TRANSPARENCY, REPRODUCIBILITY AND IMPACT

PLACING OPEN SCIENCE INTO PRACTICE

Simone Sacchi, Ph.D.
Open Science Librarian
European University Institute

Simone Sacchi, Ph.D.
Open Science Librarian
European University Institute

19 October 2022

EUROPEAN UNIVERSITY INSTITUTE

CIVICA
THE EUROPEAN UNIVERSITY OF SOCIAL SCIENCES

Before **we start**…

Welcome everyone!!!!

Before **we start**…

This training is intentionally an overview

…intended to trigger your curiosity!

EUI EUROPEAN UNIVERSITY INSTITUTE

CIVICA
THE EUROPEAN UNIVERSITY OF SOCIAL SCIENCES

Before **we start**...

Being a hybrid session, if you have questions please either raise your hand or type the questions in the chat

Let's **begin**!

How many of you are already <u>familiar</u> with the general idea of **Open Science**?

How do you **feel** about this definition?

"Open science describes the practice of carrying out scientific research in a completely transparent manner, and making the results of that research available to everyone. Isn't that just 'science'?"

Watson M., When will "open science" become simply "science"? Genome Biol. 2015;16:101

How do you **feel** about this definition?

*"Open science describes the practice of carrying out scientific research in a **completely transparent manner**, and making the **results of that research available to everyone**. Isn't that just 'science'?"*

Let's start with a **short story**....
(which may sound pretty 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 Science** sauce

(and address some of the issues above)

A **revised** approach [1]

*She **collects data** and **design 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 they are collected, **she uses the repository to safekeep her 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. 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
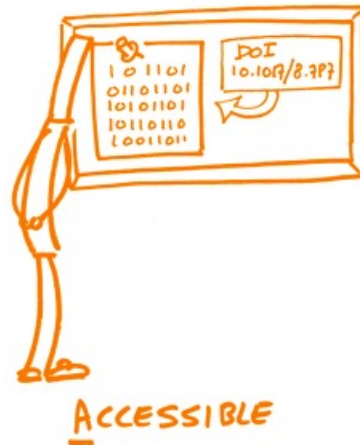
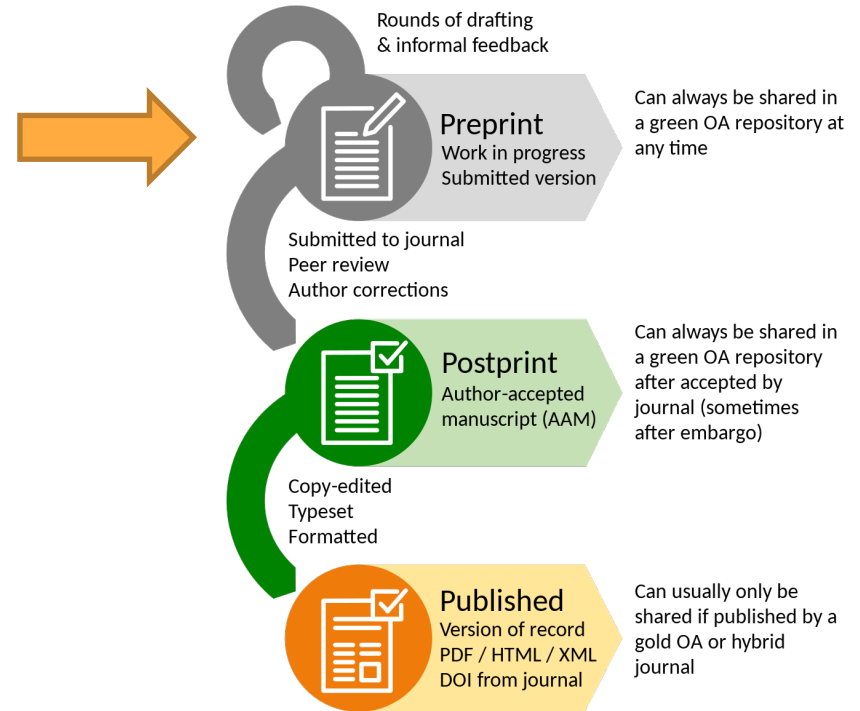Preprints and Working Papers are your friends

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
- Almost all journals now accept that preprints (or working papers) are shared early on
  - No issues in being accepted, no embargo, no delay

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."

