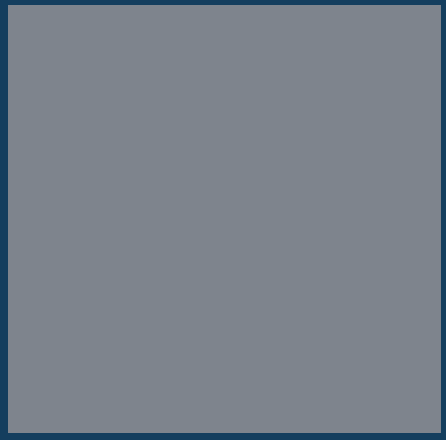


Open Science: How do we do it?

Andy Barrett and Walt Meier

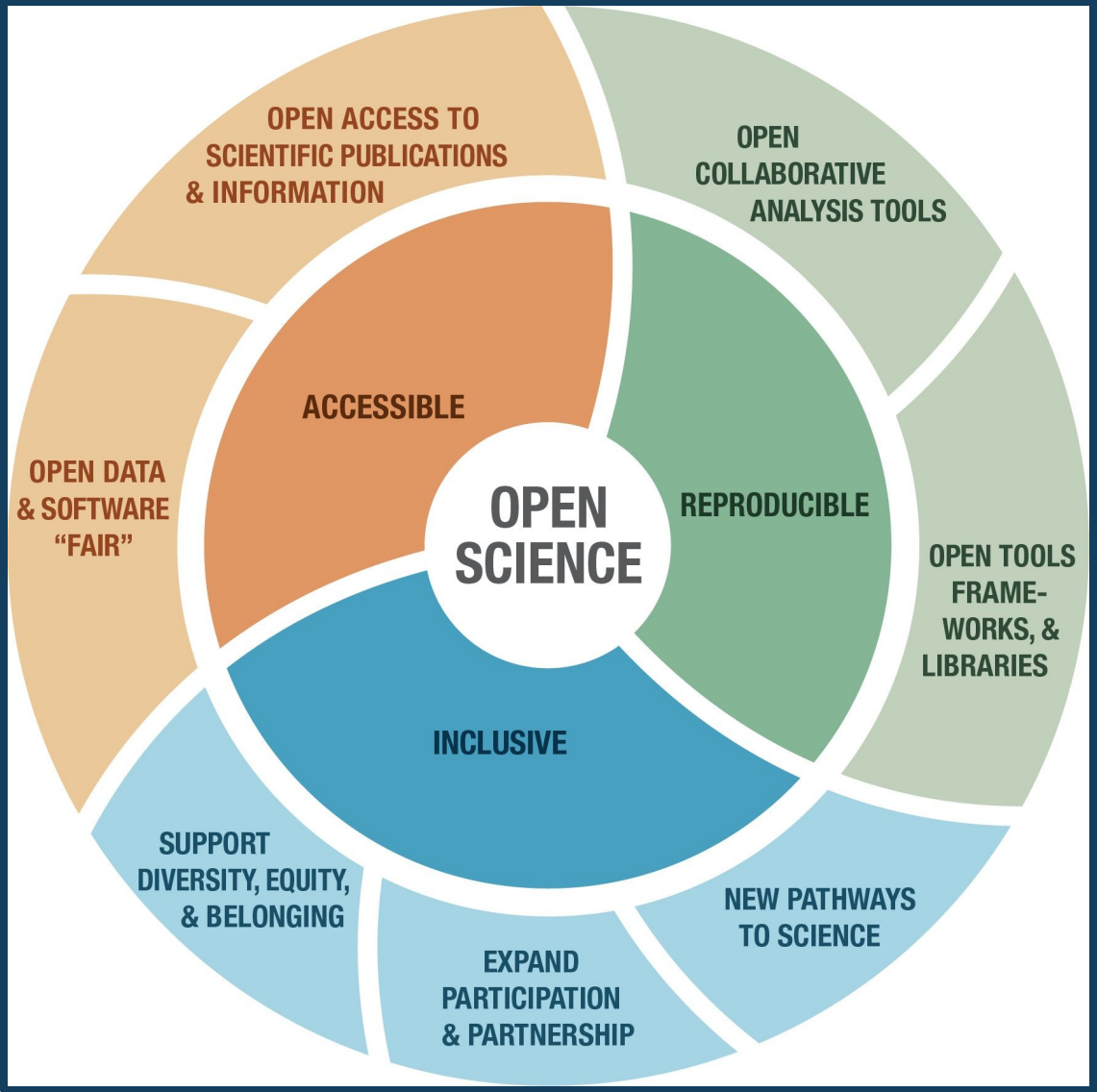
NSIDC National Snow and Ice Data Center

Advancing knowledge of Earth's frozen regions

CIRES

What is Open Science

- Open Science is the opening-up of the scientific process from “ideation” to results.
- Open sharing of results, data and code, and other parts of the process makes science:
  - transparent and reproducible;
  - increases accessibility and participation.



https://doi.org/10.1029/2020EA001562

Open Science is not just an aspiration…

2023 is the Federal Year of Open Science and NASA's Year of Open Science.

