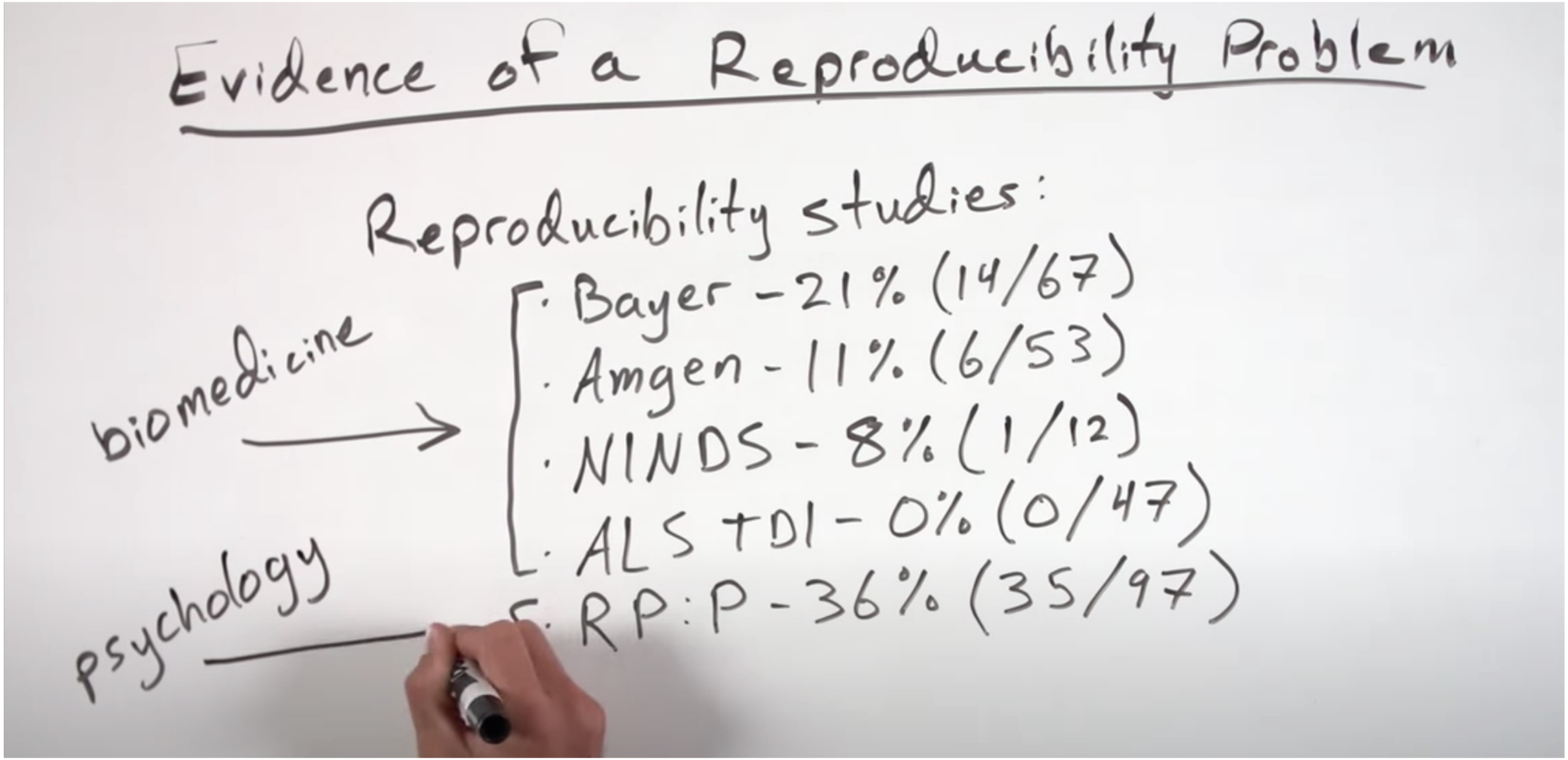
# Open GeoSciences

Lecture 10/5/22

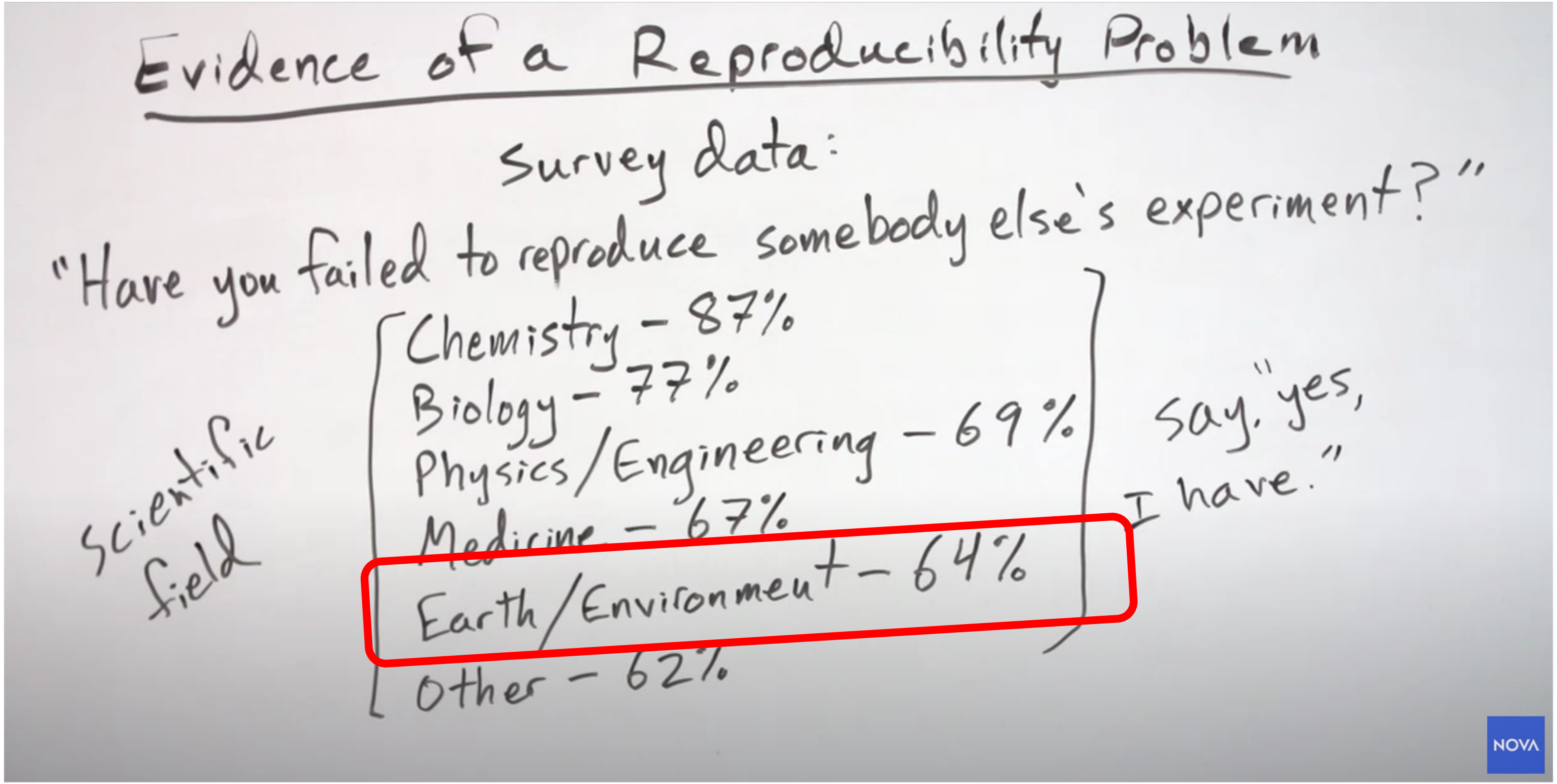
Notes are inspired by EarthDataScience.org, The Turing Way

# Reproducible Science: a crisis




https://www.youtube.com/watch?v=NGFO0kdbZmk&ab_channel=NOVAPBSOfficial

# Reproducible Science: a crisis




https://www.youtube.com/watch?v=NGFO0kdbZmk&ab_channel=NOVAPBSOfficial

Donoho 2010

# Reproducible Science

“The actual scholarship is the full software environment, code and data, that produced the result” **Claerbout** and Karrenbach (1992)

http://sepwww.stanford.edu/sep/jon/

Is reproducibility extra work? Yes in the short run, No in the long run. And it’s worth it:

- Improved work and work habits (people check your codes!)
- Improved teamwork
- Greater impact:
  - less inadvertent competition (if it’s transparent and already done, why doing it again?)
  - More Acknowledgements with better citation and licensing system
- Greater continuity and cumulative impact: if we can reproduce the work, we can better train and advise, and especially be more effective at it
- Research is a public good: public funds pay for the research and should be able to access it.

