Transparency, Reproducibility and Impact: Placing Open Scholarship Into Practice

Cara Bradley, Kaetlyn Phillips and Christina Winter

This work is an adaptation of the presentation "Transparency, Reproducibility and Impact" by Simone Sacchi, licenced under a Creative Commons CC-BY 4.0 International Licence.

Introductions

Before we start...

This session is intentionally an overview

...intended to trigger your curiosity!

Let’s begin

How many of you are already familiar with the general idea of **Open Scholarship**?

What do you think about this definition?

Open scholarship is the application of open practices throughout the teaching, learning, research and scholarly environment. Examples of open practices include:

- The use of open copyright licenses that grant permission for others to freely access, reuse, redistribute, or build upon your scholarly work.
- Using networks, web-based tools, and the Internet to facilitate collaboration.
- Making research processes, data, and code transparent in order to increase the reproducibility and accessibility of research.
- Creating and using educational resources that are free of costs and copyright restrictions in order to lower student costs and access barriers for learning materials such as textbooks.

- Integrating principles of social justice into academic processes in order to make scholarship more equitable and inclusive.
- Engaging in pedagogical practices where learners are creators of knowledge and have agency in their own education.
- Connecting and engaging with communities, networks, and the broader public.

Open scholarship is an umbrella term which encompasses open access, open research, open science, open data, open education, open pedagogy and all other forms of openness in the scholarly and research environment.

- UBC Program for Open Scholarship and Education, https://pose.open.ubc.ca/home-page/getting-started/what-is-open-scholarship/

Let’s start with a **short story**…. (which may sound familiar)

A pretty **common story**...

*A researcher intends to **answer her research question** by proving certain hypotheses.*

She **collects data** and **designs her methodology** along the way. She also **writes/modifies some** code to analyse that data, saving everything on her laptop (occasionally backed up by her department)

Once she has some results, she **writes and submits her paper** to a journal. She might include her data – a growing number of journals require this – but she probably doesn't include her code.

Time passes...

*The journal sends her **anonymous reviews**.*

She **revises her paper** to satisfy them, she might also **modify the code** she wrote earlier, and resubmits.

More time passes...

*The **paper is eventually published**.*

It might include a link to an online copy of her data, but the paper itself is behind a **paywall**: only people who have personal or institutional access will be able to read it.

Do you see any **issue**?

- **Reproducibility and trust** with empirical social science research
- **Lack of full disclosure and openness** can lead to distortion by false-positive results, data mining and tendentious reporting, etc.
- **Equitable and timely access** to (publicly funded) research
  - By the academic community for the academic community
  - For the society (including business & innovation) to address global challenges
- **Broadest reach, visibility and impact** for individual researchers
    - Helps with their career
    - Boosts collaboration
  - Last but not least: **what about safekeeping your work**?

Let’s see how we can season this story with some **Open Scholarship** sauce

(and address some of the issues above)

A **revised** approach [1]

*She **collects data** and **designs her methodology** along the way. She also **writes/modifies some** code to analyse that data, saving everything on her laptop (occasionally backed up by her department)*

- The researcher **pre-registers her study** and **prepares a DMP**
- She creates a **version control repository** for her work (e.g. GitHub):
  - As data are collected, **she uses the repository to safekeep the data**
  - As she perfects her analysis, **she pushes changes to her code to that repository**
  - The repository also becomes **her hub for collaboration**

Pre-register your study

What is a **pre-registration**?

A researcher **posts her/his research hypotheses and the planned research design** (i.e., methodology) in a publicly available registry (e.g. AsPredicted, OSF) ahead of the beginning of the project

- Sharing research design + plans beforehand complements disclosures and openness of research outputs
- It takes time? Yes... It takes more time? No...
- Mostly applies to confirmatory empirical research, but...
- It does not discourage exploratory and/or observations research!

Why pre-registration?

- Study **pre-registration** carries a series of benefits:
    - Rounds out the body of evidence by creating a “paper trail” of unpublished studies in an area
    - Reduces the risk of data mining and other tendentious presentation of results (“cherry-picking”)
    - Makes open data and disclosure more effective allowing to cross-check published claims against original research plans
    - Encourages scholars to more carefully think through their design increasing research quality and reducing “waste” of funding on poorly conceived projects.

Plan the management of your research data

Research Data **Management**

Research Data Management (RDM) is a general term that covers how you organize, structure, store, and care for the information used or generated during a research project.

