinson,
Executive Director
ESIP



https://twitter.com/UNOOSA/status/803692453358424064

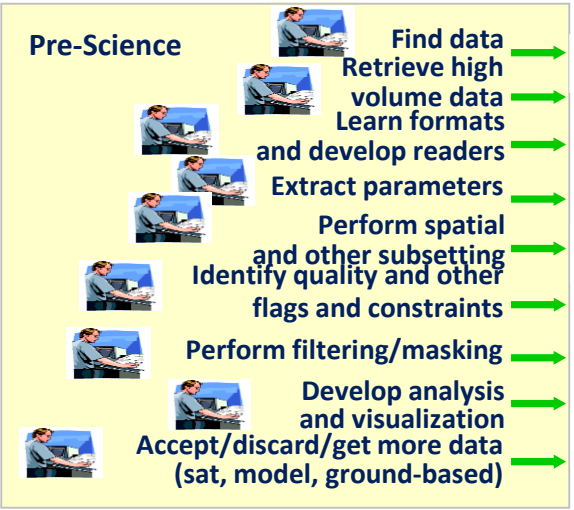
Kansas Agricultural Smoke, April 12, 2003





Making Data Matter

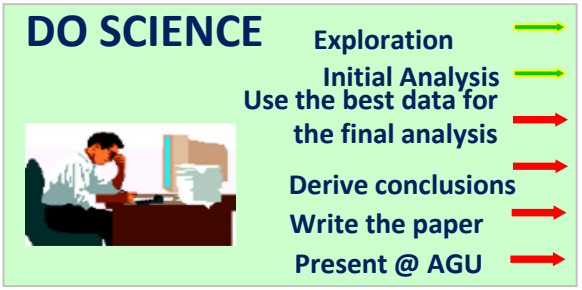
Old Way



Jan

Mar

Jun



Sept

Dec

[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;*

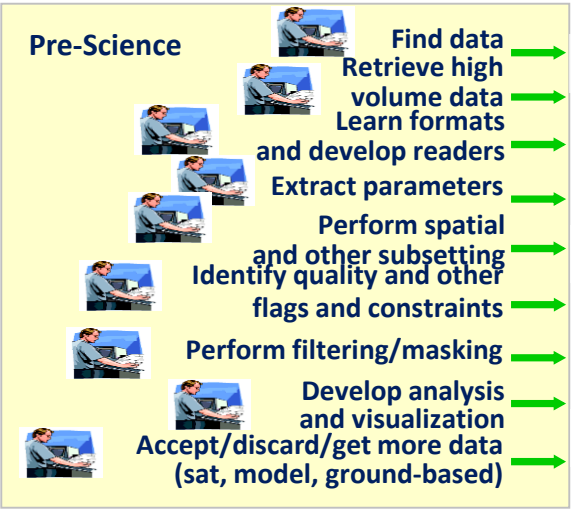
*if he finds them good, he can not **merge** them with other data"*

Information Technology and the Conduct of Research: The User's View (1989)