Results of Federally funded research now must be open.

New NASA missions and research awards must comply with NASA's Open Science Policy SPD-41a.

Outline

- Overview of open science initiatives.
- How we can meet policy requirements.
- Thoughts on implementing Open Science.

U.S. Open Science Efforts

- Year of Open Science
- White House OSTP memorandum
- NASA SPD-41a
- NASA OSSI
- NASA TOPS
- NSF Geosciences Open Science Ecosystem
  - https://beta.nsf.gov/funding/opportunities/geosciences-open-science-ecosystem-geo-ose
- NOAA/NCAR Open Hackathon
  - https://www2.cisl.ucar.edu/events/ncarnoaa-open-hackathon

NSIDC

Year of Open Science

A personal take on science and society

World view

By Chelle Gentemann

Why 2023 is the US Year of Open Science

Here's how NASA is incentivizing open science, and how you can too.

I'm thrilled to be the Transform to Open Science lead for NASA, which has a 60-year legacy of pushing the limits of how science is used to understand the Universe, planetary systems and life on Earth. Much of NASA's success can be attributed to a culture of openness for the public good. Since the 1990s, the agency has been a leading advocate for full and open access to data and algorithms.

That culture is needed now more than ever. Humanity is facing many intersecting challenges, from the COVID-19 pandemic to climate change and food and water insecurity. To combat them, we must find breakthroughs faster,

> "Active inclusion of diverse groups of people is a crucial element of truly open science."

Science isn't happening just at NASA. But agency partnerships take years to negotiate. I needed a side door. Knowing this, NASA invited me to become one of their representatives to the National Science and Technology Council Subcommittee on Open Science, which coordinates and advances open science across federal agencies. There, I helped to catalyse the Year of Open Science. First, we agreed on a definition: open science is the principle and practice of making research products and processes available to all, while respecting diverse cultures, maintaining security and privacy, and fostering collaborations, reproducibility and equity. Next, we set four goals for each agency involved in the Year of Open Science: to develop a strategic plan for open science; improve the transparency and equity of reviews; account for open-science activities

Chelle Gentemann, NASA TOPS mission lead

Gentemann, C., *Nature*, 13 January 2023
https://doi.org/10.1038/d41586-023-00019-y

OSTP Memorandum

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

August 25, 2022

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Dr. Alondra Nelson
Deputy Assistant to the President and Deputy Director for Science and Society
Performing the Duties of Director
Office of Science and Technology Policy (OSTP)

SUBJECT: Ensuring Free, Immediate, and Equitable Access to Federally Funded Research

This memorandum provides policy guidance to federal agencies with research and development expenditures on updating their public access policies. In accordance with this memorandum, OSTP recommends that federal agencies, to the extent consistent with applicable law:

1. Update their public access policies as soon as possible, and no later than December 31st, 2025, to make publications and their supporting data resulting from federally funded research publicly accessible without an embargo on their free and public release;
2. Establish transparent procedures that ensure scientific and research integrity is maintained in public access policies; and,
3. Coordinate with OSTP to ensure equitable delivery of federally funded research results and data.

https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf

NASA SPD-41a

Science Mission Directorate (SMD) Policy Document 41a

- Peer-reviewed publications openly available without embargo period
- Research data and software shared at time of publication or the end of the funding award
- Mission data released asap; unrestricted mission software developed openly
- Science workshops and meetings held openly → enable broad participation