- **Planning how your data will be looked after** – many funders now require data management plans as part of applications
- **How you deal with information on a day-to-day basis** over the lifetime of a project
- **What happens to data in the longer term** – what you do with it after the project concludes

Data Management **Plans**

- **Data Management Plans (DMPs)** are documents that describe:
  - How data is **generated,** used, **elaborated,** and **organised**
  - How data, and data subjects, are protected
  - How data, code, and ancillary elements are described and documented
  - How data is stored and secured, and how long it will be retained
  - How dataset authorship and credit are assigned
  - How data is preserved and under what terms, can be shared
- DMPs are more and more **required by funders and publishers**

See also The Research Data Lifecycle by Thomas Bourke, EUI Library: https://www.eui.eu/Research/Library/ResearchDataServices/StepByStep

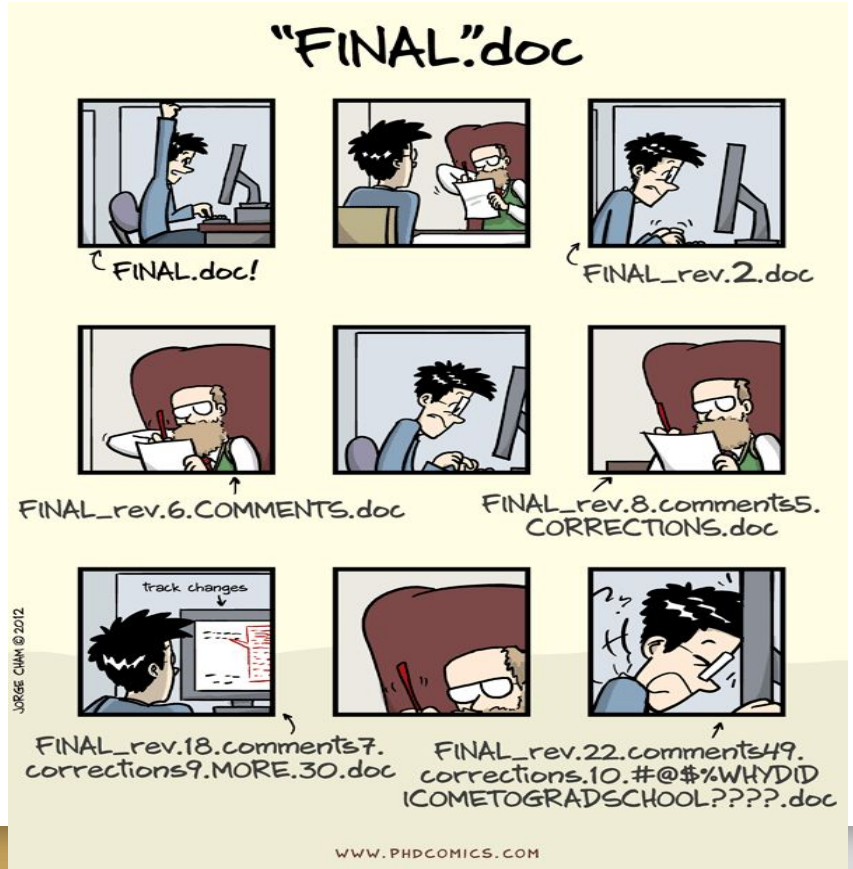
Keep track of changes in your work and open up your process

Version control for research

We’ve all been in this situation

at some point….

It seems ridiculous to have multiple nearly-identical versions of the same document…. *Yet we all do it!!!!!!*



“Piled Higher and Deeper” by Jorge Cham, http://www.phdcomics.com

VCSs **benefits**

- A version control system (VCS) **keeps track of changes** for us, creating different versions of files in a repository
- Online services (e.g. GitHub, GitLab, BitBucket), allow repositories to be kept in sync across computers, **facilitating collaboration** among different people (or different YOUs).
- Keeping an (open) track of changes in your project allows others to **understand your research process**, building trust

Version control is like an unlimited 'undo': you never lose anything and can always roll back

A **revised** approach [2]

*Once she has some results, **she writes and submits her paper** to a journal. She might include her data – a growing number of journals require this – but she probably doesn't include her code.*

- As soon as data are stable **she stores the final dataset in an open repository** (e.g. Borealis Dataverse or Cadmus or Zenodo)

  - so it is assigned a **persistent identifier** (e.g. DOI)
- When she's happy with her paper, **she posts a version in a preprint server** (e.g. SocArXiv) with links to data and code to invite feedback from peers.