# Reproducible Science

... is when anyone (including others and your future self) can understand and replicate the steps of an analysis, applied to the same or even new data.

- Transparency in data collection, processing and analysis methods, and derivation of outcomes.
- Publicly available data and associated processing methods.
- Transparent communication of results.

http://openscience.org/what-exactly-is-open-science/, Gezelter (2009)
https://www.earthdatascience.org/courses/intro-to-earth-data-science/open-reproducible-science/get-started-open-reproducible-science/

# Reproducible Science

- ***Computational reproducibility***: When detailed information is provided about code, software, hardware and implementation details. Use scientific programming to process data, automate tasks. Avoid GUI-based workflow that require manual processing.
- ***Empirical reproducibility***: When detailed information is provided about non-computational empirical scientific experiments and observations. In practice, this is enabled by making the data and details of how it was collected freely available.
- ***Statistical reproducibility***: When detailed information is provided, for example, about the choice of statistical tests, model parameters, and threshold values. This mostly relates to pre-registration of study design to prevent p-value hacking and other manipulations.

https://the-turing-way.netlify.app/reproducible-research/overview/overview-definitions.html

# Reproducible vs replicable science

- **Reproducible**: the *same* analysis steps performed on the *same* dataset consistently produces the *same* answer.
- **Replicable**: the *same* analysis performed on *different* datasets produces qualitatively similar answers.

- **Robust**: the work is not dependent on the specificities of the programming language chosen to perform the analysis, results are qualitatively the same.
- **Generalisable**: tresult is not dependent on a particular dataset nor a particular version of the analysis pipeline

| | | Data | |
|---|---|---|---|
| | | Same | Different |
| Analysis | Same | Reproducible | Replicable |
| | Different | Robust | Generalisable |