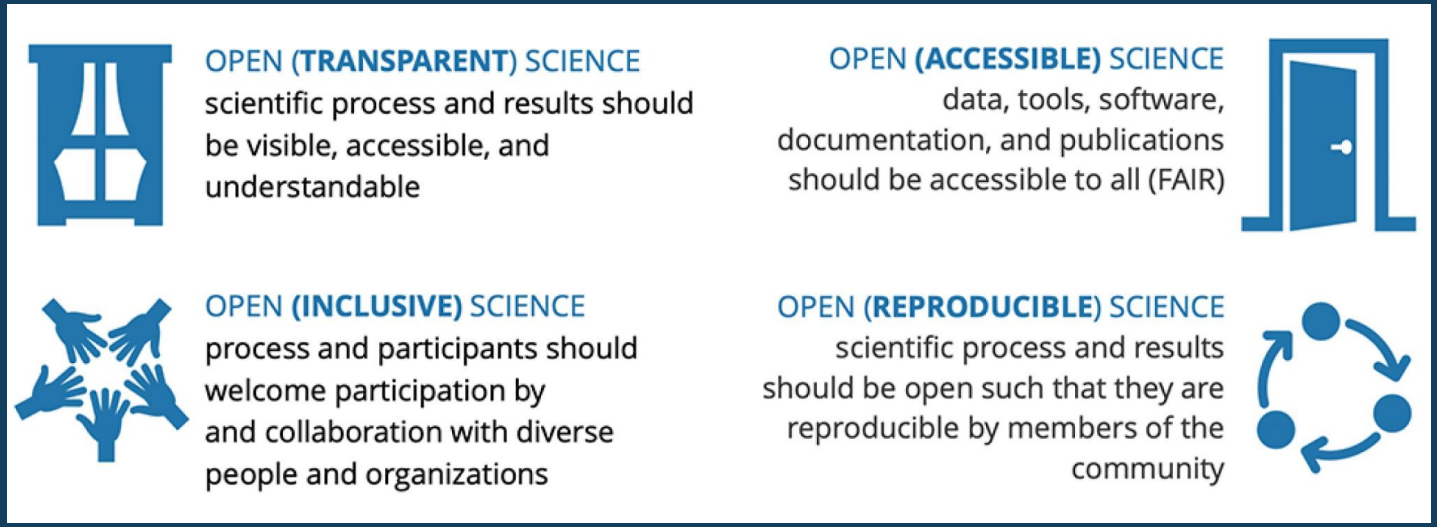
NSIDC

https://science.nasa.gov/science-red/s3fs-public/atoms/files/SMD-information-policy-SPD-41a.pdf

NASA OSSI

Open Source Science Initiative

Aims to implement NASA's "Strategy for Data Management for Groundbreaking Science 2019-2024"



https://science.nasa.gov/open-science-overview

NSIDC

NASA TOPS
Transform to OPen Science

- 5-year program to engage scientists and aid transition of NASA research to open science
- Goal is to get 20,000 scientists certified for open science
- Areas of action:
  - *Engagement:* community building - forums, panels, conferences, publications
  - *Capacity Sharing:* curriculum, workshops, hackathons
  - *Incentives:* open science badges
  - *Moving Toward Openness*

NSIDC

https://nasa.github.io/Transform-to-Open-Science/

TOPS Strategic Objectives

1. Accelerate major scientific discoveries
2. Broaden participation by historically excluded communities
3. Increase understanding and adoption of open science principles and techniques

NSIDC

https://nasa.github.io/Transform-to-Open-Science/

TOPS Activities

- AGU Fall 2022 - official “kick off”
- AMS 2023 - first training sessions
- Training: OpenCore, ScienceCore
- ROSES TOPS Training RFP (ROSES 2022 F.14) - proposals under review
  - More training calls in future years

NSIDC

https://nasa.github.io/Transform-to-Open-Science/

TOPS OpenCore

- Basics of open science, developing workflows
- 5 modules, 2.5 hours each
  a. Ethos of Open Science
  b. Open Tools and Resources
  c. Open Data
  d. Open Software
  e. Open Results

NSIDC

https://nasa.github.io/Transform-to-Open-Science/

TOPS Resources

- https://nasa.github.io/Transform-to-Open-Science
- https://github.com/nasa/Transform-to-Open-Science
- Presentation: https://doi.org/10.5281/zenodo.5621674

NSIDC

How do we do Open Science
Some guiding principles

- There are many ways to do Open Science.
- Open Science is kind science.
- Remember: Accessible, Reproducible, Inclusive.
- Follow FAIR principles.
- Open Science benefits you as well..

Making Publications Open

Several approaches

- Publish in an Open Access Journal
- Use institutional archive (embargo period)
- Use a pre-print server. https://essopenarchive.org/, https://eartharxiv.org/
- Use https://scholar.colorado.edu/

NSIDC

Making Data Open and FAIR

Submit data to long-lived, publicly accessible archives.

Data is **F**indable and **A**ccessible.

Archives

- Assign persistent identifiers: DOI
- Help create metadata.
- Ensure data is **I**nteroperable and **R**eusable

NSIDC

Putting the IR in FAIR

Interoperability

- Use standard formats:
  - NetCDF, GeoTiff, CSV, HDF5
- Follow metadata conventions
  - CF Conventions
  - OGC

Reusable: Make sure data and metadata are accurate and fully documented, and licensed.

NSIDC

Making code open

Software “documents” analysis as well as allowing reproduction of results.

Code can be archived in the same repositories as data.

Github is a collaborative tool not an archive.