  - She may post several revisions before finally submitting her paper to a journal. She may also share a reproducible version

Register (and store) your data in an open repository

Why **register your data**?

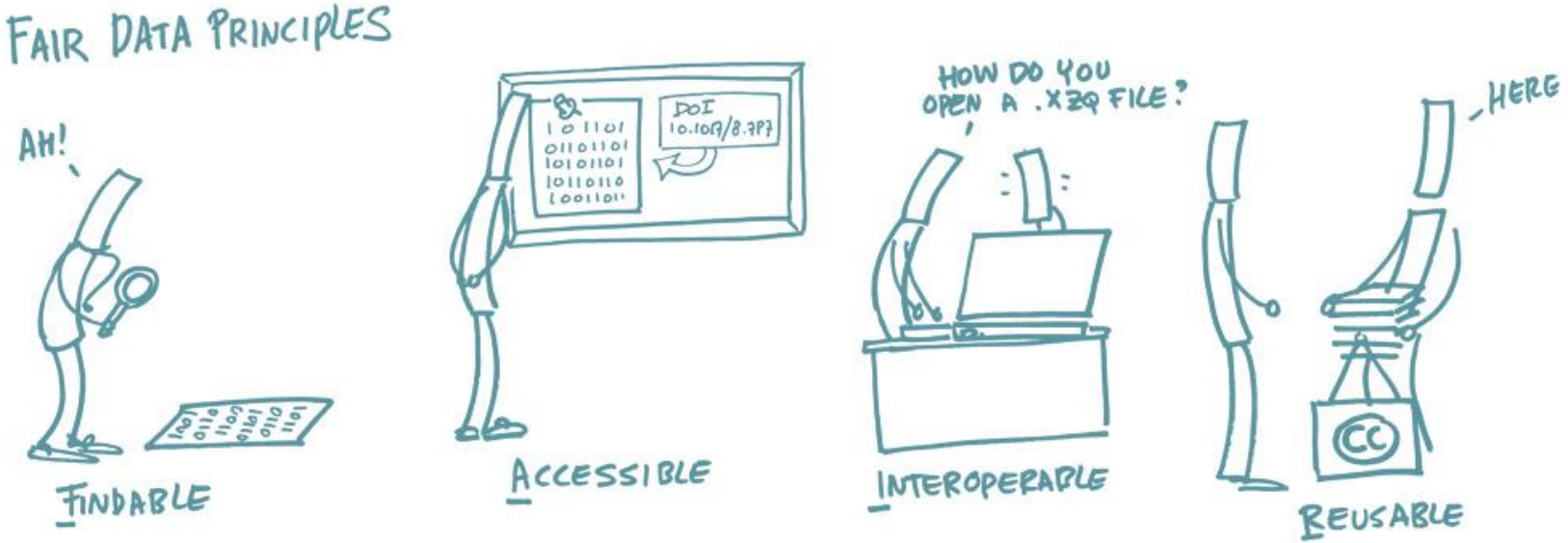
- If you **register and store your research data in an (open) repository** you make them first class citizens in the scholarly communication landscape

- - - -

They are assigned a **persistent identifier**
They are **curated and preserved**
They can be **cited**
They can be **reused** and you be credited for that

Remember: as open as possible but as closed as necessary

The **FAIR** principles



Preprints and Working Papers are your friends

Why sharing a **preprint**

- Preprints are **version(s) of papers shared before an "official" submission** and (eventual) publication in a journal
- Very common in certain fields (e.g. Physics) are becoming a fact also in the Social Sciences
    - They carry a "**time stamp**" and provide immediate access
    - Make the **"trail" of your research more transparent**
    - While waiting for the "official" peer reviews, **research is already discussed, visible, and alive** in the academic community
    - No issues in being accepted, no embargo, no delay
- Check a journal's policy on preprints - example https://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/overlapping-publications.html

A **revised** approach [3]

The journal sends her anonymous reviews. She revises her paper to satisfy them, she might also modify the code she wrote earlier, and resubmits.