**Fig. 3** How the Turing Way defines reproducible research

# Reproducible Science in scientific journals

Biostatistics

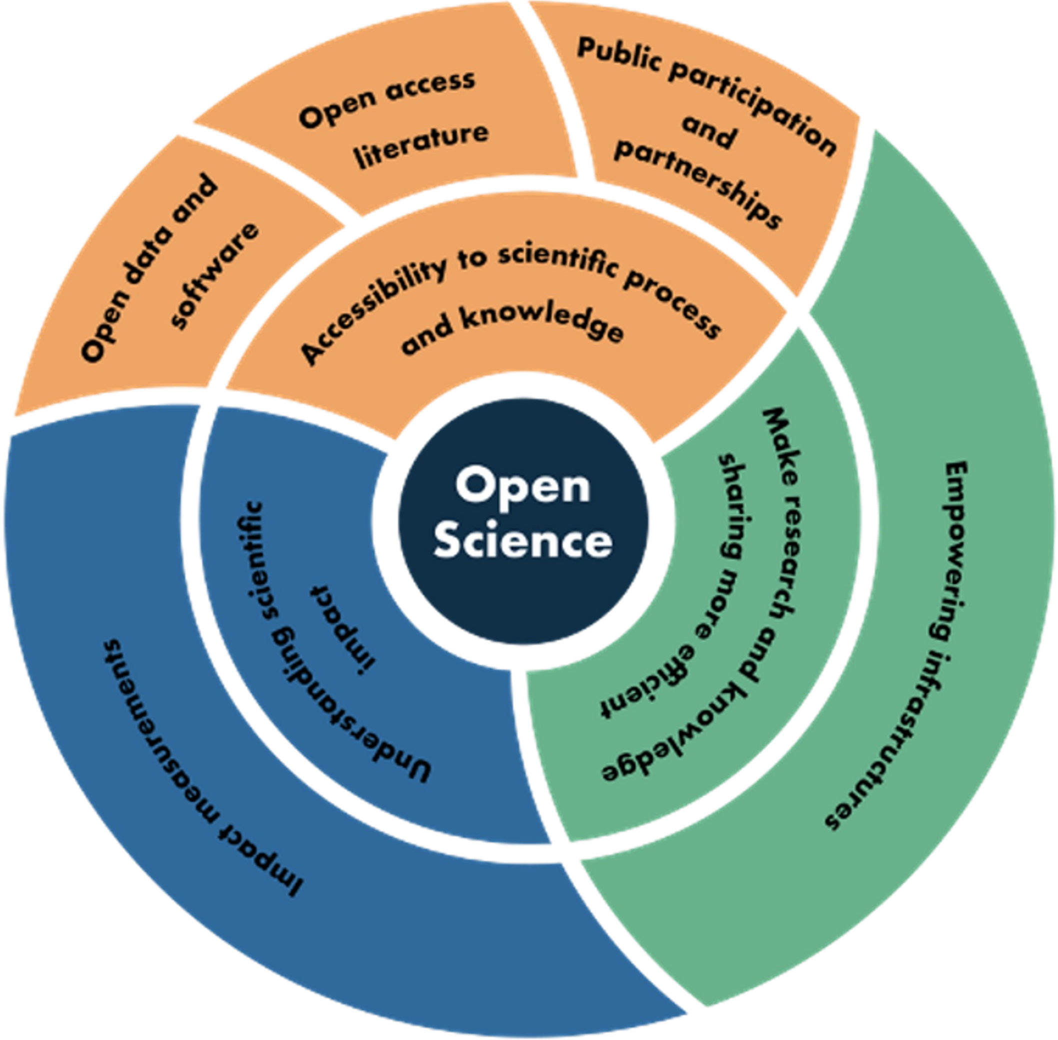
## Reproducible Research

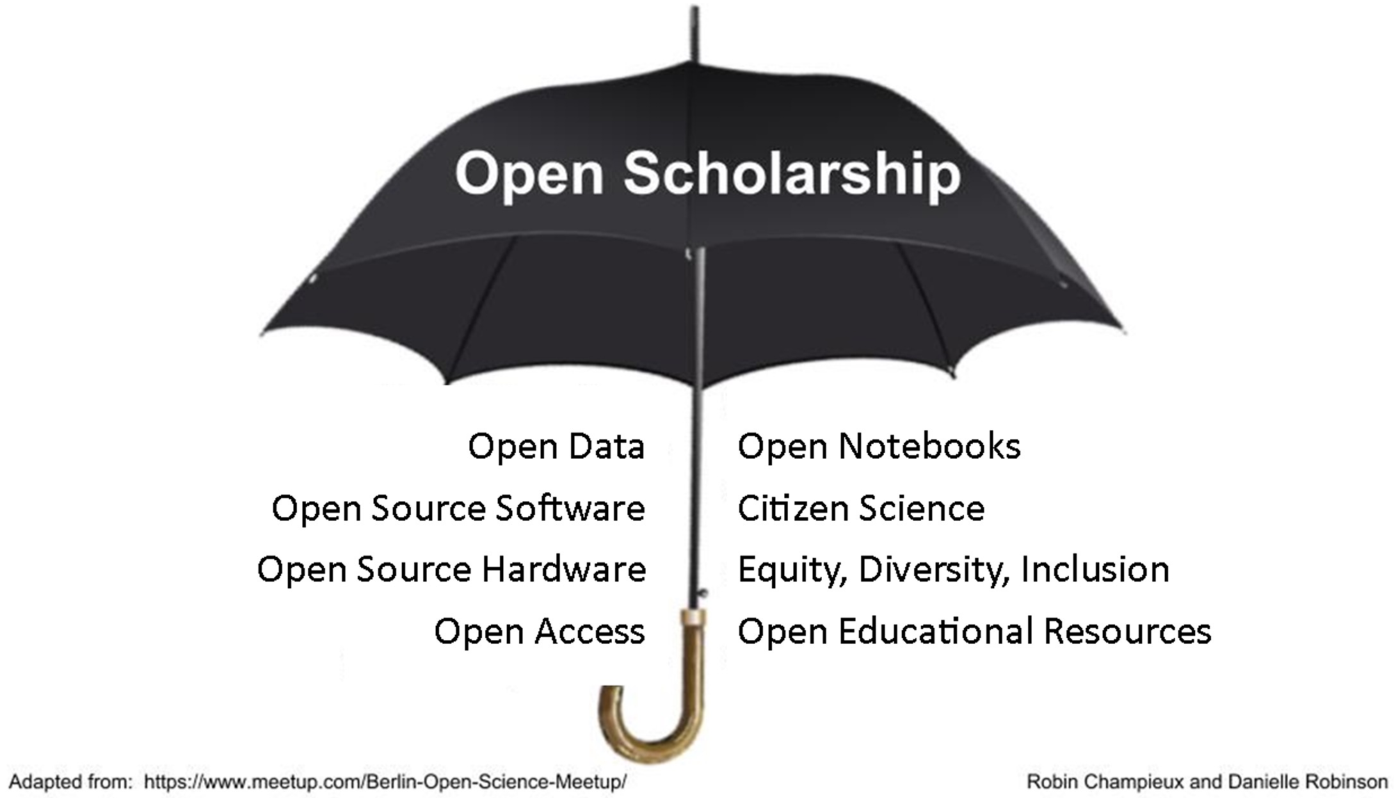
Our reproducible research policy is for papers in the journal to be kite-marked D if the data on which they are based are freely available, C if the authors' code is freely available, and R if both data and code are available, and our Associate Editor for Reproducibility is able to use these to reproduce the results in the paper. Data and code are published electronically on the journal's website as Supplementary Materials.

## Code Availability

Authors are strongly encouraged to submit code supporting their publications. Authors should submit a link to a Github repository and to a specific example of the code on a code archiving service such as Figshare or Zenodo.

**Reproducibility is not yet mandated by journals, but there is a move toward it. Today, the definition of reproducibility is still in debate**


Open Science
Accessibility to scientific process and knowledge
Open data and software
Open access literature
Public participation and partnerships
Make research and knowledge sharing more efficient
Empowering infrastructures
Understanding scientific impact
Impact measurements




Open scholarship means open to *everyone* without discrimination based on factors such as race, gender, sexual orientation, or any number of other factors.

# Open Science

***Open Data***: Documenting and sharing research data openly for re-use.

***Open Source Software***: Documenting research code and routines, and making them freely accessible and available.

***Open Hardware***: Documenting designs, materials, and other relevant information related to hardware, and making them freely accessible and available.

***Open Access***: Making all published outputs freely accessible for maximum use and impact.

# Open Science

***Open Notebooks***: An emerging practice, documenting and sharing the experimental process of trial and error.

## *Open Scholarship:*

***Open educational resources***: Making educational resources publicly available to be re-used and modified.

***Equity, diversity, inclusion***: Ensuring scholarship is open to anyone without barriers based on factors such as race, background, gender, and sexual orientation.

***Citizen science***: The inclusion of members of the public in scientific research.

# Open Science: **Open Data**

1. Protect your raw data (do not modify it)
2. Make your data available (online, free access, reliable archive, under an open license)
3. Make your data easy to understand (data+metadata+documentation)
4. Make your data easy to use (data stored in standard formats)
5. Make your data citeable (Digital Object Identifier )

# Open Science: **Open Data**

Host your data on public repository

Files < 50GB, free

Data
Software

Data set < 300GB
File size < 50GB
$120 for data < 50GB
Curated by librarians
**Best for RAW DATA**

Curated by
librarians
**RAW DATA**

**PANGAEA.**
Data Publisher for Earth &
Environmental Science

Data < 20GB
free

Wilkinson et al., 2016

# Open Science: **Open Data**

**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

"Good data management and stewardship is not a goal in itself, but rather a pre-condition supporting knowledge discovery and innovation."

# Open Science: **Open Data**

**Domain specific archives**

IRIS, UNAVCO

Curated, static, not diverse, follows FAIR guidelines, has established licenses, is citeable.

**General data repositories archives** [less integrated, less curated, citeable]

Zenodo, Figshare, Dryad, Pangaea, Dataverse, Mendeley Data, DANS, EUDat, DataHub

**Data Journals**

Earth system science data

**Registry of Research data Repositories**

https://www.re3data.org/

# Open Science: **Open Data**

**requirement for Authors @ AGU journals**

**Data** must be **deposited in repository** that supports FAIR principles.

**Data availability statement** : data citation + statement in acknowledgement

**Data citations:** Follows ESIP's guidelines.

Example:

Cline, D., R. Armstrong, R. Davis, K. Elder, and G. Liston. 2003. CLPX-Ground: ISA snow depth transects and related measurements ver. 2.0. Edited by M. A. Parsons and M. J. Brodzik. NASA National Snow and Ice Data Center Distributed Active Archive Center. https://doi.org/10.5060/D4MW2F23. Accessed 2008-05-14

# Open Science: **Open Data**

requirement for [Authors @ AGU journals]

Select a repository:

- **Domain/disciplinary repository:** IRIS for seismic data, UNAVCO for geodetic data etc, NASA, NOAA, USGS, ..
- **Computing center** : HPC or institutional clusters may have their own data management storage and preservation recommendations.
- **Institutional repository**: ResearchWorks @ UW ( ~ xGBs)
- **General repository:** Zenodo, Figshare, Dryad, Pangaea, etc etc

# Open Science: **Open Data**

### Requirement for [Authors @ Nature Journals]

A condition of publication in a Nature Portfolio journal is that **authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications.**

# Open Science: **Open Data**

## Requirement for Authors @ Nature Journals

A condition of publication in a Nature Portfolio journal is that **authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications.**

## Requirement for Authors @ AAAS Journals

The *Science* Journals generally require **all data underlying** the results in published papers to be **publicly and immediately available**. Post-publication embargoes are not permitted, nor are **stipulations for readers to contact the authors** (rare exceptions involving third-party datasets must be discussed with the editor prior to publication and explained in detail in the acknowledgments).
**Meet community standards + machine-readable formats**

# Data Journals

EGU: Earth System Science Data (OA)

Wiley: Geoscience Data Journal (OA)

Nature: Scientific Data (OA)

Seismological Society of America: SRL Data Mine

Seismica (OA) - Tektonika (OA) - Volcanica (OA)

Do you know of other?

# Open Science: **Open Source Software**

A <u>spectrum</u> of levels for OSS :

1. In journal supplements (e.g. a zip file attached to the paper or on zenodo)
2. On Github with 1 commit and a README
3. On GitHub with documentation, contributing guide, test
4. On GitHub… + package repository (conda-forge, Pypi)
5. On GitHub, … + automated testing and deployment (Github Actions, continuous integration)

***A code without a license cannot be reused or modified or extended***

Acknowledgement: inspired by Dr. Leonardo Uieda, SPIN [talk]

# Open Science: **Open Source Software**

Why?

- Others can re-use your code and benefit from your knowledge and savoir-faire (no need to reinvent the wheel!)
- It forces you to be be organized, clean, structured. You gain in work efficiency and productivity!
- Helps debugging codes (e.g. people posting github issues!)
- Meet people by participating in a OSS development!
- Find mentor and teach each other
- Create a public reputation for good software practices

How?