Adapted from Leptoukh, 2012

Making Data Matter

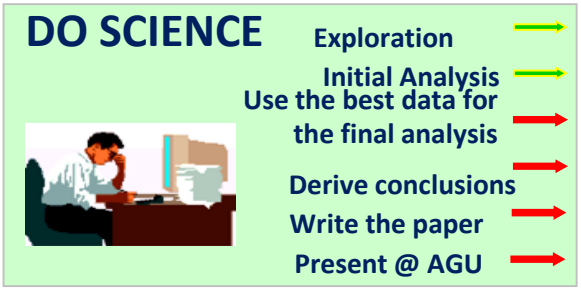
Old Way



Jan

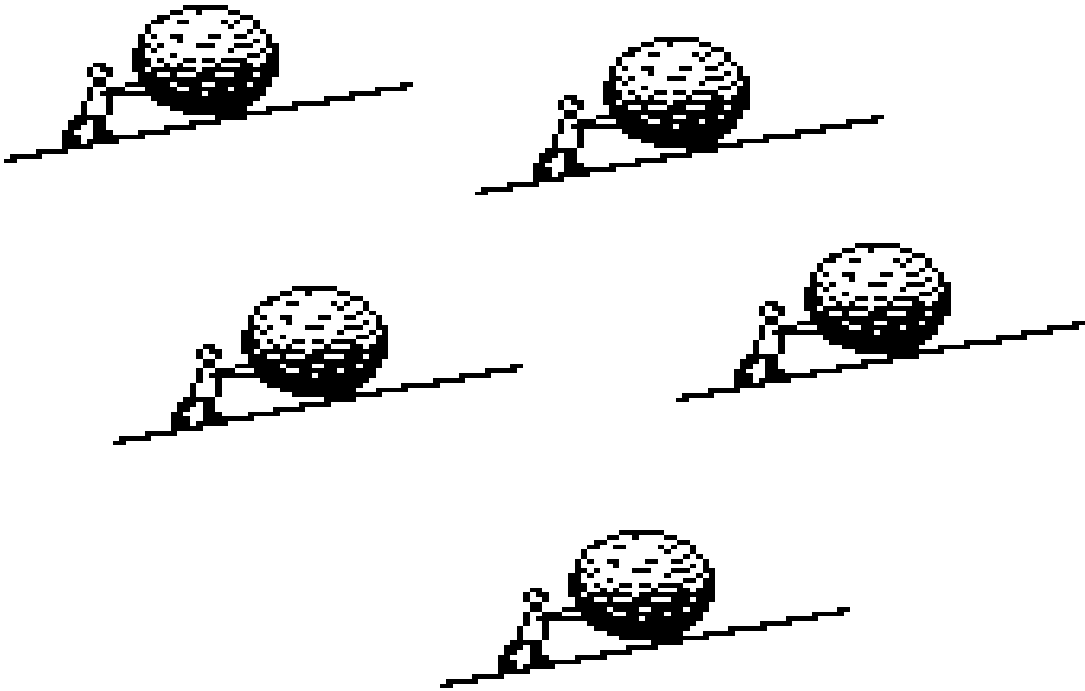
Mar

Jun



Sept

Dec



Adapted from Leptoukh, 2012

FAIR Guiding Principles

FAIR is…

Findable

Accessible

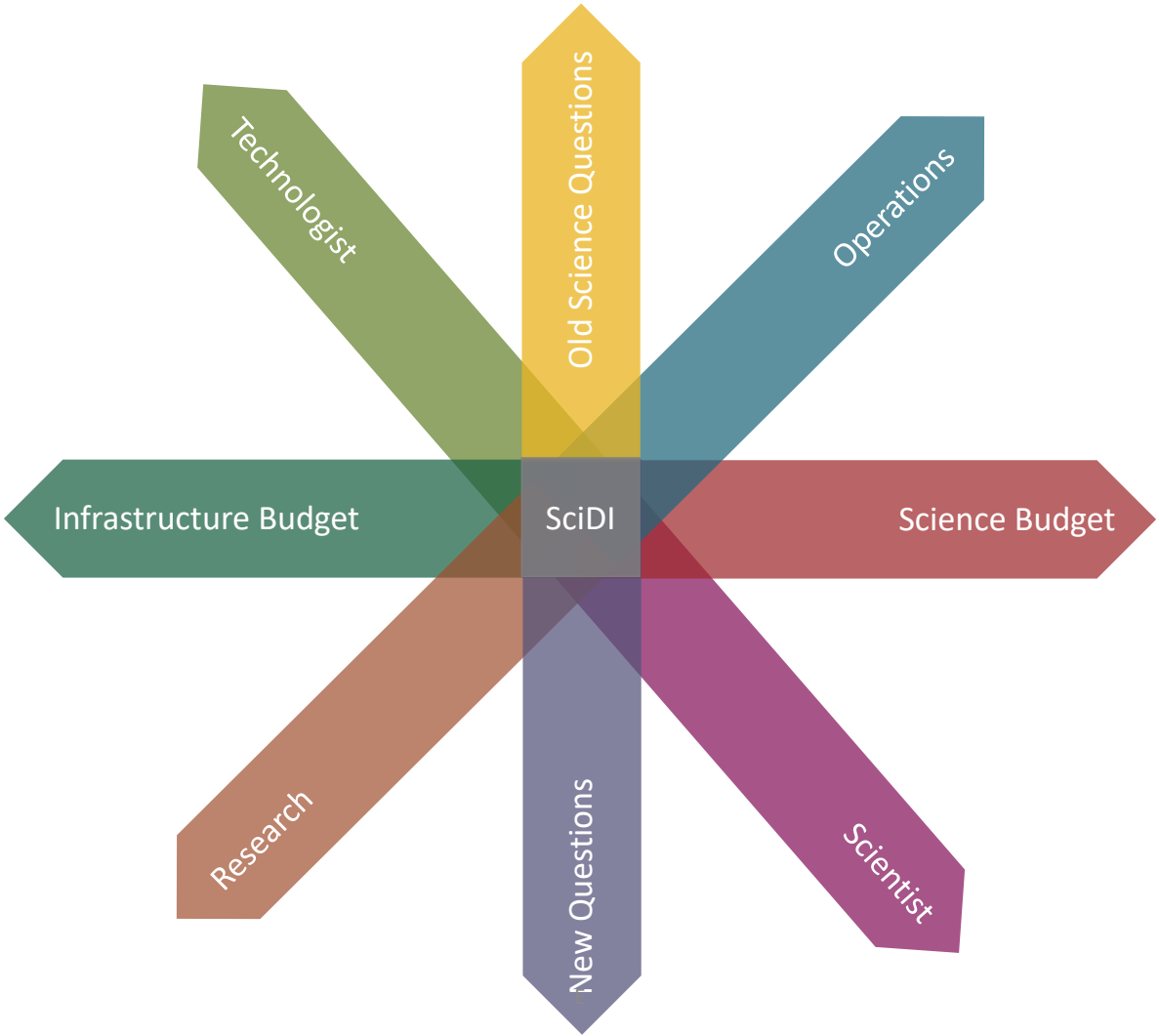
Interoperable

Reusable

Article in Nature journal *Scientific Data*: Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).

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*


Technologist
Old Science Questions
Operations
Infrastructure Budget
SciDI
Science Budget
Research
New Questions
Scientist

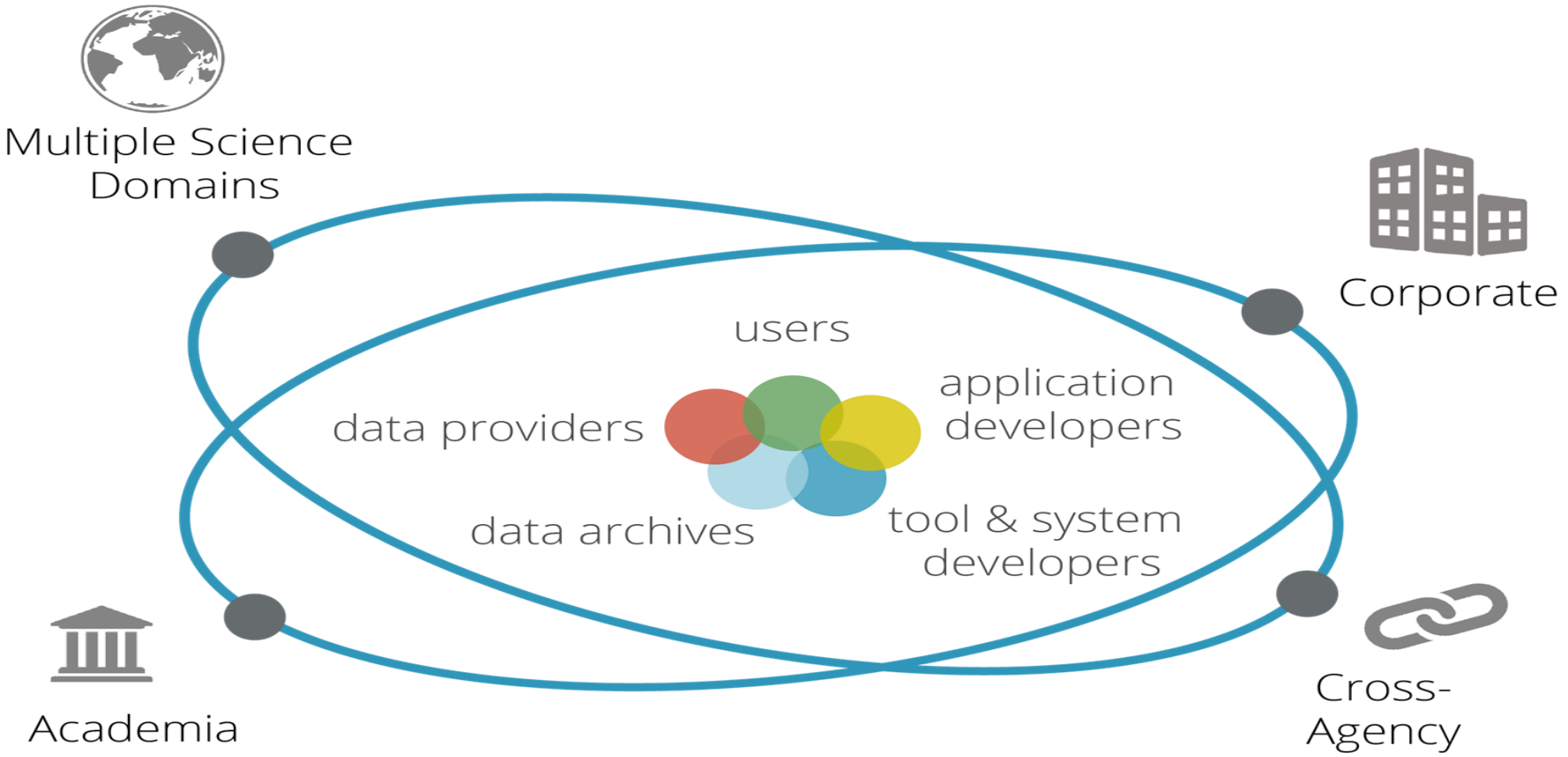
There is an urgent need to improve the **[Global Collaborative]** infrastructure supporting the **(re)**use of scholarly data.