Rounds of drafting & informal feedback

Preprint
Work in progress
Submitted version

Can always be shared in a green OA repository at any time

Submitted to journal
Peer review
Author corrections

Postprint
Author-accepted manuscript (AAM)

Can always be shared in a green OA repository after accepted by journal (sometimes after embargo)

Copy-edited
Typeset
Formatted

Published
Version of record
PDF / HTML / XML
DOI from journal

Can usually only be shared if published by a gold OA or hybrid journal

CIVICA
THE EUROPEAN UNIVERSITY OF SOCIAL SCIENCES

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

CIVICA
THE EUROPEAN UNIVERSITY OF SOCIAL SCIENCES

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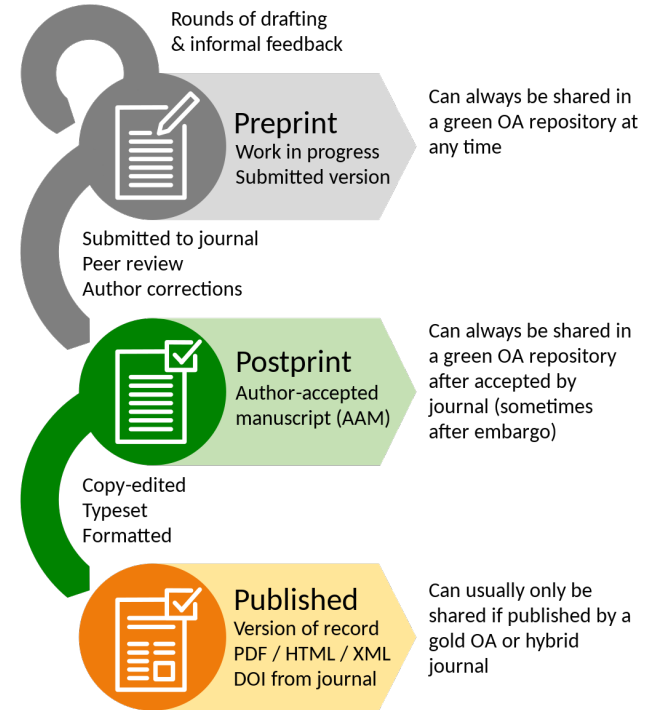
- When the paper is published (possibly in an Open Access journal), it includes links to her preprint, to her **data**, her **code** repository, and to her **study pre-registration**
- This makes it much **easier for other scientists to use her work** as starting point for their own research and validate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

Open Access publishing
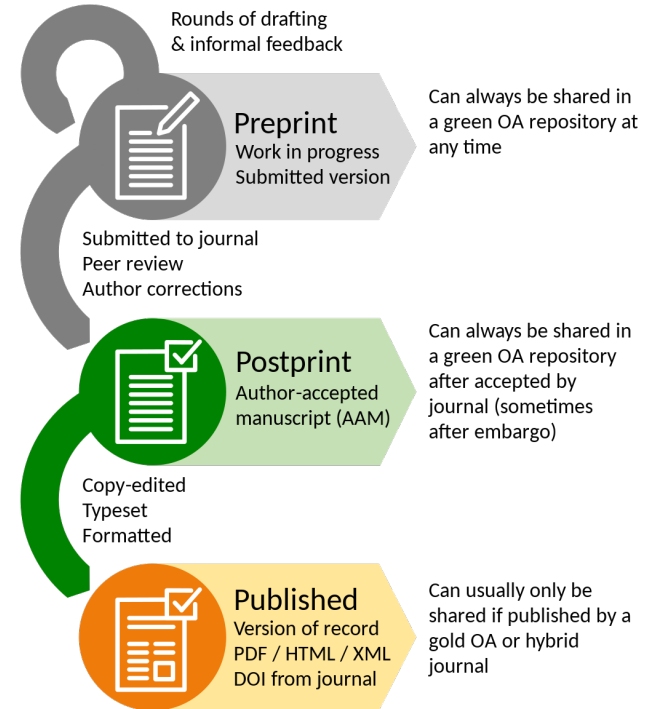
Open Access **options**

- 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."