- Use version control (hosting services: GitHub, GitLab, BitBucket)
- Publish your code on Zenodo (integration with github) and get a DOI for it.
- Follow best practices for OSS
- Contribute or lead an open source project:
    - helps to build and maintain good coding habits.
    - demonstrates technical ability
    - Provides a favorable reputation

# Open Science: **Open Source Software**

Use standard, operating system agnostic, data formats (CSV, MD, TXT).
Avoid PDF, word docs, proprietary file formats like excel.
File naming tips:

- Name your files consistently
- Keep it short but descriptive
- Avoid special characters or spaces to keep it machine-compatible
- Use underscores to keep it human-readable. Avoid capitals or case-sensitive characters
- Use consistent date formatting, for example ISO 8601: YYYY-MM-DD to maintain default order
- Include a version number when applicable
- Share/establish a naming convention when working with collaborators
- Record a naming convention in your data management plan

```
* 2020-image.jpg
* 2019-image.jpg

* 2018-image.jpg
```

| | |
|---|---|
| Myabstract.docx | 2020-06-08_abstract-for-sla.docx |
| Joe's Filenames Use Spaces and Punctuation.xlsx | Joes-filenames-are-getting-better.xlsx |
| figure 1.png | Fig01_scatterplot-talk-length-vs-interest.png |
| fig 2.png | Fig02_histogram-talk-attendance.png |

# Computational Workflows

Create a modular code with functions:

- Good for repetitive tasks. 1-3 inputs, 1-2 outputs. One function, one task. Keep it small and focused. Combine functions in a wrapper function
- Easier to update code. Easier to test code.
- Python has guidelines to write in *good style*

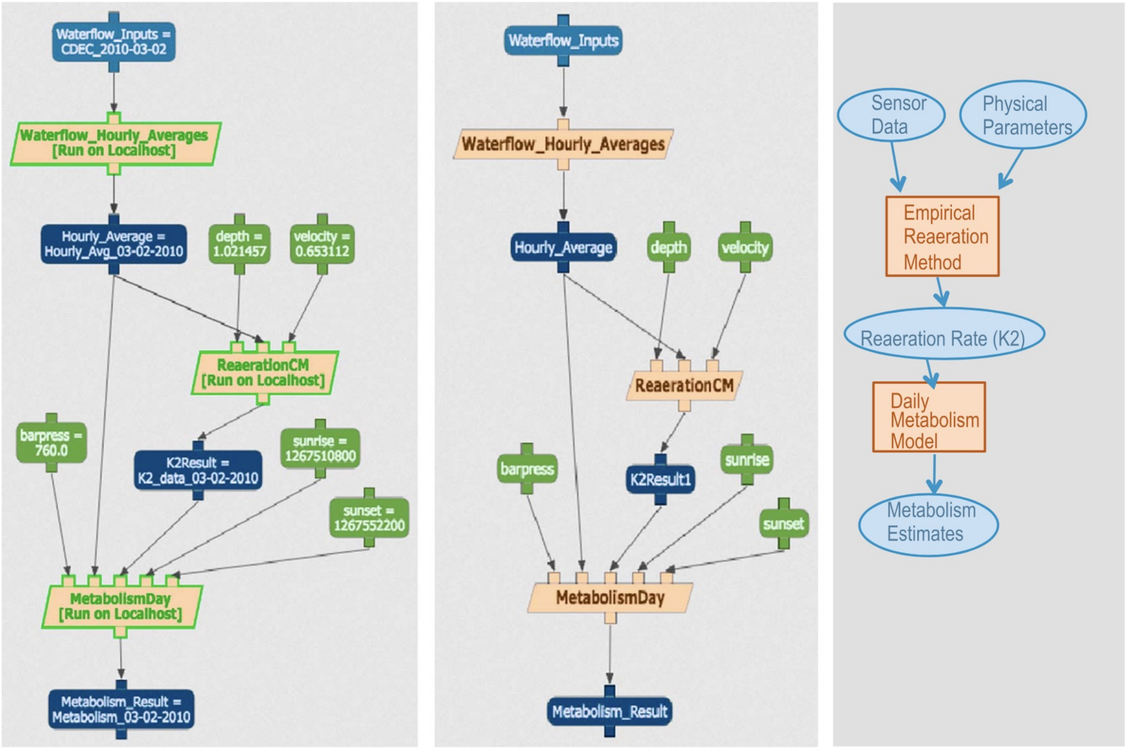
Expand your code to add checks and tests.

```
# Example of a simple function
def add_five(x):
    """Add the numeric value 5 to an
input value.

    Parameters
    ----------
    x : int or float

    Returns
    ----------
    int or float
        Input data with value increased
by 5.
    """

    return (x + 5)

# How do we call a Python function

again?

add_five(3)
```

# Organize the project directory: make it easy to find code, data, ...

| Organized Project | Non Organized Project |
|---|---|
| /01-scripts<br>    01-clean-data.py<br>    02-run-model.py<br>    03-create-plots.py<br>/02-data<br>    /raw-data<br>        /landsat-imagery<br>        /fire-boundary<br>/03-output-graphics<br>    study-area-map.png<br>/04-final-paper<br>    fire-paper.pdf | file1-new.py<br>file1.py<br>plotting-test.py<br>data-file.txt<br>/old-stuff<br>testoutput1.txt<br>testoutput2.csv |

# Document the workflow (text, flow chart, ...)




(a) Computational provenance of an execution run

(b) Reusable workflow corresponding to the general method

(c) High-level workflow diagram

- Make a workflow diagram in your paper.
- Draw a sketch that can be saved into a JSON file ([https://asset.isi.edu/](https://asset.isi.edu/))

# Open Science: **Open Access**

# **A crisis**

How librarians, pirates, and funders are liberating the world's academic research from paywalls.

By Brian Resnick and Julia Belluz | Updated Jul 10, 2019, 3:58pm EDT

*Illustrations by Javier Zarracina*

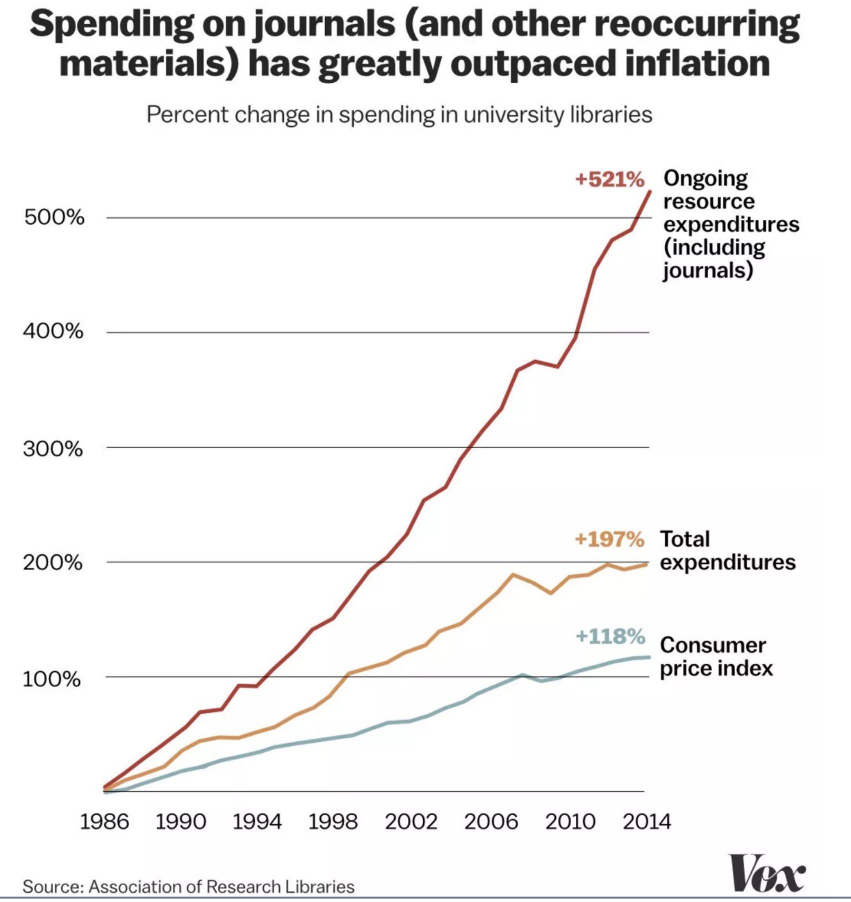
SHARE

The Highlight BY Vox

The **27,500 scientists** who work for the University of California generate **10 percent** of all the academic research papers published in the United States.