Archive the version of code used for the publication.

Always add a permissive license.

NSIDC

Most scientists aren't programmers

Focus on **clean**, understandable code.

- Use comments.
- Descriptive variable and function names
  - e.g. `air temperature` not `t`
- Prefer widely used packages and libraries to writing your own routines.
- Prefer non-proprietary tools (Python, R, NCL, Julia).
- …but clean, understandable Matlab and IDL are better than nothing.

NSIDC

Mindset, tools and workflows

Practice Open Science from the start of a project, not just at the end.

Overcome the “hero” mindset. Open Science is about collaboration.

Share tools, tips and “tricks” with your group.

Don’t duplicate work. Use collaborative tools.

- GitHub for version control of code and documents.
- Shared file systems for data.
- Document workflows

NSIDC

Implementation: some simple steps

1.	Shortlist possible practices for your project (you can't do everything!)
2.	Consider what you can adopt yourself versus what you group needs to adopt
3.	Talk to your research group about what to implement
4.	Create an implementation plan
5.	Be prepared to reassess and adapt; what is working and not working?
6.	Document and share best practices

From: Heise V, Holman C, Lo H, Lyras EM, Adkins MC, Aquino MRJ, et al. (2023) Ten simple rules for implementing open and reproducible research practices after attending a training course. PLoS Comput Biol 19(1): e1010750. https://doi.org/10.1371/journal.pcbi.1010750