- Our researcher most likely will be following a similar path, but:
    - Remember that **her preprint is already out there** making an impact
    - If she changes her code, **the history of changes is tracked** in her version control repository
    - She might have submitted to a journal that adopts an **open peer review** approach

A **revised** approach [4]

The paper is eventually published. It might include a link to an online copy of her data, but the paper itself is behind a paywall: only people who have personal or institutional access will be able to read it.

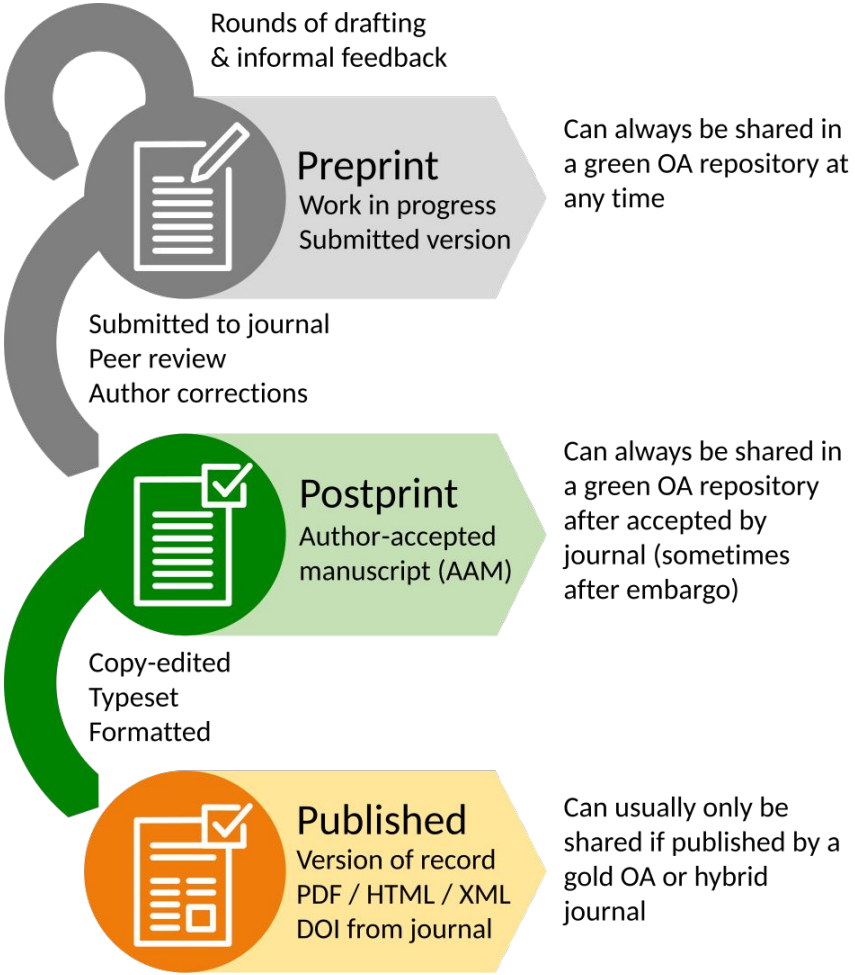
- When the paper is published (possibly in an Open Access journal), it includes links to her preprint, to her **data**, her **code** repository, and to her **study pre-registration**
- This makes it much **easier for other researchers to use her work** as starting point for their own research and validate her findings.
- Her research is **transparent, reproducible, and impactful**