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

ESIP Vision

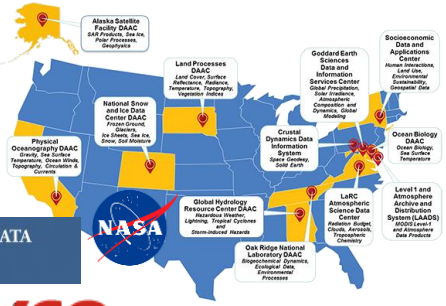
Leaders in promoting
*the **collection, stewardship and (re)use***
Of Earth science data, information and knowledge
that is responsive to societal needs.

ESIP COMMUNITY



Types of ESIP Members

Type I: Data Centers



NCEI
NOAA NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION

COUNCIL OF DATA FACILITIES

UNAVCO

GDC
Geological Data Center

BCO-DMO
Biological & Chemical Oceanography Data Management Office

Type II: Researchers and Tool Developers

CSIRO

IMOS

AuScope

NCI

SGCI
Science Gateways Community Institute

Type III: Application Developers

esri

UNM University Libraries

Kitware

Element 84

Type IV: Strategic Partners

NASA

NOAA
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
US DEPARTMENT OF COMMERCE

USGS
science for a changing world

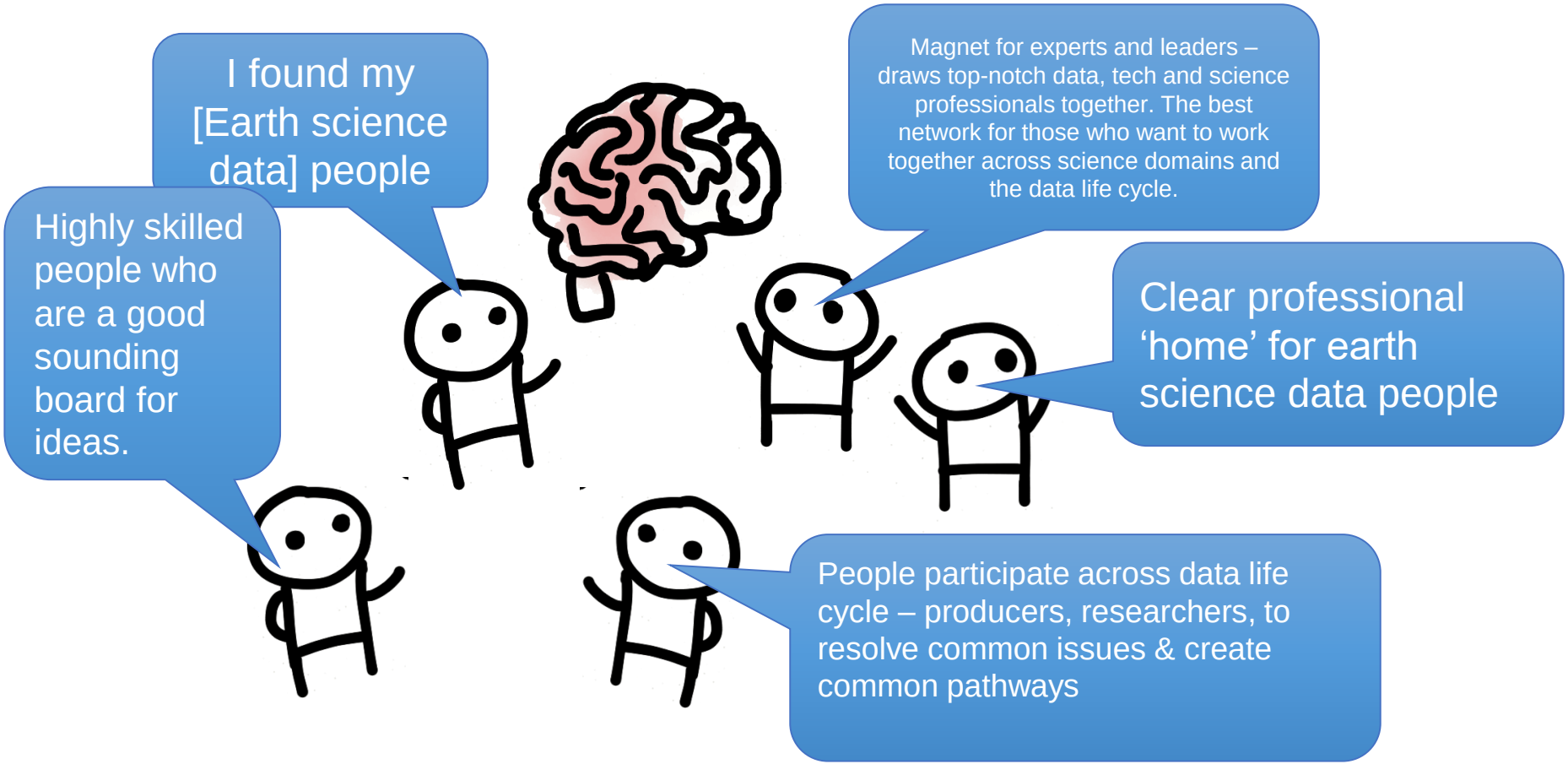
MOU Collaborators

OGC®
Making location count.

