



# Wasn't Open FAIR Enough?

The future of data and software publication

Jens Klump | 25 September 2019

Australia's National Science Agency



This work is licensed under a  
Creative Commons Attribution  
4.0 International License.





# Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities

“The Internet has fundamentally changed the practical and economic realities of distributing scientific knowledge and cultural heritage. For the first time ever, the Internet now offers the chance to constitute a global and interactive representation of human knowledge, including cultural heritage and the guarantee of worldwide access.

[...] Obviously, these developments will be able to significantly modify the nature of scientific publishing as well as the existing system of quality assurance.

[...]” (Berlin Declaration, 2003)



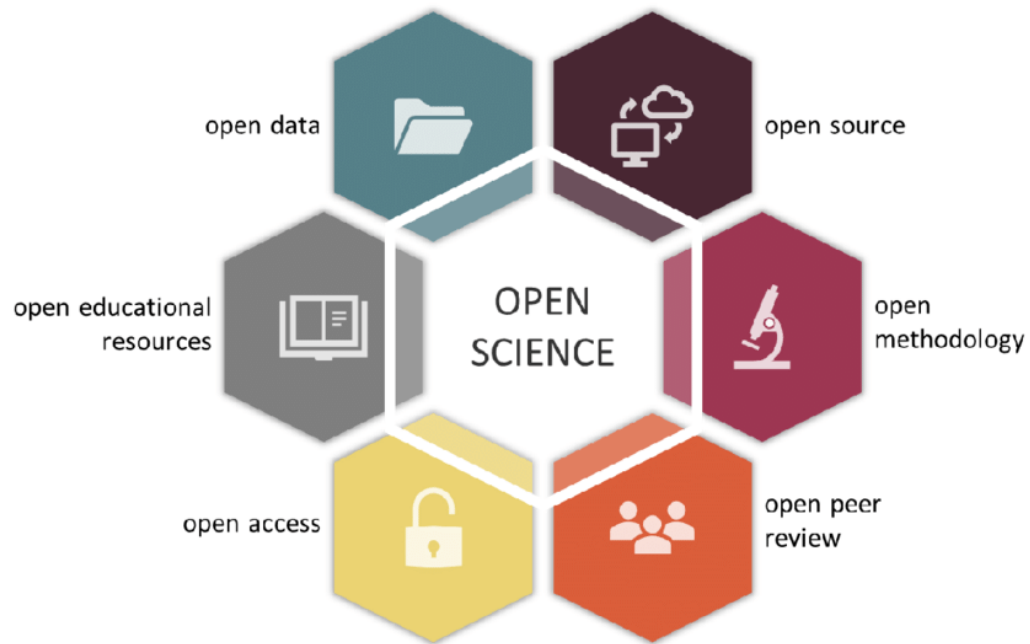
# Berlin Declaration

“Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material.”

*Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*



# Open Science



# Section 1: Open Access

**Berlin Declaration and Beyond**





# Open Access and the Berlin Declaration

- Today the value of open access to publicly funded research data and of open source software is well accepted.
- The most prominent call for Open Access came with the publication of the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities in 2003.
- 644 signatories to date.
- The Berlin Declaration is significant because from the start it was backed by major research organisations.