A first peek into

Open Access

This can be considered a first step towards Open Access

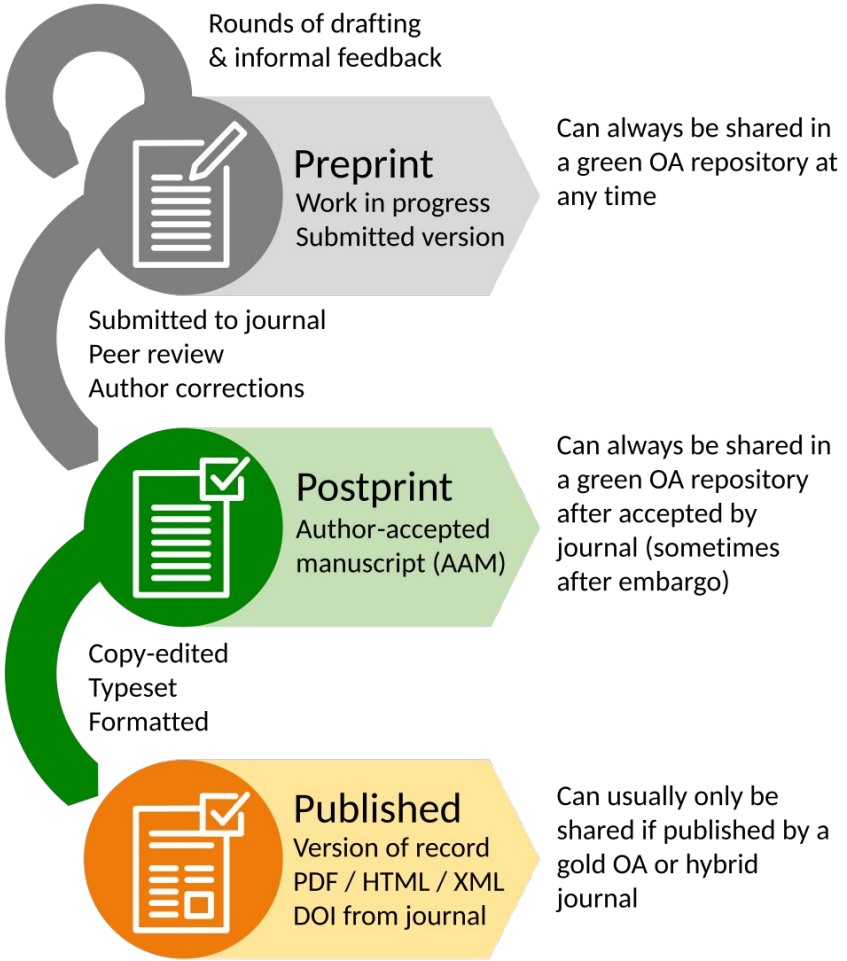
"Open Access (OA) is the free, immediate, online access to the results of scholarly research, and the right to use and re-use those results without most permission barriers."



Open Access **options**

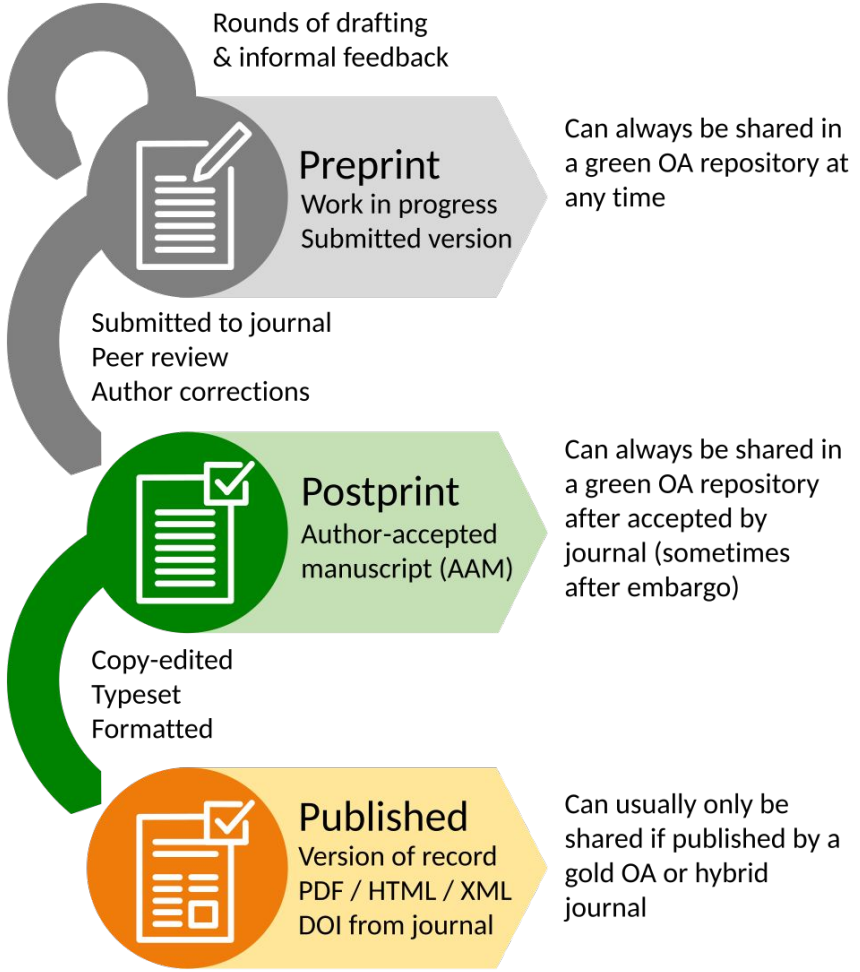
- Let's consider again this publishing workflow with Open Access options and the definition

"Open Access (OA) is the free, immediate, online access to the results of scholarly research, and the right to use and re-use those results without most permission barriers."