AGU 100
ADVANCING EARTH AND SPACE SCIENCE

And many more! ESIP has 110+ Members

What ESIP community says..



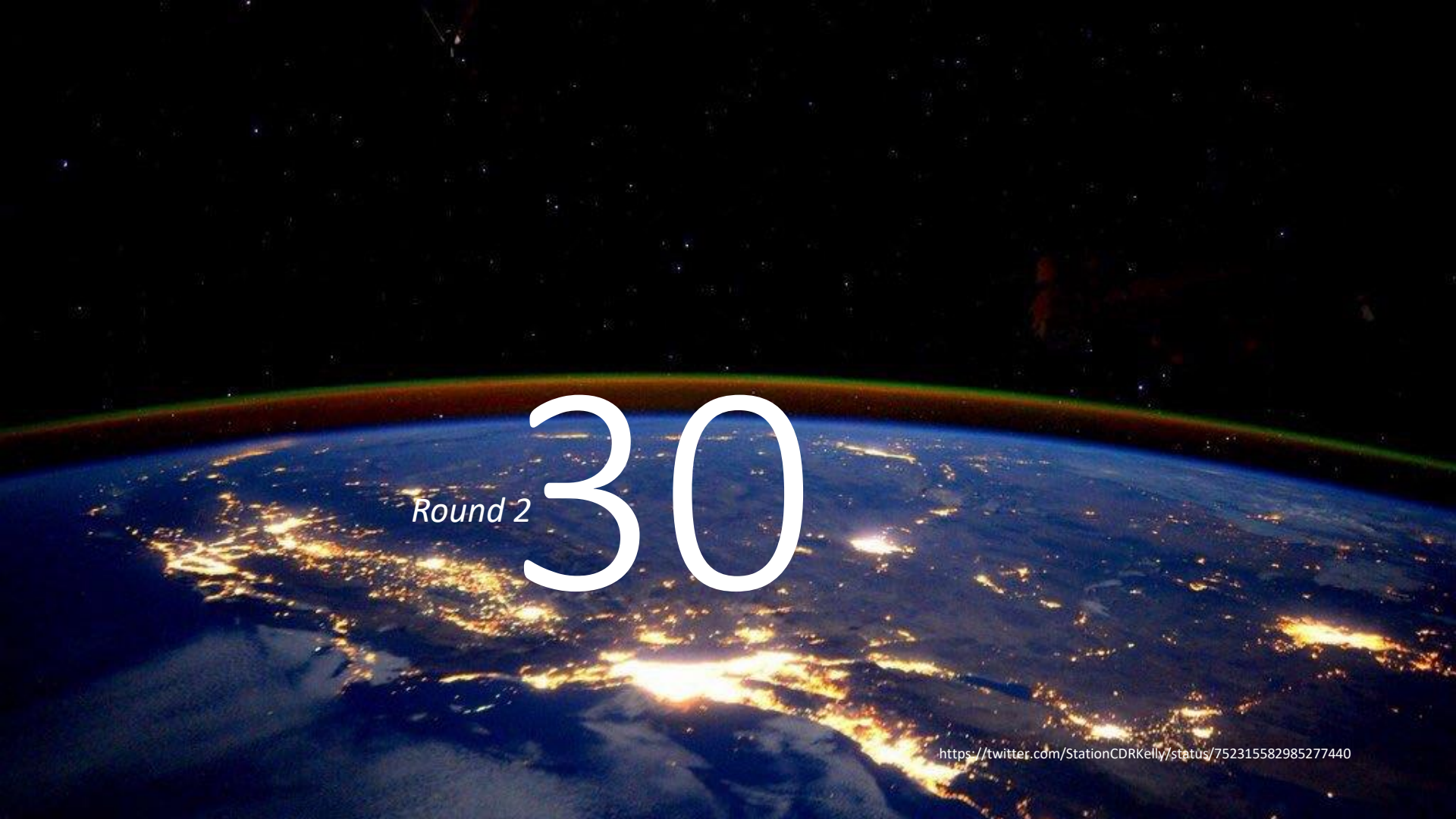
https://i.ytimg.com/vi/s6dYLtxHVZg/maxresdefault.jpg