Their university recently put them in a strange position: Starting July 10, these scientists will not be able to directly access much of the world's published research they're not involved in.

That's because in **February**, the UC system — one of the country's largest academic institutions, encompassing Berkeley, Los Angeles, Davis, and several other campuses — dropped its nearly **$11 million annual subscription** to Elsevier, the world's largest publisher of academic journals.




https://www.vox.com/the-highlight/2019/6/3/18271538/open-access-elsevier-california-sci-hub-academic-paywalls

# Open Science: **Open Access**

**Nature**: for profit journals. Nature communication is open access. Professional editors (non academics)
**Science Magazine:** non-profit journals. Science Advances is open access. Professional editors (non academics)
**AGU**: non-profit journals

Gold Open Access journals

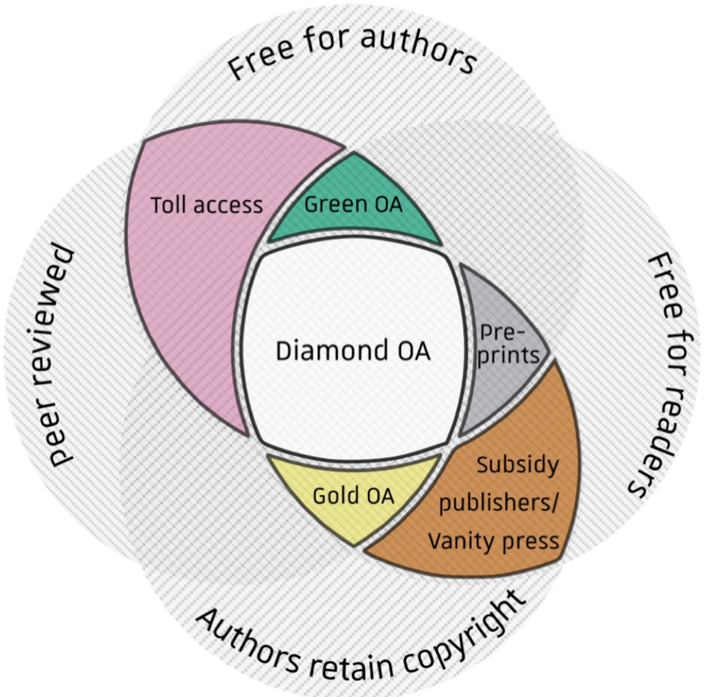
AGU has 5 fully Open Access journals: *AGU Advances*, the *Earth's Future*, *Earth and Space Science*, *GeoHealth*, and *Journal of Advances in Modeling Earth Systems*. *Space Weather* will become open access in January 2020.

- Authors must pay an Article Processing Charge (APC) for their article to be published.
- Authors can select one of several Creative Commons (CC) licenses.
- Articles published in these journals are freely available online immediately upon publication.

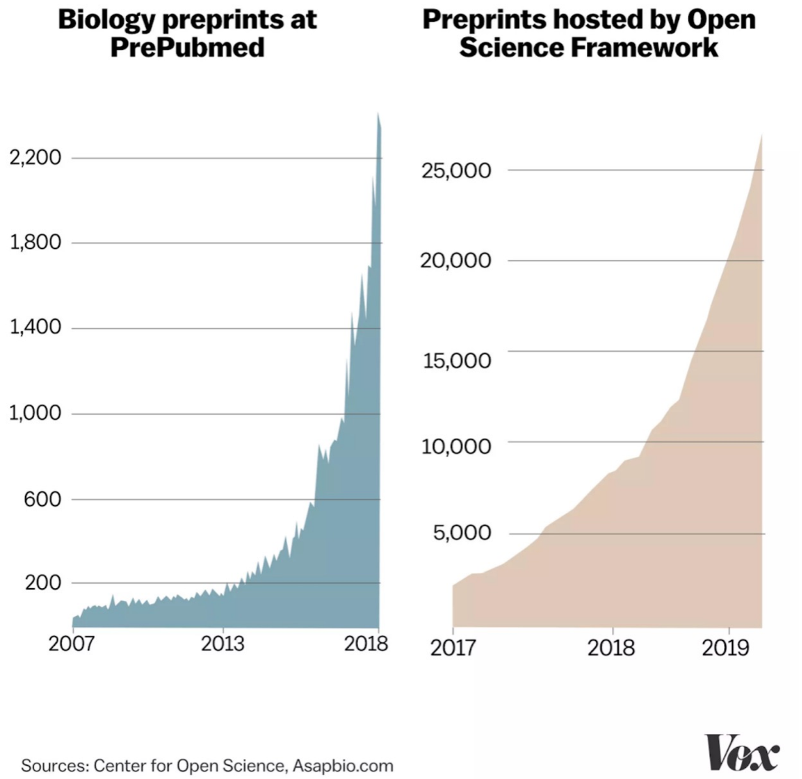
Hybrid Open Access journals

AGU has 16 hybrid journals where readers will find some articles available via Open Access while others will be behind a paywall unless their institution has a subscription.

- If an author wants their article to be available free of charge to all readers (i.e. Open Access) they will need to pay an Article Processing Charge. Authors can select this option after their article is accepted. Authors select one of several Creative Commons (CC) licenses. The article is freely available online immediately upon publication.
- Alternatively authors do not pay an Article Processing Charge to be published in the journal but their article is only accessible via subscription or purchase.



# Open Science: **Open Access and Pre-Prints**




ArXiv: https://arxiv.org/
EarthArXiv: https://eartharxiv.org/
ESSOAr:[https://www.essoar.org/](https://www.essoar.org/)