NSIDC

Next Steps…

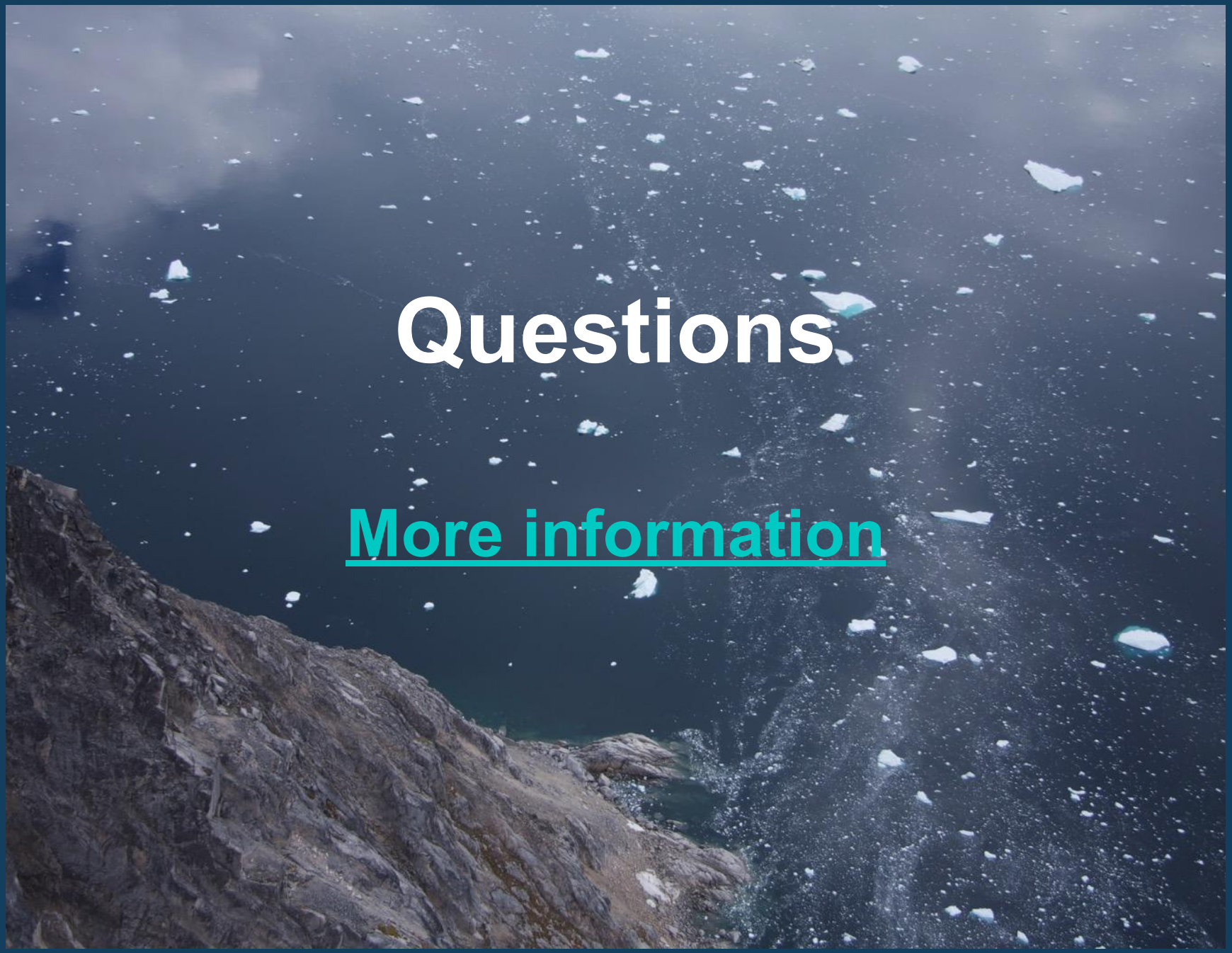
Get an ORCID orcid.org

Get a GitHub account github.org

Register for the TOPS OpenCore Curriculum

Identify a new skill to learn: maybe Github or Python, R or some other non-proprietary language.

NSIDC


Questions
More information


Extra Slides

The cost of Open Science

- Maintenance of repositories
- Learning new tools
- It takes time to prepare data for curation and sharing
- Writing “clean code” takes time
- Open Access publication costs

Be kind, keep tabular data Tidy

- Each row is an observation
- Each column is a variable
- A table has one header row

Standard tools in many languages can read this format easily shortening time to analysis

Excel worksheets with fancy formatting look nice but are not easily read by other tools.

Wickham, H. . (2014). Tidy Data. *Journal of Statistical Software*, *59*(10), 1–23. https://doi.org/10.18637/jss.v059.i10

	temperature	humidity	wind_speed
Ipswich	3.0	85	5
Norwich	1.0	87	2
Cambridge	-4.0	98	3

```
$ more weather data.csv
,temperature,humidity,wind_speed
Ipswich,3.0,85,5
Norwich,1.0,87,2
Cambridge,-4.0,98,3
```

NSIDC

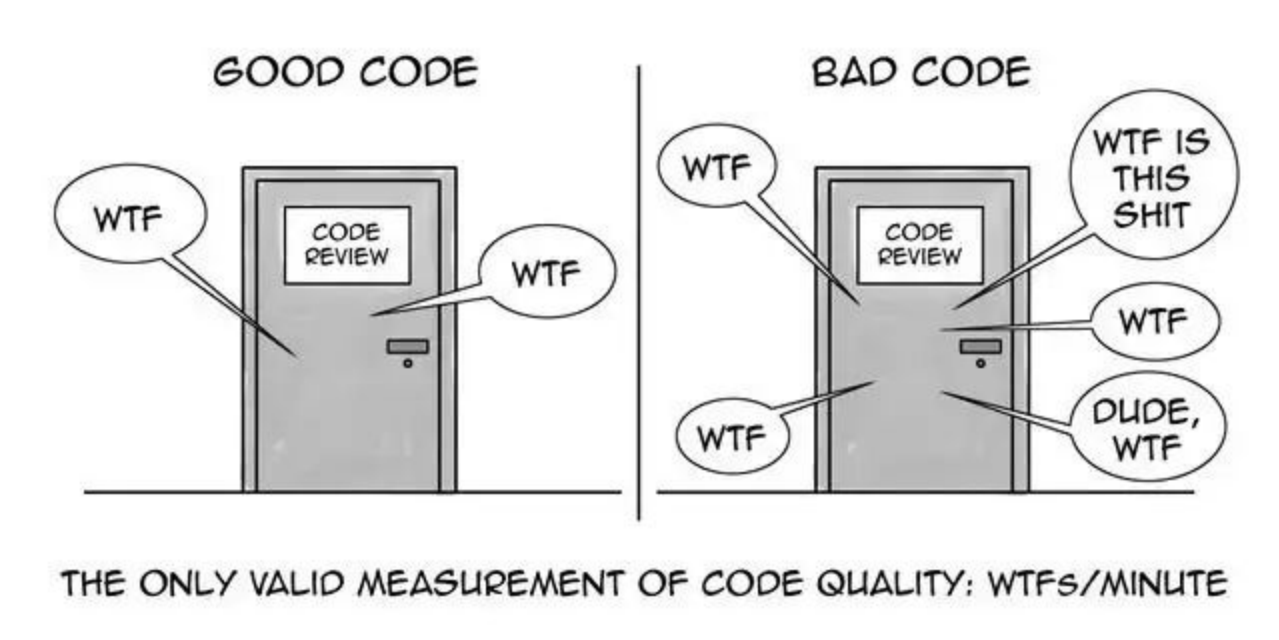
Credit where credit is due

Collaboration means that we need to give credit to all contributors, not just co-authors.

Keep track of who contributed to what.