Round 1
30
https://twitter.com/StationCDRKelly/status/662269393515204608

Round 1

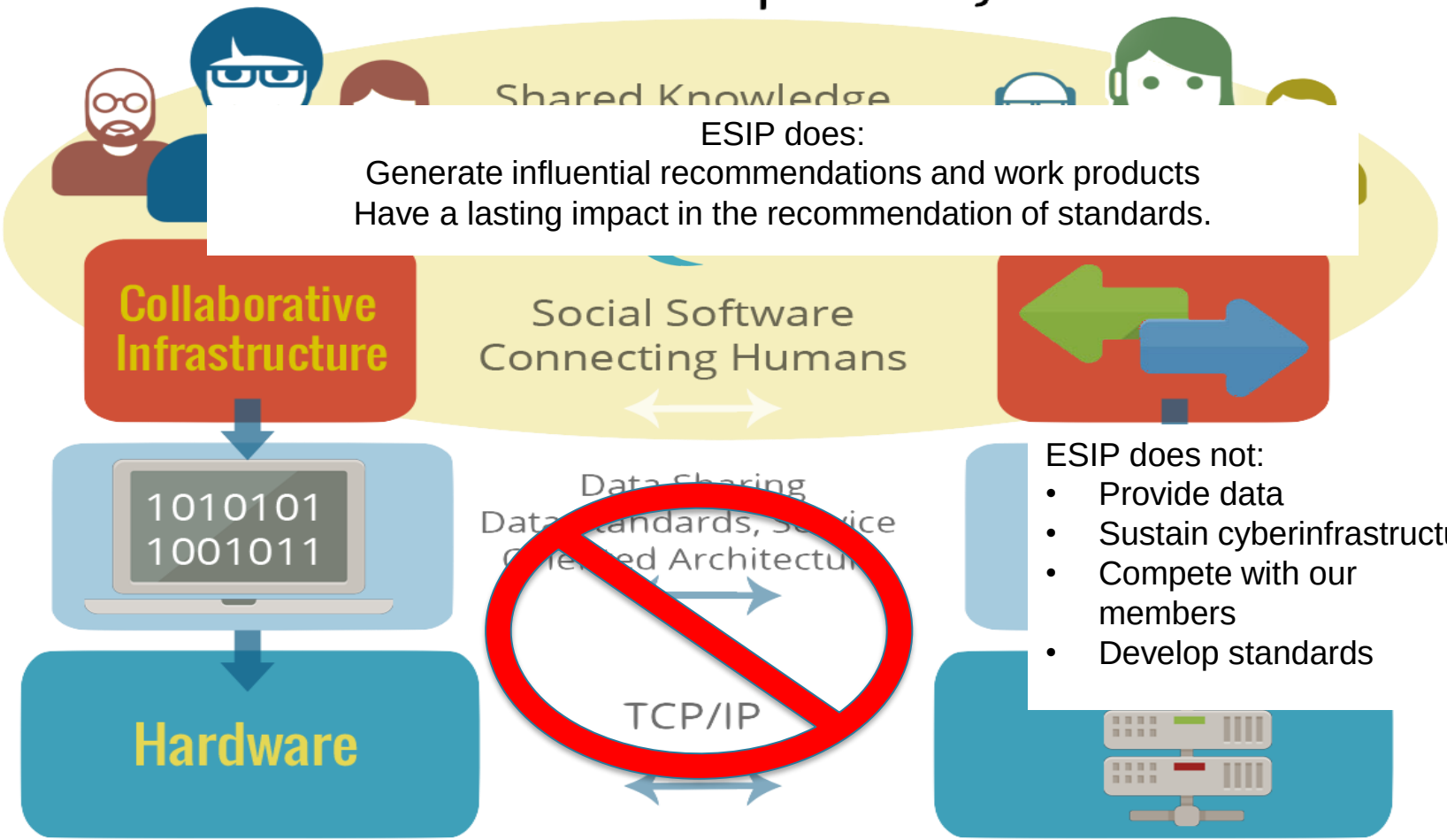
15

https://twitter.com/StationCDRKelly/status/751479254404198400


Round 2
30
https://twitter.com/StationCDRKelly/status/752315582985277440


Round 2
15
https://twitter.com/StationCDRKelly/status/802301531588476928

Information Interoperability Stack



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

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:
ESIP Software and Services Citation Cluster. (2019). Software and Services Citation G
and Examples. Ver. 1. ESIP. https://doi.org/10.6084/m9.figshare.7640426.
Figshare link:
https://esip.figshare.com/articles/Software_and_Services_Citation_Guidelines_and_E
640426

Document Status

Approved by ESIP Assembly Meeting 16 January 2019

Data Citation Guidelines for Earth Science Data Version 2

Suggested Citation:

ESIP Data Preservation and Stewardship Committee. 2019. *Data Citation Guidelines for Earth Science Data. Ver. 2.* Earth Science Information Partners. https://doi.org/10.....

Table of Contents

Resource Object Citation Cluster

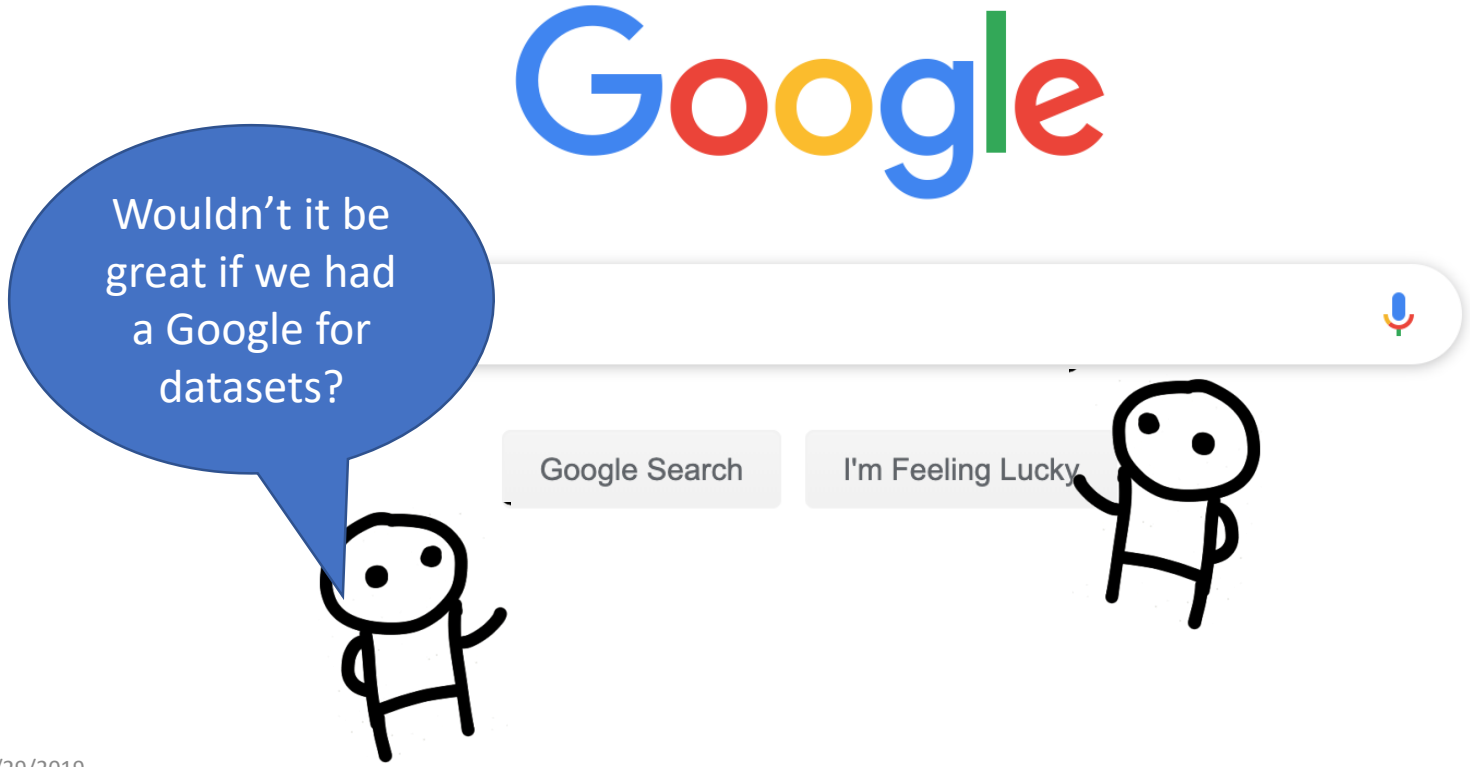
- New cluster spun of the Data Stewardship Committee and in combination with the Software and Services Citation Cluster
- First task: update data citation guidelines.
  - Refinement core concepts and issues and mapped concepts to more metadata dialects
  - New guidance on “Dynamic data citation” notably 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.