Find open access journals:
https://www.doaj.org/

**Be wary of non-peer reviewed work**
**Be wary of peer reviewed work.**

# Open Science: **Diamond Open Access Earth science Journals**

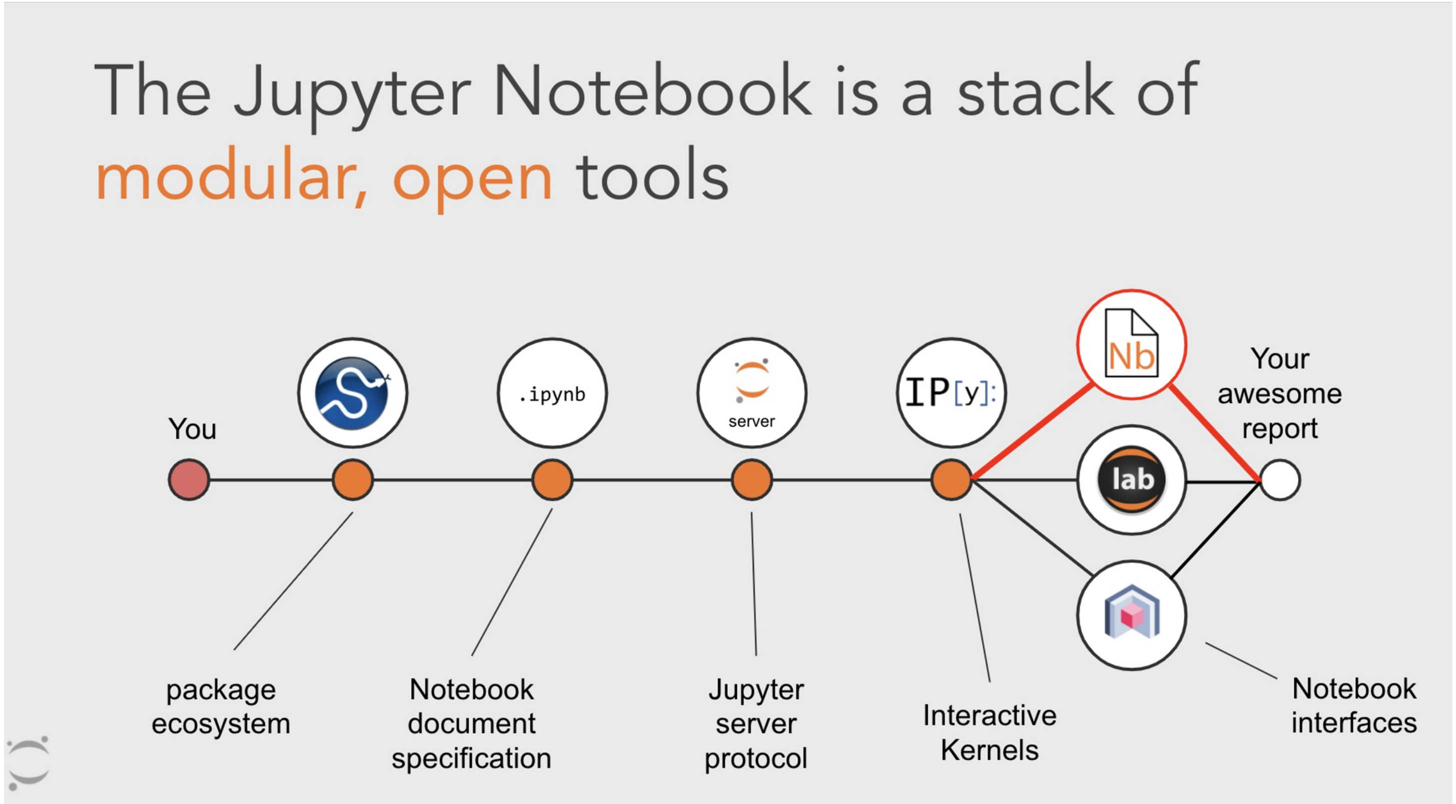
Volcanica

tektonika

SEISMICA

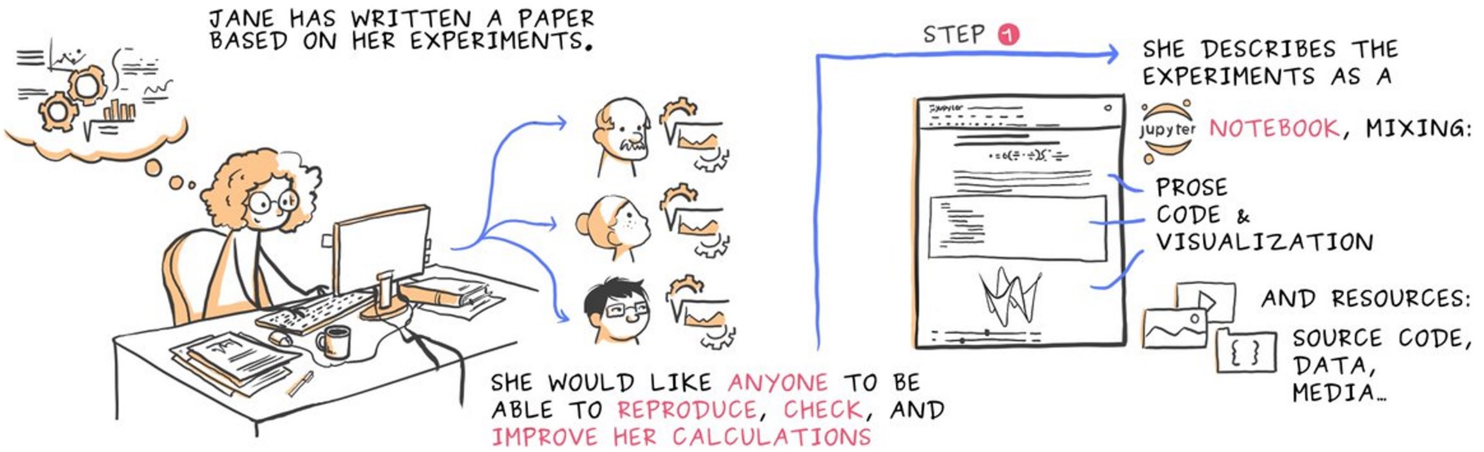
**Do you know of others?**

# Open Science: **Open Notebooks**

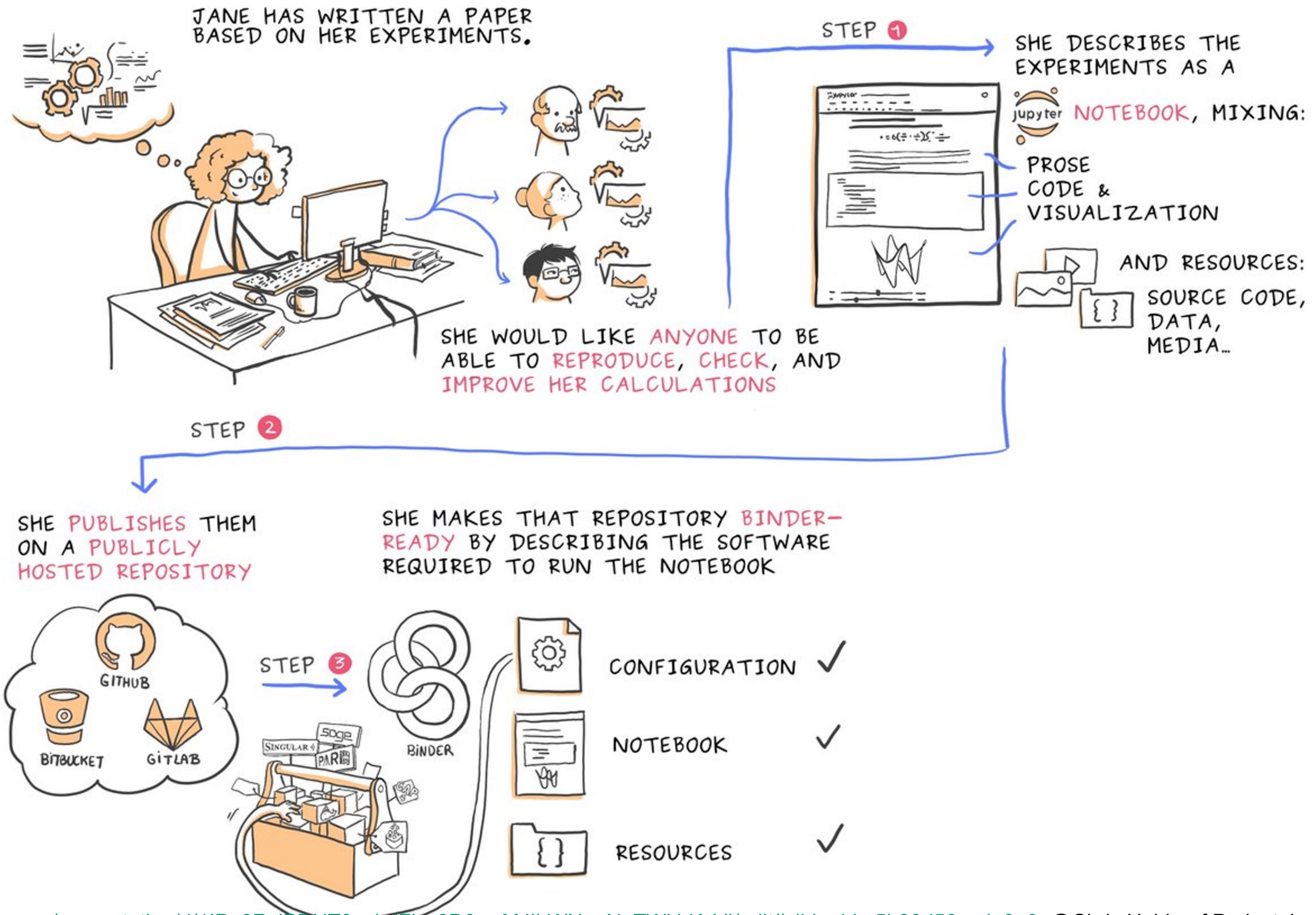



**Jupyter Notebook:** A browser-based application that allows you to create and share documents (i.e. Jupyter Notebook files) that contain live **code**, **equations**, **visualizations** and **narrative text**.