Open Access **benefits**

- Immediate visibility and dissemination
- Equitable access to knowledge
- Accelerated discovery
- Greater transparency
- Increased possibility to be cited
- Public enrichment
- Compliancy with funder requirements
- Greater opportunities for collaboration
- Greater potential impact



Is Open Access **free**?

- Free at the point of read/use
- Nothing is free in absolute terms, different business models
- Business models:

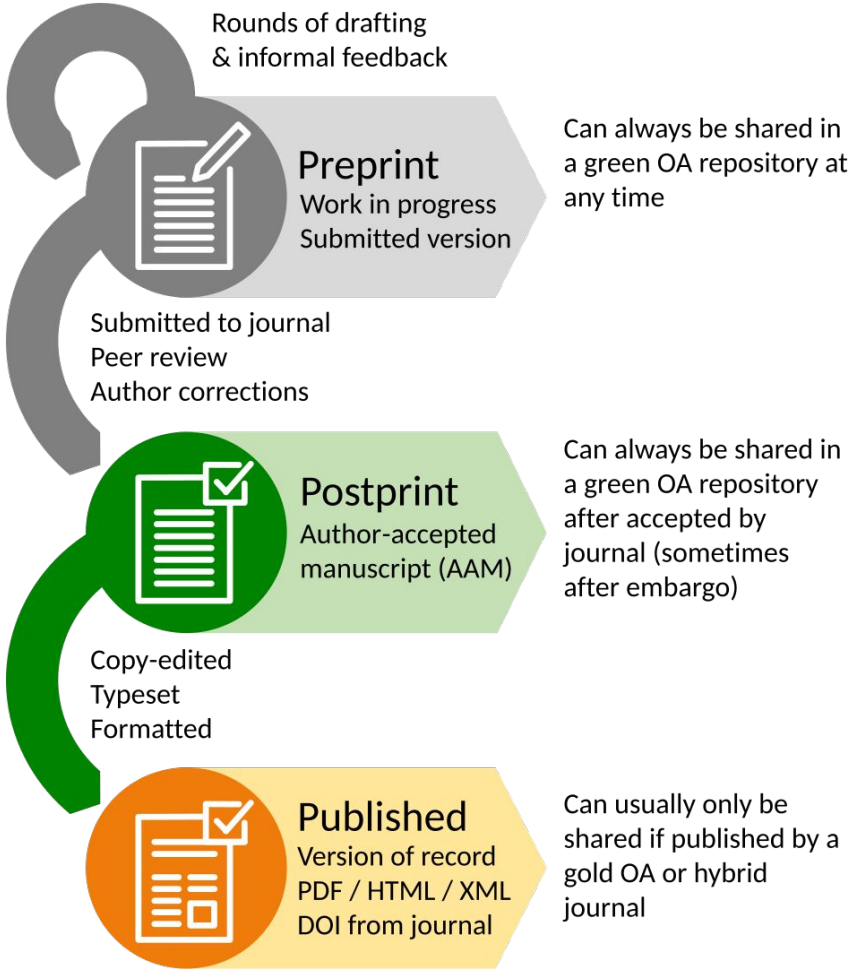
○ **Gold** (Article Processing charges)

Author pays (set aside money if you can)
Look also for agreements publishers/libraries!