Rounds of drafting & informal feedback

Preprint
Work in progress
Submitted version

Can always be shared in a green OA repository at any time

Submitted to journal
Peer review
Author corrections

Postprint
Author-accepted manuscript (AAM)

Can always be shared in a green OA repository after accepted by journal (sometimes after embargo)

Copy-edited
Typeset
Formatted

Published
Version of record
PDF / HTML / XML
DOI from journal

Can usually only be shared if published by a gold OA or hybrid journal

35

CIVICA
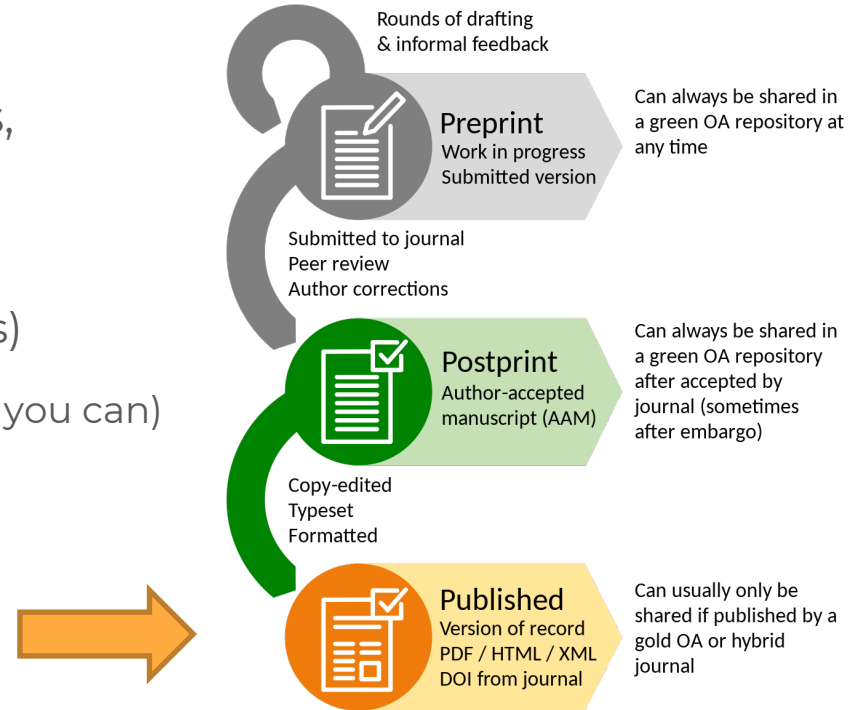THE EUROPEAN UNIVERSITY OF SOCIAL SCIENCES

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

Rounds of drafting & informal feedback

Preprint
Work in progress
Submitted version

Can always be shared in a green OA repository at any time

Submitted to journal
Peer review
Author corrections

Postprint
Author-accepted manuscript (AAM)

Can always be shared in a green OA repository after accepted by journal (sometimes after embargo)