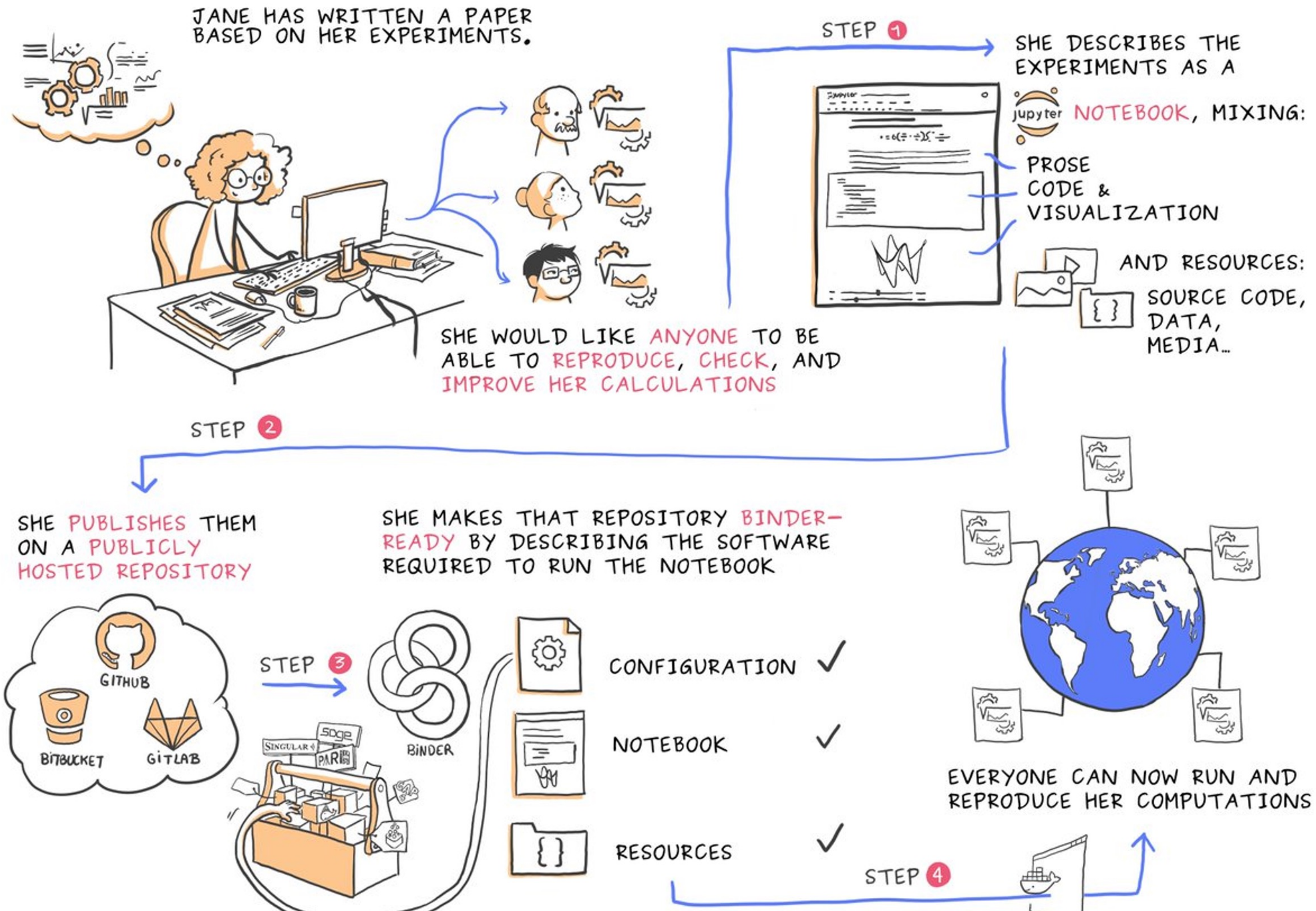

https://docs.google.com/presentation/d/1IP_0RclPDHT3ezbdFjxc8D0xwMJIhWNnnNgTWkbKyVI/edit#slide=id.g5b30d56ecd_0_0, @Chris Holdgraf Project Jupyter




https://docs.google.com/presentation/d/1IP_0RclPDHT3ezbdFjxc8D0xwMJIhWNnnNgTWkbKyVI/edit#slide=id.g5b30d56ecd_0_0, @Chris Holdgraf Project Jupyter




https://docs.google.com/presentation/d/1IP_0RclPDHT3ezbdFjxc8D0xwMJlhWNnnNgTWkbKyVI/edit#slide=id.g5b30d56ecd_0_0, @Chris Holdgraf Project Jupyter




https://docs.google.com/presentation/d/1IP_0RclPDHT3ezbdFjxc8D0xwMJIhWNnnNgTWkbKyVI/edit#slide=id.g5b30d56ecd_0_0, @Chris Holdgraf Project Jupyter

# Checklist to open science research: **Share, Publish, Archive**

**Share**: provide access to information

**Publish**: render the data and/or code citable and discoverable

- Raw [unless already available] or starting data set, metadata, data cleaning/manipulation steps, analysis scripts, source codes, readme
- Don’t re-publish, don’t publish pre-existing restrictive license, password and private keys

**Archive**: long-term preservation of data and/or code:

- Is there a domain specific repos? Are there backup and replication policies? Is there a plan for long-term preservation?
- **It needs to be findable**
- **It needs to be citable (DOI)**

Earth Lab, https://www.earthdatascience.org/slide-shows/share-publish-archive/#4

# The Geoscience Paper of the Future

Gil et al (2016)

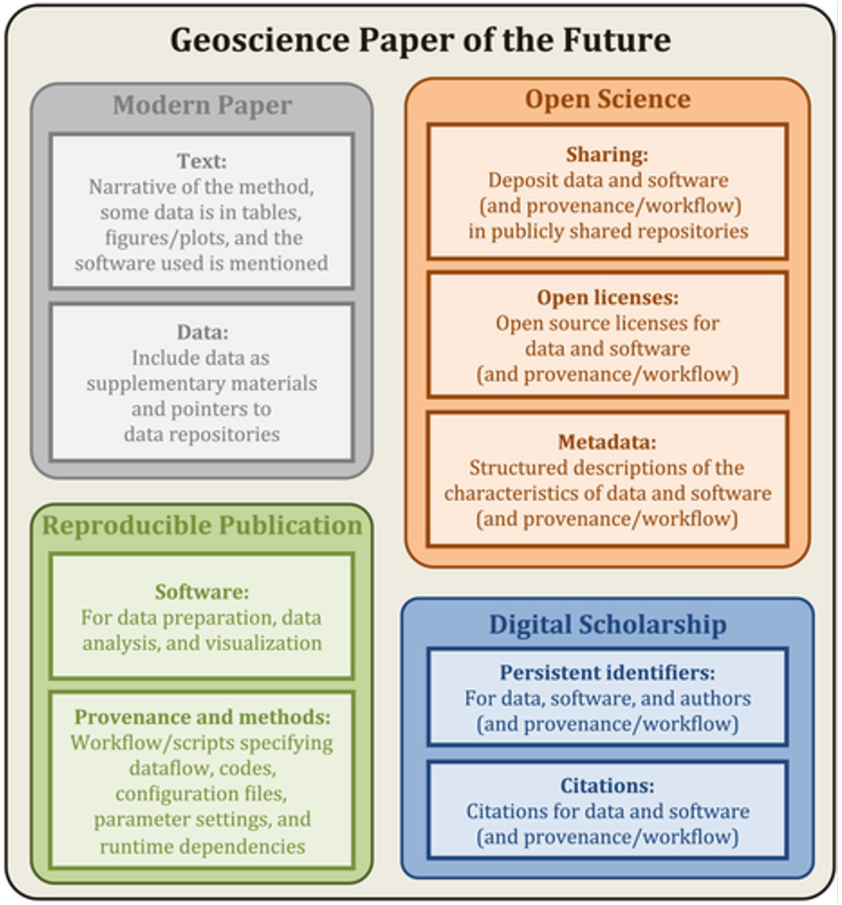



**Table 1.** A Proposed Checklist for GPF Authors, With 20 Recommendations That Can Guide Them to Assemble the Information That Should Be Included in a GPF