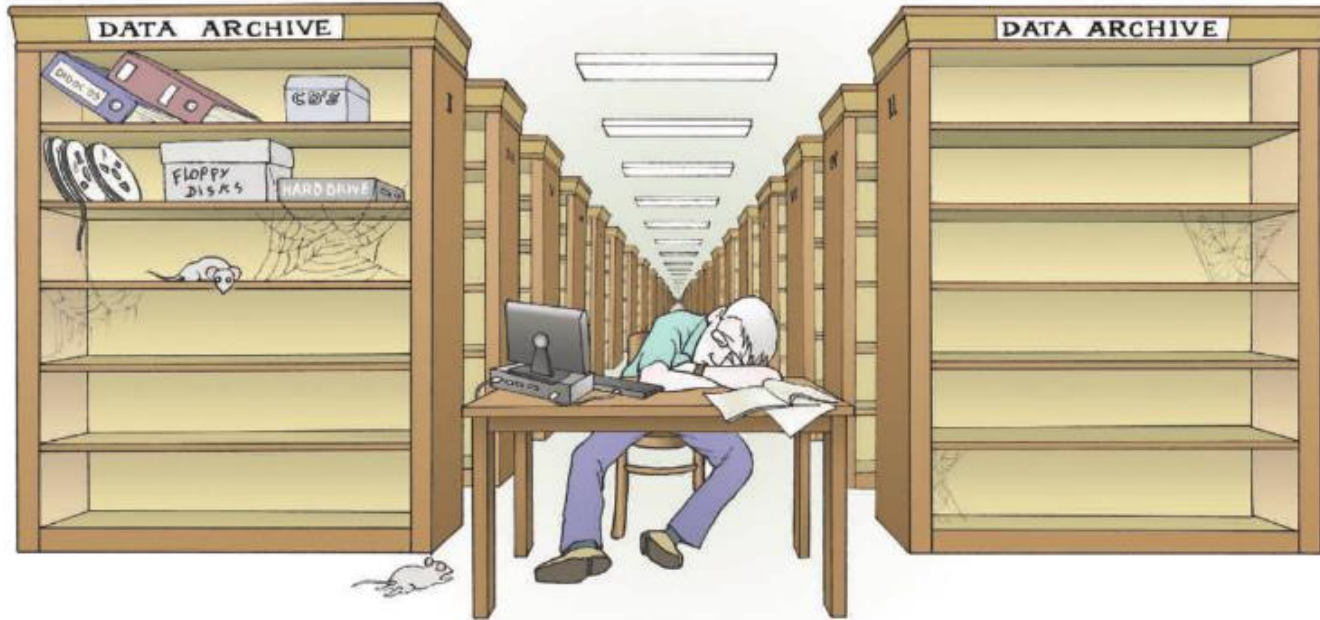
# More Policy on Open Access

- The Berlin Declaration was followed by other policies calling for open access to publicly funded research data.
- OECD Principles and Guidelines (2006) asking OECD member states to mandate Open Access.
- More policy papers followed.
- It all looked pretty good for Open Access.





# Impact of Open Access Policies on Data:



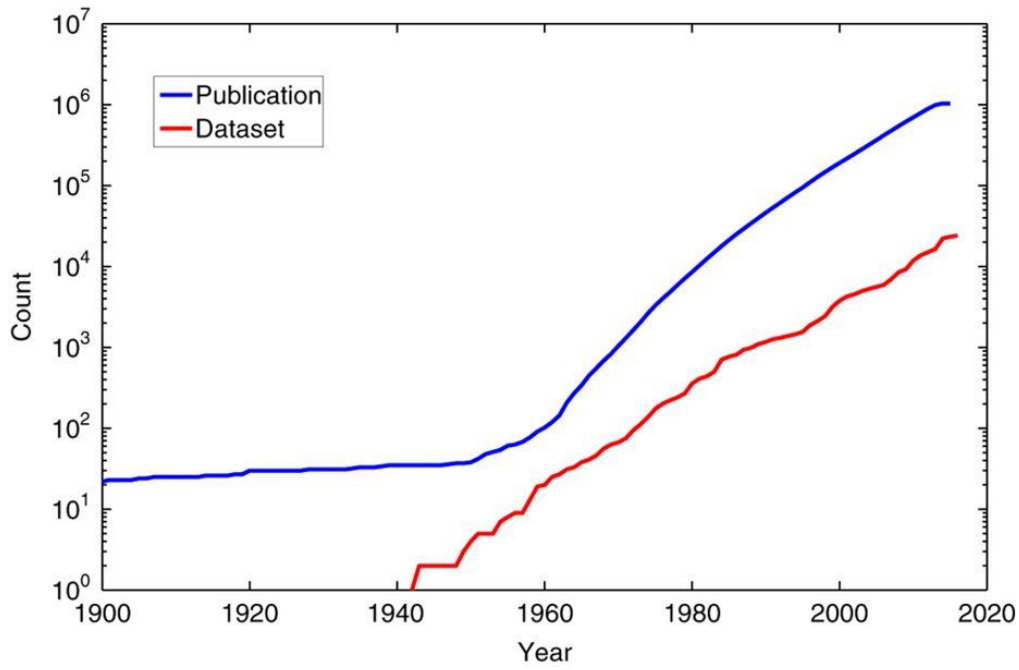
Nelson. 2009. 'Data Sharing: Empty Archives'. *Nature* 461 (7261): 160–63.  
<https://doi.org/10.1038/461160a>.