Copy-edited
Typeset
Formatted

Published
Version of record
PDF / HTML / XML
DOI from journal

Can usually only be shared if published by a gold OA or hybrid journal
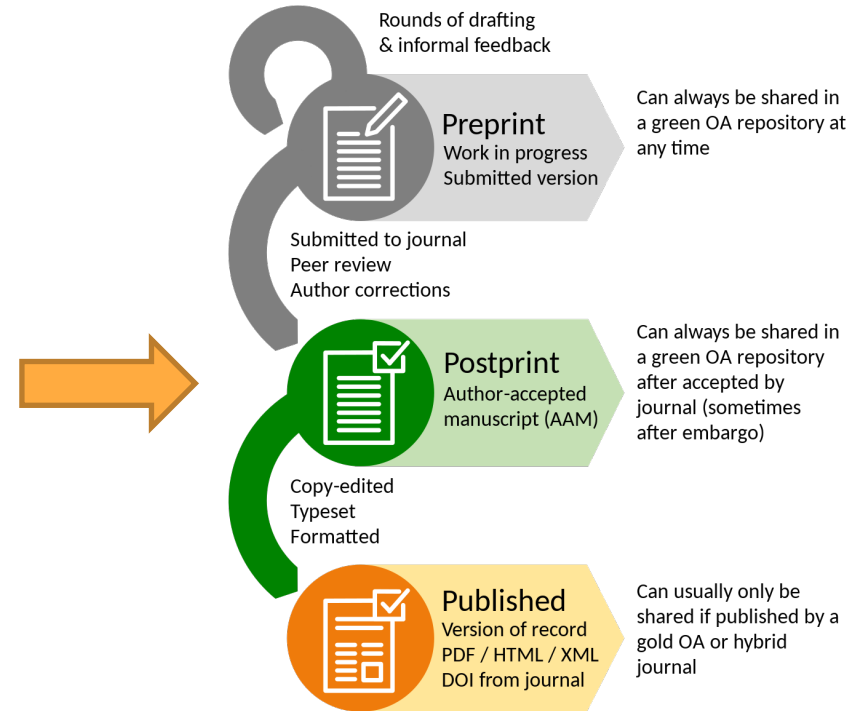
Is Open Access **free**?

- 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)

Rounds of drafting & informal feedback

Preprint
Work in progress
Submitted version

Can always be shared in a green OA repository at any time

Submitted to journal
Peer review
Author corrections

Postprint
Author-accepted manuscript (AAM)

Can always be shared in a green OA repository after accepted by journal (sometimes after embargo)

Copy-edited
Typeset
Formatted

Published
Version of record
PDF / HTML / XML
DOI from journal

Can usually only be shared if published by a gold OA or hybrid journal

CIVICA
THE EUROPEAN UNIVERSITY OF SOCIAL SCIENCES

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

Rounds of drafting & informal feedback

Preprint
Work in progress
Submitted version

Can always be shared in a green OA repository at any time

Submitted to journal
Peer review
Author corrections

Postprint
Author-accepted manuscript (AAM)

Can always be shared in a green OA repository after accepted by journal (sometimes after embargo)

Copy-edited
Typeset
Formatted

Published
Version of record
PDF / HTML / XML
DOI from journal

Can usually only be shared if published by a gold OA or hybrid journal

Let's take a breath...

Open Science **is** (and **is not**)...

Open Science is *not* a goal (in itself)
Open Science 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 Science: **challenges**

...or why it is not yet "just science"?

- 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?""
  - "I do not know how to do it (but I am not going to tell you)"
  - "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…

You can make your workflow more open by …
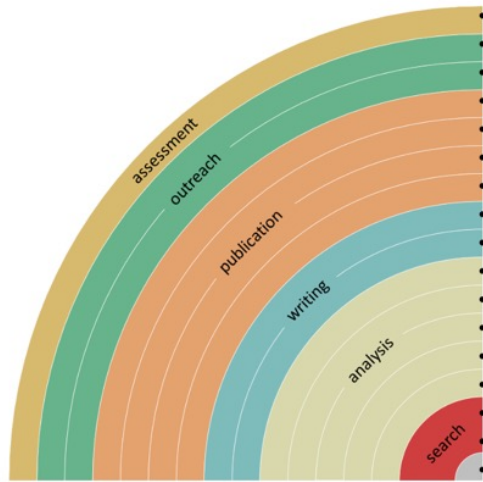
- adding alternative evaluation, e.g. with altmetrics
- communicating through social media, e.g. Twitter
- sharing posters & presentations, e.g. at FigShare
- using open licenses, e.g. CC0 or CC-BY
- publishing open access, 'green' or 'gold'
- using open peer review, e.g. at journals or PubPeer
- sharing preprints, e.g. at OSF, arXiv or bioRxiv
- using actionable formats, e.g. with Jupyter or CoCalc
- open XML-drafting, e.g. at Overleaf or Authorea
- sharing protocols & workfl., e.g. at Protocols.io
- sharing notebooks, e.g. at OpenNotebookScience
- sharing code, e.g. at GitHub with GNU/MIT license
- sharing data, e.g. at Dryad, Zenodo or Dataverse
- pre-registering, e.g. at OSF or AsPredicted
- commenting openly, e.g. with Hypothes.is
- using shared reference libraries, e.g. with Zotero
- sharing (grant) proposals, e.g. at RIO