Data + Community = Action

This is important because:

- evidence suggests that citing data in related publications increases **the citation rate of those publications**
- routine citation of data **acknowledges data as a first class research output** and facilitates reproducible and transparent research
- citations for published data can be included in CVs and biographical sketches along with journal articles, reports and conference papers
- only cited data can be counted and tracked (in a similar manner to journal articles) **to measure impact.**

Source: https://www.ands.org.au/working-with-data/citation-and-identifiers/data-citation


Google
Wouldn’t it be great if we had a Google for datasets?
Google Search
I'm Feeling Lucky

Dataset Discovery

Google dataset discovery

'Facilitating Dataset Discovery using new developments within schema.org' @ Summer ESIP 2013, Peter Fox & Doug Newman

7/2013

ESDSWG Search Relevance WG

Continued to discuss and move forward

Chris Lynnes, Lewis McGibbney,

2015

2014

NASA RDFa Mark-up carried over from ECHO to CMR

2016

ESIP Semantic Tech
Provides guidance for publishing schema.org as JSON-LD for the sciences

Beth Huffer, Doug Fils, Lewis,

P418

Separate NSF-related work for data facilities to implement dataset schema.

Doug Fils, Adam Shepherd, Eric Lingerfelt

2017

1/2018

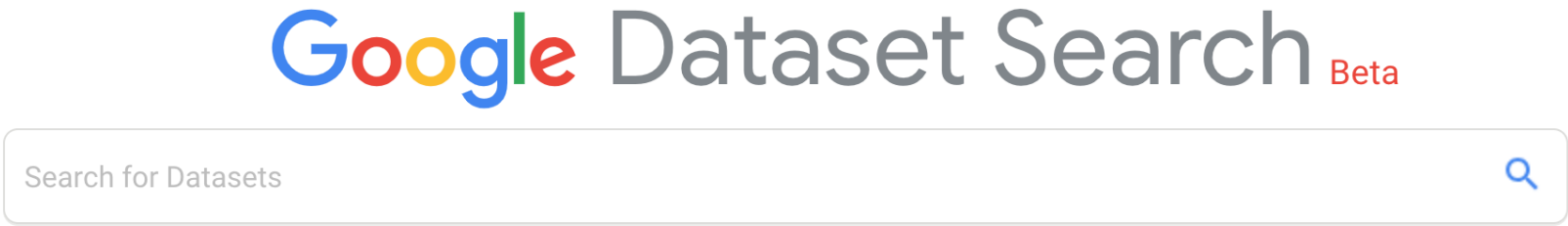
Google Dataset Search

Finally Google is utilizing Dataset schema for search; Google showed up at ESIP Winter Meeting 2018

Google Dataset Search is finally live

9/2018

Source: Doug Newman, ESIP Winter 2019 https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20190000098.pdf

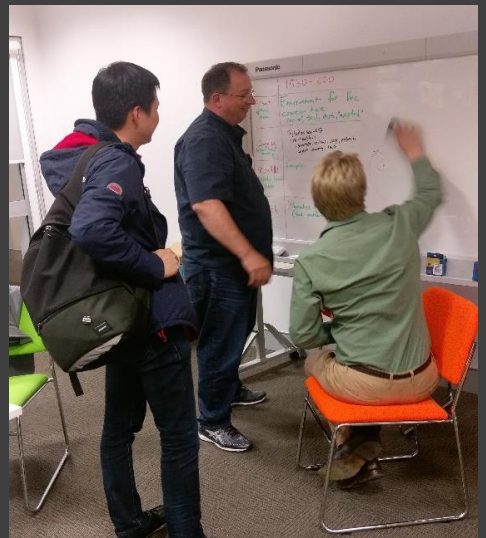


Data + Community = Action

This is important because:

- Community was ready
- Implementation of Dataset family of Schema.org on dataset landing pages was do-able
- Google provided validation & testing tools and community improved

TBD: Is this how scientist will actually find data?

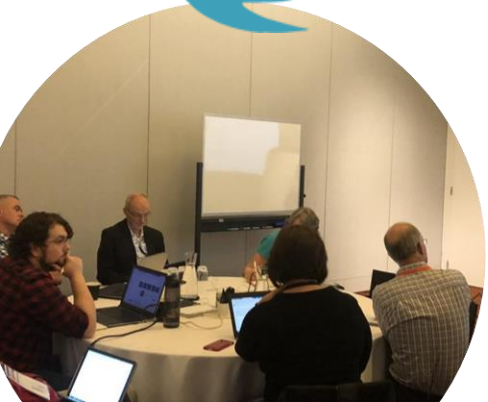


2017 Cutting edge science symposium – Linking environmental data & samples

- Canberra, May 2017 – 3 days
- ~70 participants, 18 international, **ESIP co-chair**
- Geoscience, marine, ecosystems, biodiversity, climate
- Field trip, presentations, un-conference (breakouts), anti-conference, up-goer challenge