# Data Publication Rates are Still Too Low

Publication of data lags publications of papers by 90%.

Even though the number of data publications is rising exponentially, it is lagging behind the overall growth in publications.



Aryani et al. 2018



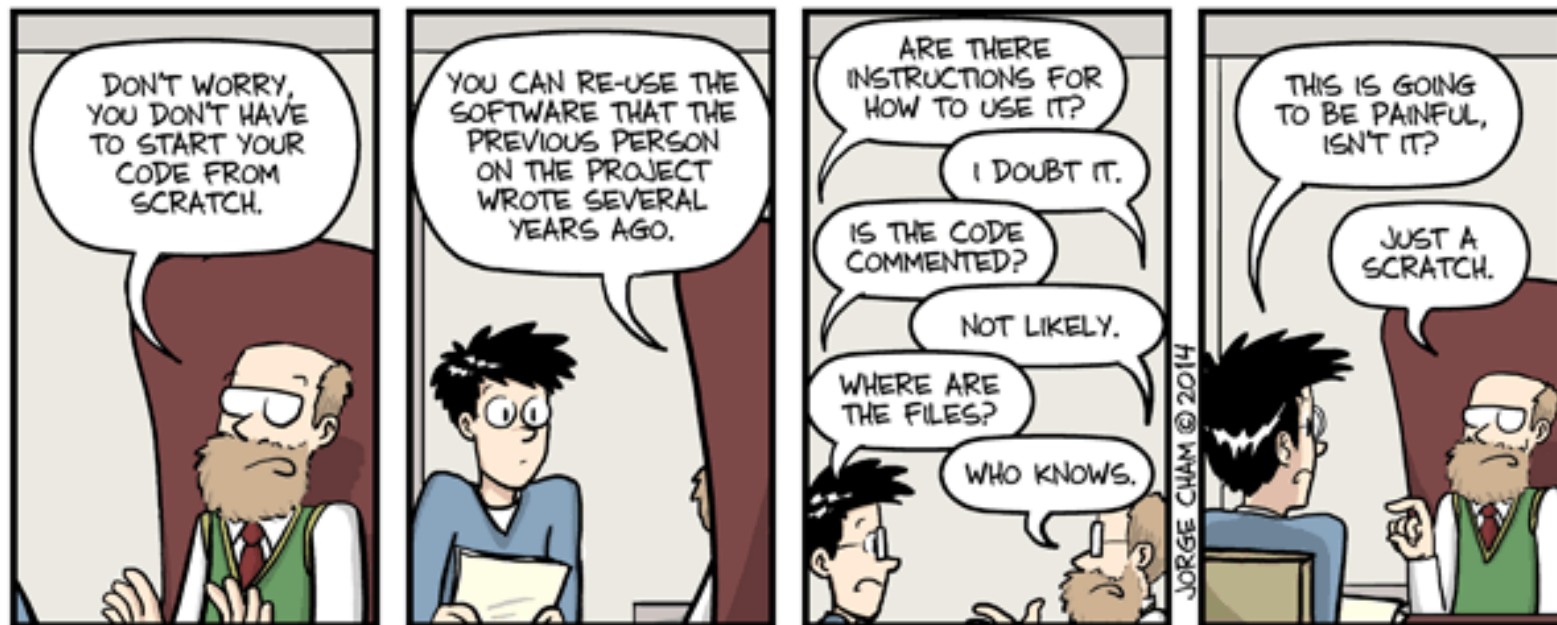
# Section 2:

# Open Source

**Free and Open Source Software for  
Research**

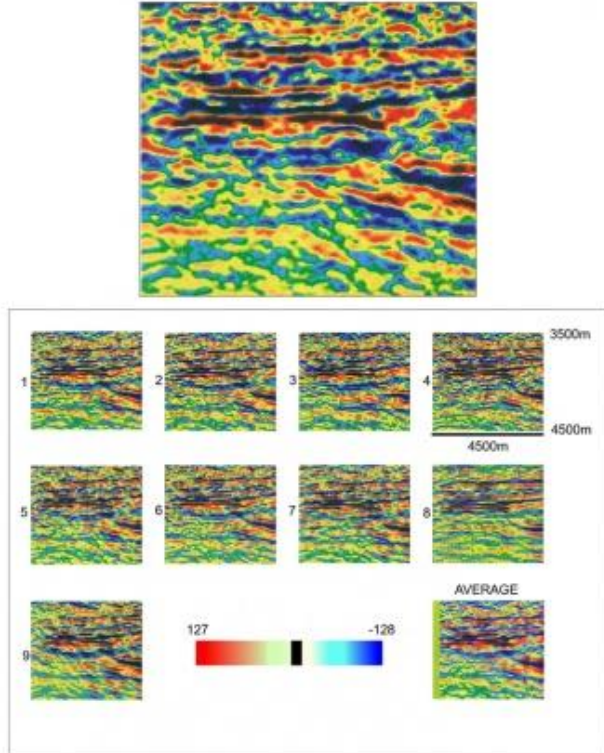


# This Is Not Reuse



WWW.PHDCOMICS.COM

# Reproducibility Crisis



- Example: seismic profiles computed on different hardware- and software stacks.
- Different system configurations and code versions produce significantly different results.
- This is in conflict with the basic principle of reproducibility of scientific results.

# Barriers to Reproducible Research

My code is not good enough for others to reuse.

I might still want to commercialise my code.

People will misinterpret my data.

Researchers only produce code for single-use.\*

I don't know where to find a repository for my code.

It's too much effort to clean up my code and data.

Will I get credited for my work?

I still want to reuse my material for further papers.

\* Wegwerf-Code

# How did it Come to This?

- Factor 1: sharing research data and code is too difficult in practice. Authors don't know what to do.
- Factor 2: There is little prestige in making data and code available. It is the journal publication that still gives you tenure.

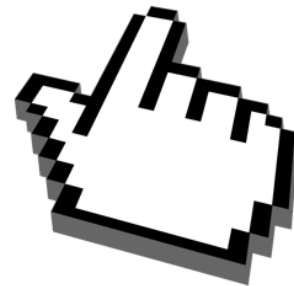


Elephant in the room: The journal publication carries the academic prestige because it showcases the "brilliant" idea that adds to a researcher's reputation.

# Section 3:

## FAIR Principles

Taking Action on Data Sharing





# Taking Action



- FAIR is not necessarily open, but it lays the foundations to making research data sharing actionable.
- The FAIR principles spell out the elements required for sharing research data at scale.
- Further discipline-specific guidance is needed.

# COPDESS and AGU FAIR Data

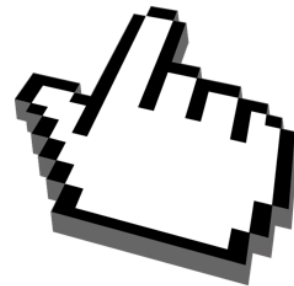
- 2014: Coalition on Publishing Data in the Earth and Space Sciences (COPDESS) Statement of Commitment.
- The Enabling FAIR Data effort and this commitment builds on the COPDESS Statement of Commitment and provides the recommendations and guidelines for implementing the full research data ecosystem.