assessment, outreach, publication, writing, analysis, search

Bianca Kramer & Jeroen Bosman https://101innovations.wordpress.com

DOI: 10.5281/zenodo.1147025

any **baby step** matter…

That's all folks!
(for today)

Questions?

(you can also email me at simone.sacchi@eui.eu)

Resources [1]

Open Science (general):

- EUI Library Open Science web pages: https://www.eui.eu/Research/Library/PublishingAndOpenScience
- Open Science Training Handbook: https://open-science-training-handbook.github.io/Open-Science-Training-Handbook_EN//
- McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., … Yarkoni, T. (2016). How open science helps researchers succeed. *ELife*, *5*, e16800. https://doi.org/10.7554/eLife.16800

Ethics, transparency and reproducibility of research:

- EUI Ethics and Integrity in Academic Research: https://www.eui.eu/ServicesAndAdmin/DeanOfStudies/Ethics-and-Integrity-in-Academic-Research
- Miguel, E., The Scientific Ethos, Misconduct and Transparency (BITSS RT2 2021): https://osf.io/kh2g3/
- Christensen, G., Freese, J., and Miguel, E. (2019). Transparent and Reproducible Social Science Research: How To Do Open Science. University of California Press, Oakland. https://opac.eui.eu/client/en_GB/default/search/detailnonmodal/ent:$002f$002fSD_ILS$002f0$002fSD_ILS:480702/one

Resources [2]

Pre-registration and Pre-Analysis Plan

- Miguel, E., Pre-registration and Pre-Analysis Plans (BITSS RT2 2021): https://osf.io/gqywv/
- Examples of registries/repositories:

  https://aspredicted.org/

  https://osf.io/registries

Version control, Git, Github (with a touch of reproducible writing):

- Version Control with Git: https://swcarpentry.github.io/git-novice/
- Kuriwaki, S., Git for Students in the Social Sciences https://github.com/kuriwaki/github-demo
- Happy Git and GitHub for the user https://happygitwithr.com/
- Github Desktop: https://desktop.github.com/
- Healy, K., The Plain Person's Guide to Plain Text Social Sciences https://plain-text.co/
- R Markdown: The Definitive Guide: https://bookdown.org/yihui/rmarkdown/
- Gandrud, C. (2020). Reproducible Research with R and RStudio. CRC Press, Boca Raton, FL, 3rd edition. https://opac.eui.eu/client/en_GB/default/search/detailnonmodal/ent:$002f$002fSD_ILS$002f0$002fSD_ILS:499575/one

Resources [3]

Research Data Management and Data Management Plans
- EUI Library Research Data Services: https://www.eui.eu/Research/Library/ResearchDataServices
- EUI Guide to Good Data Protection Practice in Research: https://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/Guide-Data-Protection-Research.pdf
- EUI Guidelines on Data Protection: https://www.eui.eu/About/DataProtection
- DMPonline: http://www.dcc.ac.uk/dmponline
- DMP Evaluation Rubric: https://scienceeurope.org/media/4brkxxe5/se_rdm_practical_guide_extended_final.pdf
- FAIRaware tool: https://fairaware.dans.knaw.nl/

Open Access and Publications:
- OS office pages on TAs https://www.eui.eu/Research/Library/PublishingAndOpenScience/OAPublishing-LibraryAgreements
- Sherpa-Romeo tool to check OA policies of journal titles and publishers: https://v2.sherpa.ac.uk/romeo/
- ORCID how to connect to EUI: https://cadmus.eui.eu/ORCID/