| Category | Applicability | Recommendations |
|---|---|---|
| Data accessibility | Initial data, significant intermediate results, and final results | D1: Data sets should be published in a publicly accessible location with a permanent unique identifier<br>D2: Data sets should have a license<br>D3: Data sets should be cited in the paper |
| Data documentation | Initial data, significant intermediate results, and final results | D4: Data sets should have general-purpose metadata specified<br>D5: Data set characteristics should be explained in detail<br>D6: Data set origins and availability of related data sets should be documented |
| Software accessibility | Software used to process initial data and to generate any intermediate or final results | S1: Software should be published in a publicly accessible location with a permanent unique identifier<br>S2: Software should have a license<br>S3: Software should be cited in the paper |
| Software documentation | Software used to process initial data and to generate any intermediate or final results | S4: Software function and purpose should be described<br>S5: Software download and execution requirements should be documented<br>S6: Software testing and reuse with new data should be documented<br>S7: Software support for extensions and updates should be mentioned |
| Provenance documentation | Provenance of all computational results reported in the article, including figures, tables, and other findings | P1: Derivations of newly generated data from initial data should be provided<br>P2: Software execution traces for newly generated results should be provided<br>P3: Versions and configurations of the software should be specified<br>P4: Parameter values used to run the software should be specified |
| Methods documentation | Computational methods that are generally applicable to data other than the data in the paper | M1: Compositions of software that form a general reusable method should be specified<br>M2: Data flow across software components should be described |
| Authors Identification | Authors of the paper and of any new data and software cited in the paper | A1: Authors have a permanent unique identifier |

# The Geoscience Paper of the Future: DATA

Gil et al (2016)

**Table 2.** A High-Level Roadmap for Authors of a Geoscience Paper of the Future (GPF), With Pointers to Popular Resources for Publication and Licensing of Scientific Research Products

| Data | | |
|---|---|---|
| Available in a public repository, with metadata, a license specifying conditions of use, and citable using a unique and persistent identifier. | | |
| Data Accessibility | | What to Show in the Paper |
| Simplest Approach | Advanced Approach | |
| 1. Choose a general repository (e.g., figshare, Zenodo, Dryad, Pangaea, etc.)<br>2. Create a public entry for your data set with a persistent unique identifier<br>3. Specify basic metadata (name, authors)<br>• Including license—choose from creativecommons.org<br>4. Upload/point to the data<br>5. The repository will give you a data citation | 1. Find a repository that your community uses, if there is not one then organize one!<br>2. Create a public entry for your data set with a persistent unique identifier<br>3. Specify the basic metadata required by that repository<br>• Including license—choose from creativecommons.org<br>4. Upload/point to the data<br>5. Get a data citation from the repository | • Cite data set in the paper references<br>• Citation includes data set name, creators, publication date, repository name, persistent identifier, and time of retrieval<br>• If there is a separate paper about a data set, cite it as well<br>• Mention in the text that the persistent identifier site has the metadata and includes a detailed description of the data<br>• Mention availability of related data sets |
| Data Documentation | | |
| Simplest Approach | Advanced Approach | |
| • Data sets should have at least domain-specific keywords<br>• Include metadata and documentation that will help reuse | • Domain-specific metadata should be well documented using metadata standards for that community | |

# The Geoscience Paper of the Future: SOFTWARE

Gil et al (2016)

| Software | | |
|---|---|---|
| Available in a public repository, with documentation, a license for reuse, and a unique and citable persistent identifier. This includes any ancillary software for data reformatting, data conversions, data filtering, and data visualization. | | |
| **Software Accessibility** | | **What to Show in the Paper** |
| Simplest Approach | Advanced Approach | • Cite software in the paper references<br>  ○ Citation similar to data but includes software version<br>• If there is a software paper, cite it<br>• Mention that the persistent identifier location for your software points to its metadata<br>• Optionally, include the software metadata as supporting information |
| 1. Create a public entry for your software with a persistent unique identifier<br>2. Post on your web site and use a PURL, upload to a data repository and get a DOI<br>3. Specify basic metadata<br>4. Include license—choose from opensource.org/licenses, e.g., Apache<br>5. Specify desired citation | 1. Learn to use a code repository that allows version tracking and collaborative software development (e.g., GitHub, BitBucket, etc.)<br>2. Create a public entry for your software with a persistent unique identifier<br>3. Specify basic metadata<br>  • Include license—choose from opensource.org/licenses, e.g., Apache<br>4. Get a software citation from the repository. | |
| **Software Documentation** | | |
| Simplest Approach | Advanced Approach | |
| • Describe as much metadata as will help reuse<br>  ○ Document basic metadata including authors, contributors, version, license, and release date | • Use a software registry<br>  ○ www.ontosoft.org/portal, csdms.colorado.edu, etc.<br>  ○ Guides through questions to provide metadata<br>• Save the metadata as HTML, XML, etc.<br>• Post the metadata on your code site | |

# The Geoscience Paper of the Future: WORKFLOWS

Gil et al (2016)

**Table 2.** (continued)

| Computational Provenance and Methods | | |
|---|---|---|
| Documented for all results by explicitly describing the series of computations and their outcome with a trace log (or computational provenance record) and a high-level workflow diagram (or a formal workflow), possibly in a shared repository and with a unique and persistent identifier. | | |
| Documentation | | What to Show in the Paper |
| Simplest Approach | Advanced Approach | |
| 1. Provide a trace log or a formal computational provenance record<br>2. Provide a provenance summary in text<br>• Data + software<br>• Specify unique identifiers for data and software, versions, credit all sources<br>3. Develop a high-level workflow diagram<br>• Capture dataflow across major computational steps as a diagram | 1. Provide a computational provenance summary in text<br>• Data + software<br>• Specify unique identifiers for software versions<br>2. Develop a high-level workflow diagram<br>• Capture high-level dataflow across major computational steps<br>3. Specify the formal workflow using a workflow system, electronic notebook, etc.<br>• Command lines + parameter values<br>• Dataflow across components<br>4. Include the computational provenance record<br>• If generated automatically, preferably using a standard (e.g., PROV)<br>5. Publish the workflow and computational provenance record in a repository (e.g., myExperiment.org) or a data repository<br>6. Get a unique persistent identifier for the workflow, the computational provenance, or both | • Describe workflow in text and provide a workflow diagram<br>○ Optionally, provide the formal workflow or lab notebook, use a persistent identifier, and cite it<br>• Include a provenance summary as supplementary material, or use a persistent identifier and cite it<br>○ Optionally, include the trace log pr the computational provenance records using a standard (e.g., W3C PROV), also with a persistent identifier and cited |

# Ethics in publishing

## CRediT – Contributor Roles Taxonomy

CRediT (Contributor Roles Taxonomy) is high-level taxonomy, including 14 roles, that can be used to represent the roles typically played by contributors to scientific scholarly output. The roles describe each contributor's specific contribution to the scholarly output.

### 14 Contributor Roles

| | |
|---|---|
| Conceptualization | Resources |
| Data curation | Software |
| Formal Analysis | Supervision |
| Funding acquisition | Validation |
| Investigation | Visualization |
| Methodology | Writing – original draft |
| Project administration | Writing – review & editing |

**Conceptualization – Ideas**; formulation or evolution of overarching research goals and aims.
**Data curation** – Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later re-use.
**Formal analysis** – Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data.
**Funding acquisition** - Acquisition of the financial support for the project leading to this publication.
**Investigation** – Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.
**Methodology** – Development or design of methodology; creation of models.
Project administration – Management and coordination responsibility for the research activity planning and execution.
**Resources** – Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.
**Software** – Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.
**Supervision** – Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.
**Validation** – Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.
**Visualization** – Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.
**Writing** – original draft – Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).
**Writing** – review & editing – Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages.

# Diamond Open Access in Earthquake Sciences