# Section 4: Software Citation

Software in the Scholarly Record



# FORCE11 Software Citation Principles

- Importance
- Credit and Attribution
- Accessibility
- Unique Identification
- Specificity
- Persistence

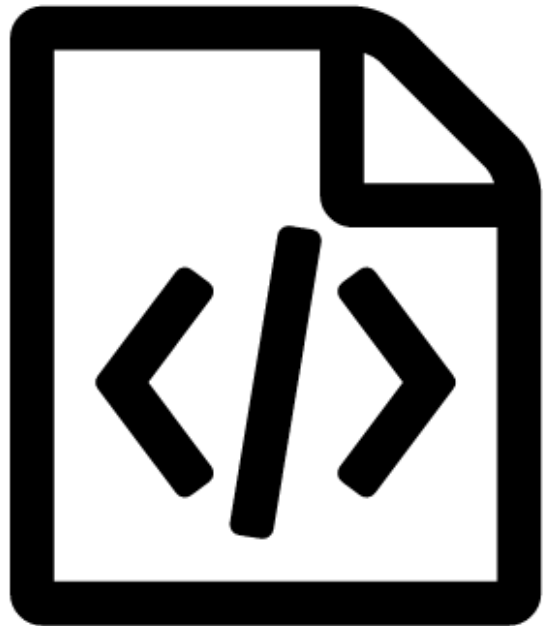


# Same, Same, But Different

- The FAIR Principles and the Software Citation Principles were both developed in the context of FORCE11, but by two separate communities.
- Both sets of principles are not yet well aligned, more work needs to be done.



# CodeMeta



- CodeMeta is a minimal metadata schema for science software and code that can be used to standardise the exchange of software metadata across repositories and organizations.
- The elements used depend on the use case for the metadata.
- <https://codemeta.github.io/>