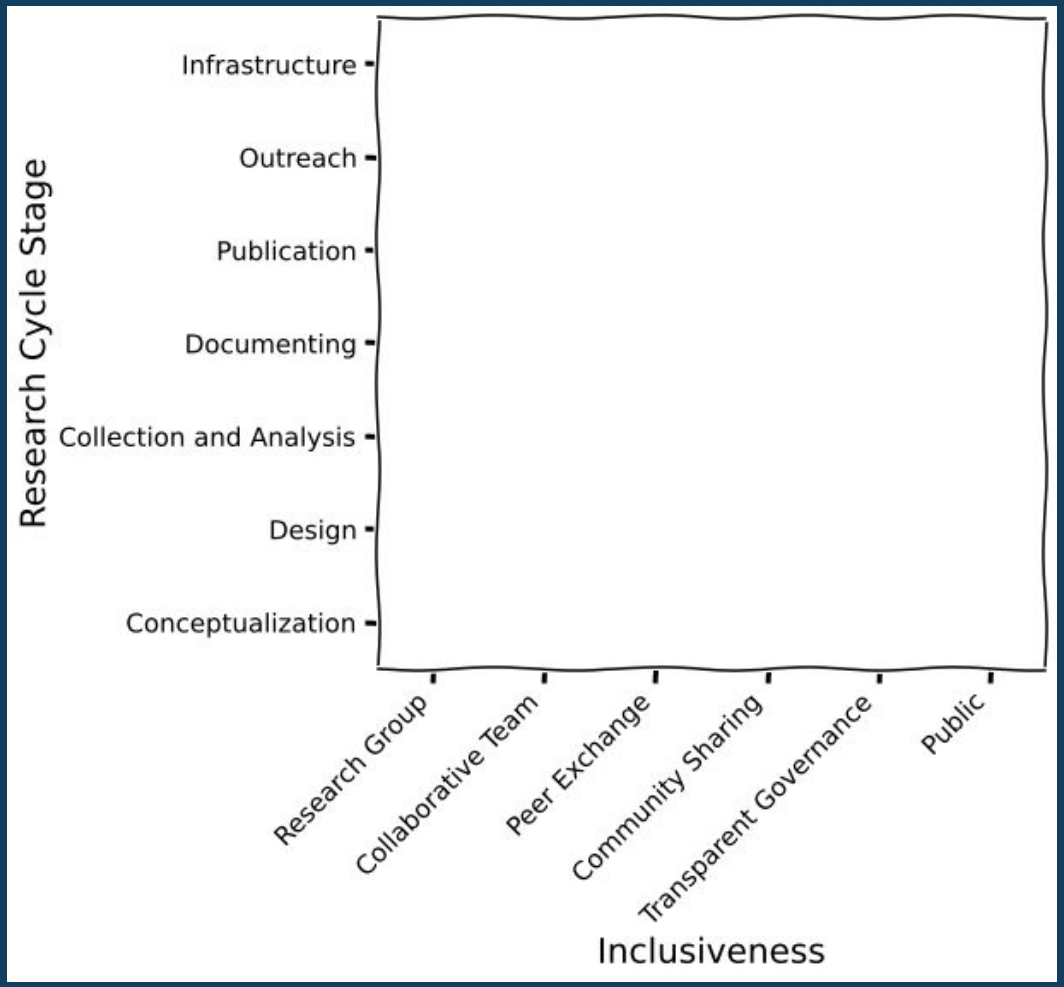
Annual assessments, promotions etc, need to incorporate other modes of contributing.