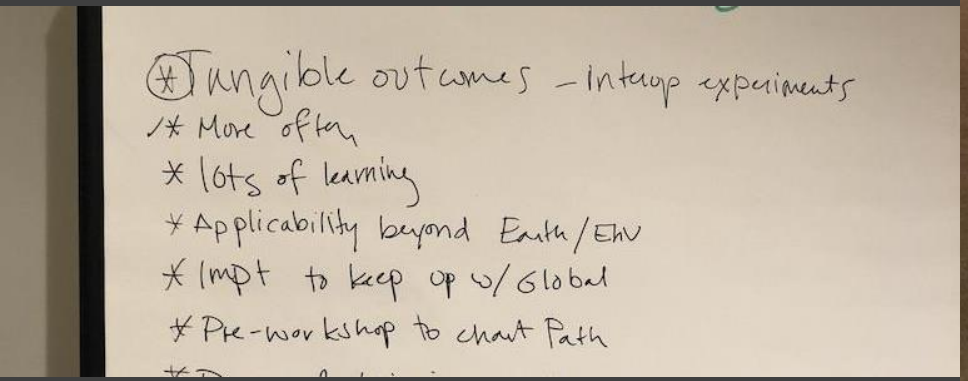
2018 E2SIP @ C3DIS

- Pre-conference workshop - http://www.c3dis.com/1725 - *09:00–18:00*
  - ~40 in the rooms (6 international), ~5 remote (2 international)
  - ESIP on organizing committee
  - 6 NCRIS facilities + CSIRO + GA + CRCSI + 2 state govts + …



6 breakouts

- Vocabularies & vocabulary services
- FAIR publishing
- FAIR services & software
- Big data/analytics
- Cloud platform for X-domain science
- Drones



Do the participants want more of this?

Next E2SIP meeting is Thursday &
More on this work in Collab Track this afternoon

2018 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**
- ORCID
- CrossRef
- CHORUS
- Scholix
- OSGeo
- Pangaea
- DataONE

And Growing!!

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

AGU 100
ADVANCING EARTH AND SPACE SCIENCE

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.

Developed by: DataCite/re3data https://repositoryfinder.datacite.org

Repository Finder

Find a repository to upload your data.