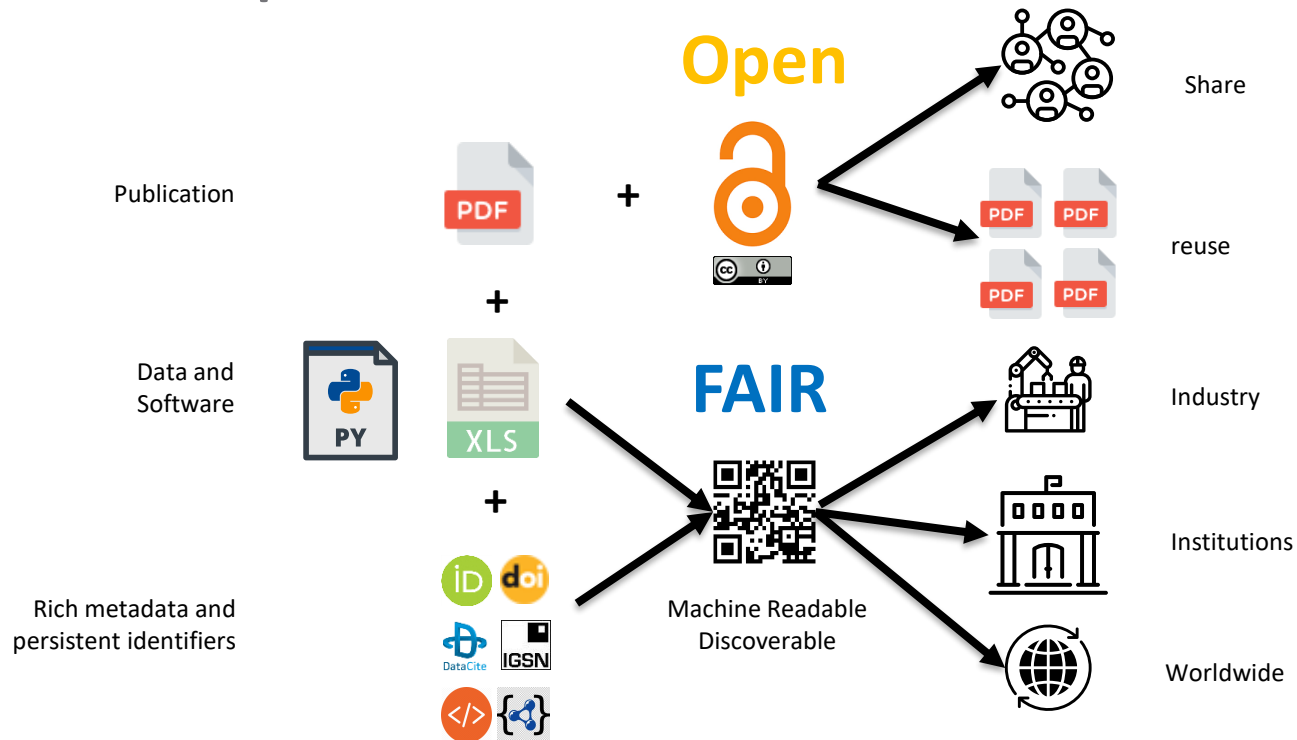
# Section 4: Next Steps

Putting it all Together



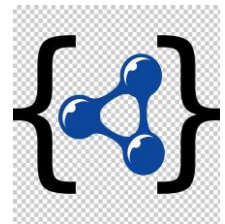


# Free – Open – FAIR



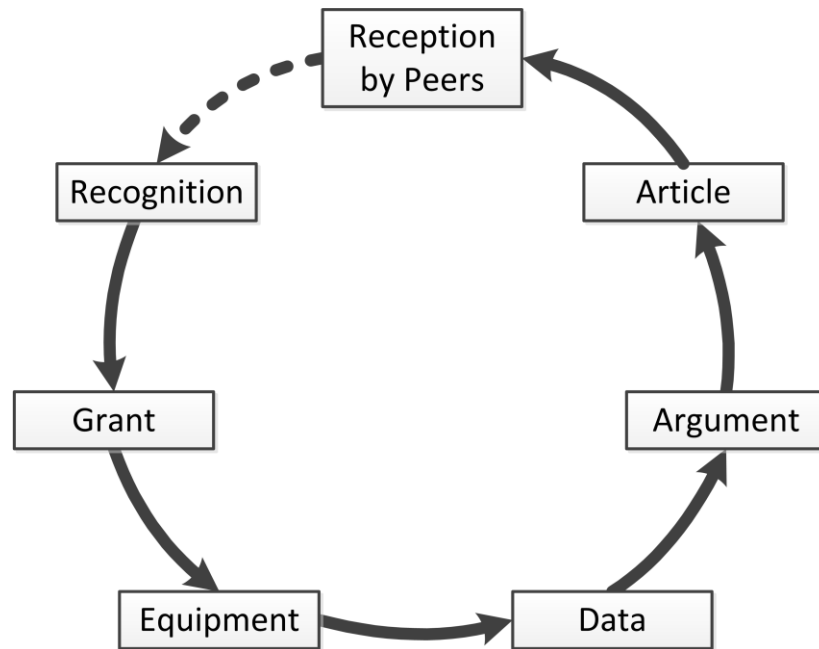
# Technical Developments

- Let the machines to the heavy lifting.
- Persistent identifiers provide anchors to people, publications, data, code, samples, instruments, ...
- Metadata become machine readable and can be harvested by standard web technologies.
- Linked data is finally coming of age (if it's done at scale and not done manually).



# Data Sharing in the Reputation Economy

- Research is a reputation economy.
- If we want to change the way research data and software are treated, we need to understand where we can intervene in the reputation economy.



# Are We There Yet?



- We're getting there.
- To change our practices around data and software we need to make them integral parts of the research ecosystem:
  - Technical integration
  - Social integration



# Thank you

## **Mineral Resources**

Jens Klump

Team Leader Geoscience Analytics

+61 8 6436 8828

[Jens.klump@csiro.au](mailto:Jens.klump@csiro.au)

[csiro.au/en/Research/MRF](http://csiro.au/en/Research/MRF)