○ **Platinum/diamond**

Subsidized costs

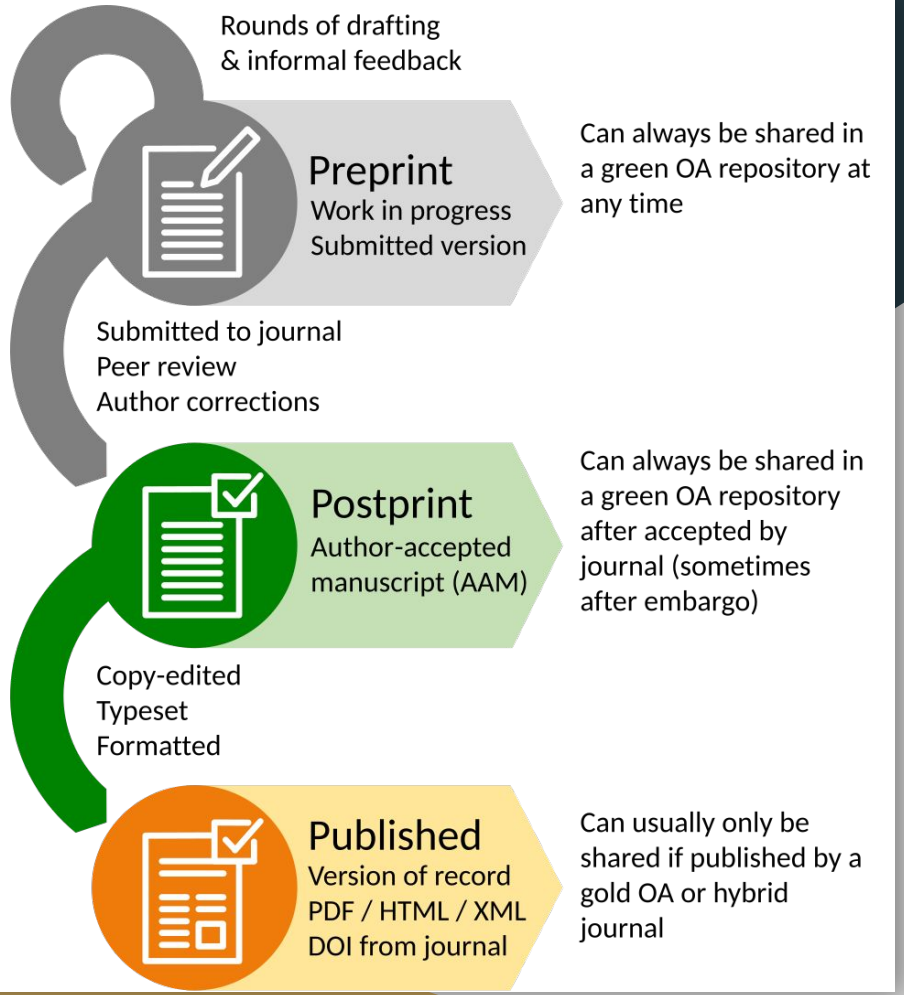
○ **Mixed** (e.g. PeerJ)



And **green** Open Access?

- Considered the “best option” by many advocates
- Share your postprint in an institutional/disciplinary repository
- Free for author/reader (but infrastructure costs are there!)
- Does not “feed” the commercial publishing system

Might carry a embargo



Let’s take a breath...

Open Scholarship **is** (and **is not**)…

Open Scholarship *is not a goal (in itself)*
Open Scholarship *is a means to achieve a spectrum of goals*

- For **you**

  - Make your work more transparent, and visible (and citable), better engage with the research community, increase collaboration

- For the **research community**(ies)
  - Foster knowledge discovery (efficiency) and reduce barriers (equity), increase reproducibility (trust) ● For the **society** at large

  - Boost innovation & economic growth, save public money, contribute to making a better world

Open Scholarship: **challenges**

...or why it is not yet “just scholarship”?

- Cultural, technical, legal barriers
    - “This is how I have always done it / how I was taught to do it”
    - “I do not have time for this, what do I get out of it?””
    - “Idonotknowhowtodoit(butIamnotgoingtotellyou)”
    - “I already have all the tools I need within my own community”
    - “I gave away my copyright / I cannot do it legally”
    - “What if [...]?”
- But times are changing rapidly...

Actually, we just **scratched** the surface...

Image credits

Slide 26: Sonja Bezjak, April Clyburne-Sherin, Philipp Conzett, Pedro Fernandes, Edit Görögh, Kerstin Helbig, Bianca Kramer, Ignasi Labastida, Kyle Niemeyer, Fotis Psomopoulos, Tony Ross-Hellauer, René Schneider, Jon Tennant, Ellen Verbakel, Helene Brinken, & Lambert Heller. (2018). Open Science Training Handbook (1.0). https://book.fosteropenscience.eu/

Slide 29: Image by Thomas Shafee, reused under a CC-BY 4.0 International, licence https://commons.wikimedia.org/wiki/File:Preprint_postprint_published.svg

Slide 33: Ibid

Slide 34: Ibid

Questions and Discussion