What is “clean” code



NSIDC

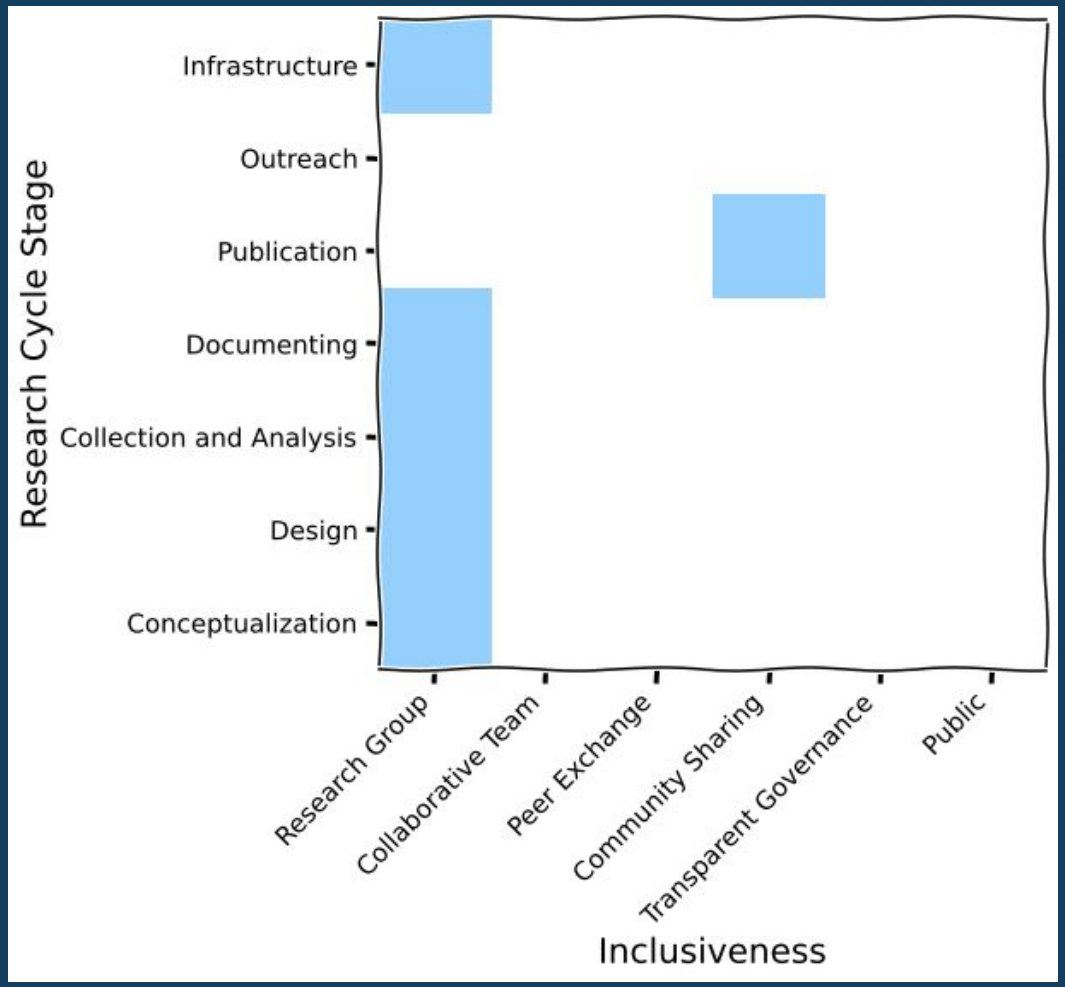
Openness by degrees



Following Whyte and Prior 2011.
DOI: 10.2218/ijdc.v6i1.18

NSIDC

Openness by degrees

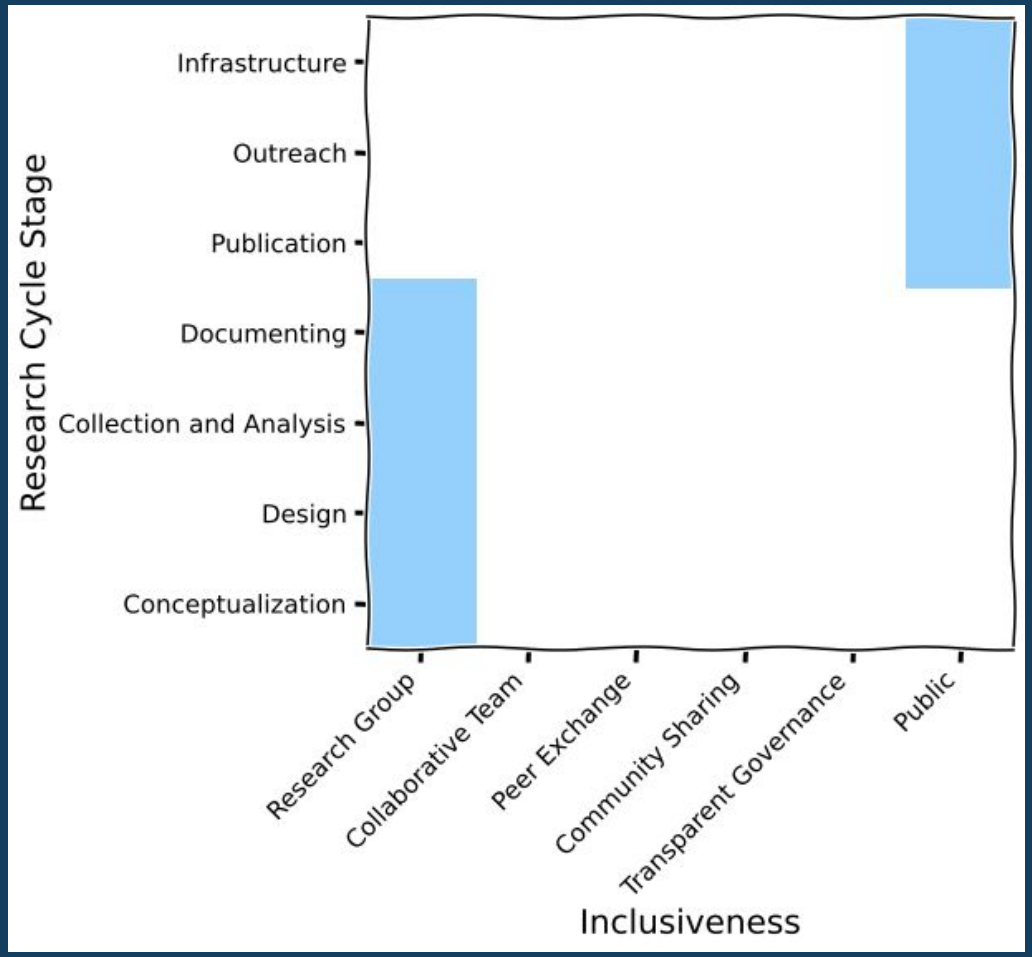


Status Quo

Following Whyte and Prior 2011.
DOI: 10.2218/ijdc.v6i1.18