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

Type to search... Search

or

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

ESIP Data Management Training

EXPLORE

Home Browse Search Submit Help About Log in

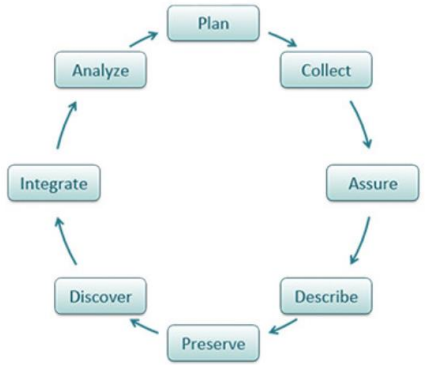
Welcome to the DMT Clearinghouse

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

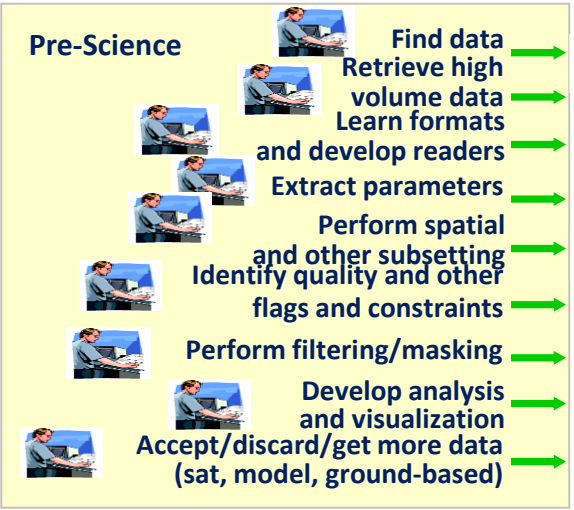


DataONE Life Cycle - https://www.dataone.org/data-life-cycle

http://dmtclearinghouse.esipfed.org

Making Data Matter

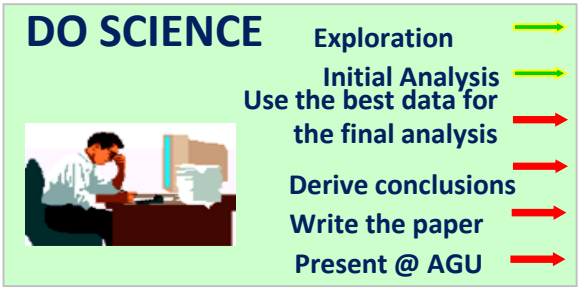
Old Way



Jan

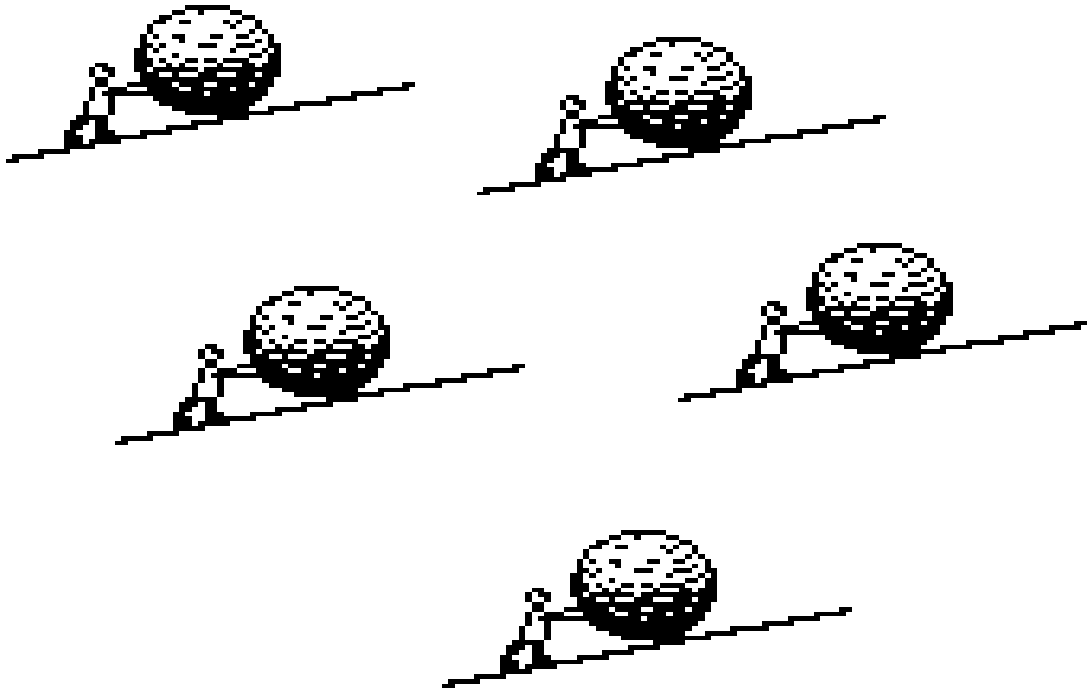
Mar

Jun



Sept

Dec



Adapted from Leptoukh, 2012

Making Data Matter Together

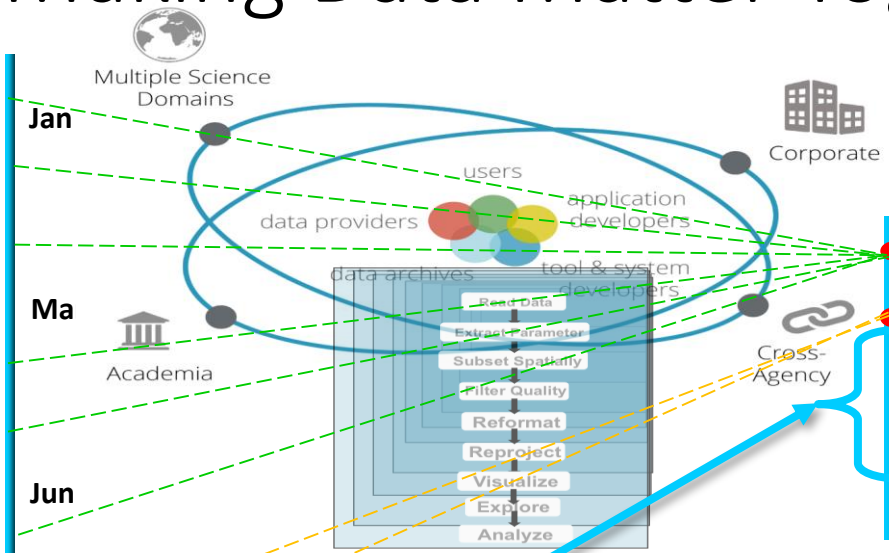
Old Way

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

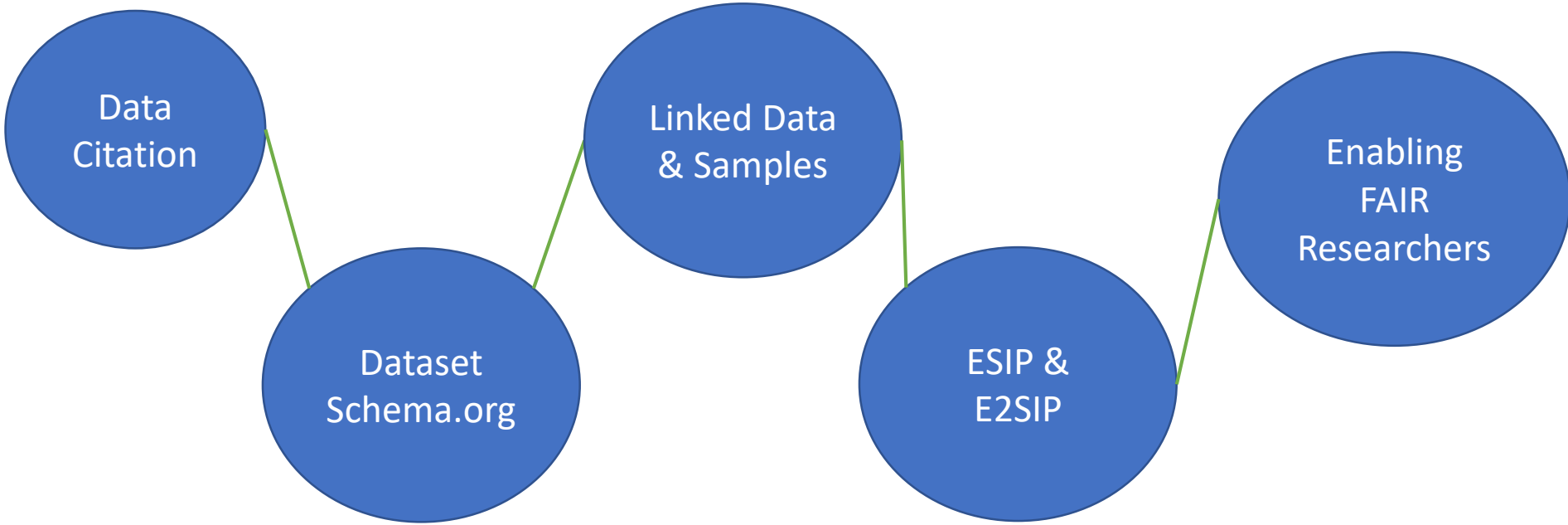


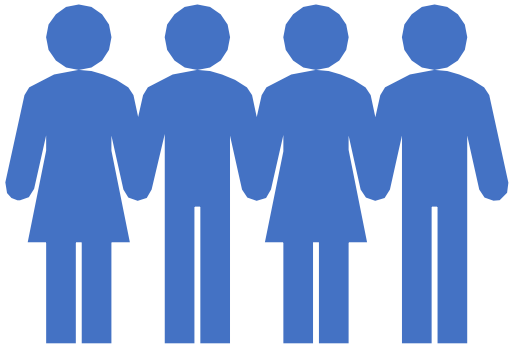
New Way

- Minutes
- Days for exploration
- Use the best data for the final analysis
- Derive conclusions
- Work with repository to manage data
- Publish data & code
- Write the paper & *cite artifacts*
- Submit the paper

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

"You can't connect the dots looking forward."
- Steve Jobs





Acknowledgements

- Peter Fox, RPI
- Simon Cox, Jens Klump, CSIRO
- Lesley Wyborn, ARDC/NCI
- Ben Evans, NCI
- Irina Bastrakova, Geoscience Australia
- Adrian Burton, Natasha Simons, Mingfang Wu, ARDC
- Helen Glaves, BGS
- ESIP Funders: NASA, NOAA & USGS
- ESIP US Community