NSIDC

Openness by degrees

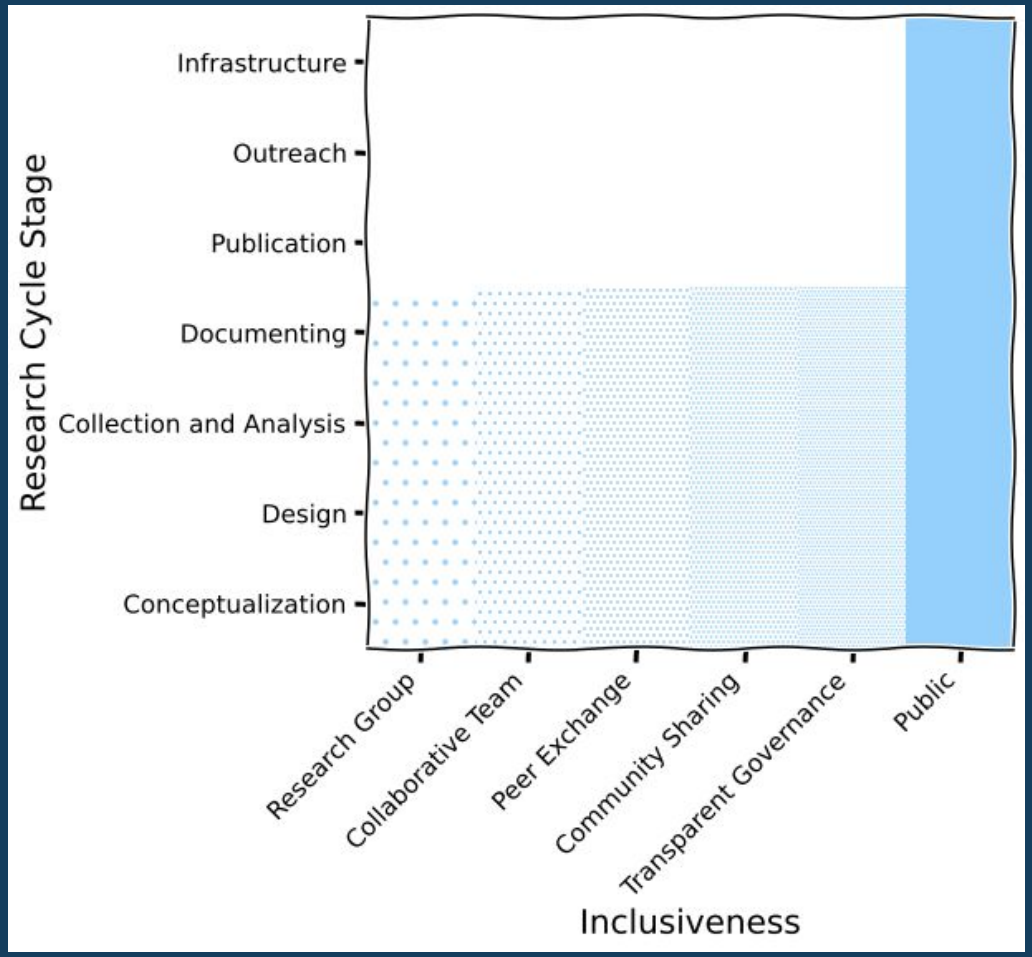


Required by SPD-41a

Following Whyte and Prior 2011.
DOI: 10.2218/ijdc.v6i1.18

NSIDC

Openness by degrees



Open Science
… of some form

Following Whyte and Prior 2011.
DOI: 10.2218/ijdc.v6i1.18

NSIDC