CSC

ICT Solutions for Brilliant Minds

Reproducibility - Rethinking the research practices

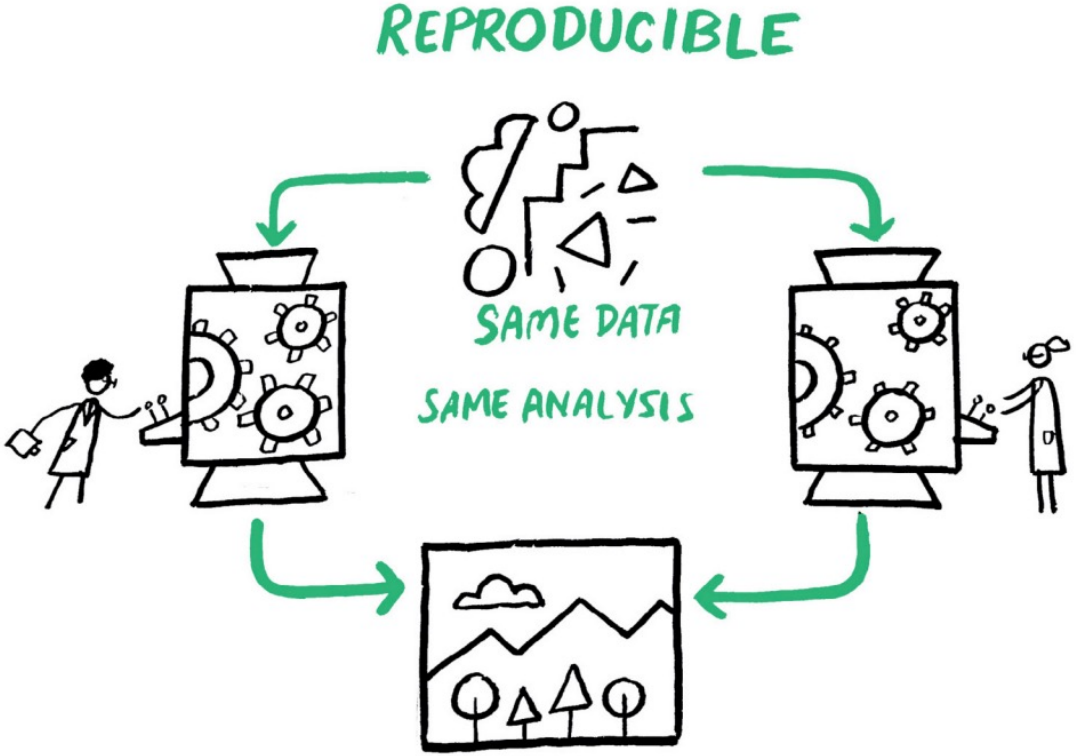
Josefine Nordling, Open Science Specialist

EOSC Finnish Forum 08.06.2022

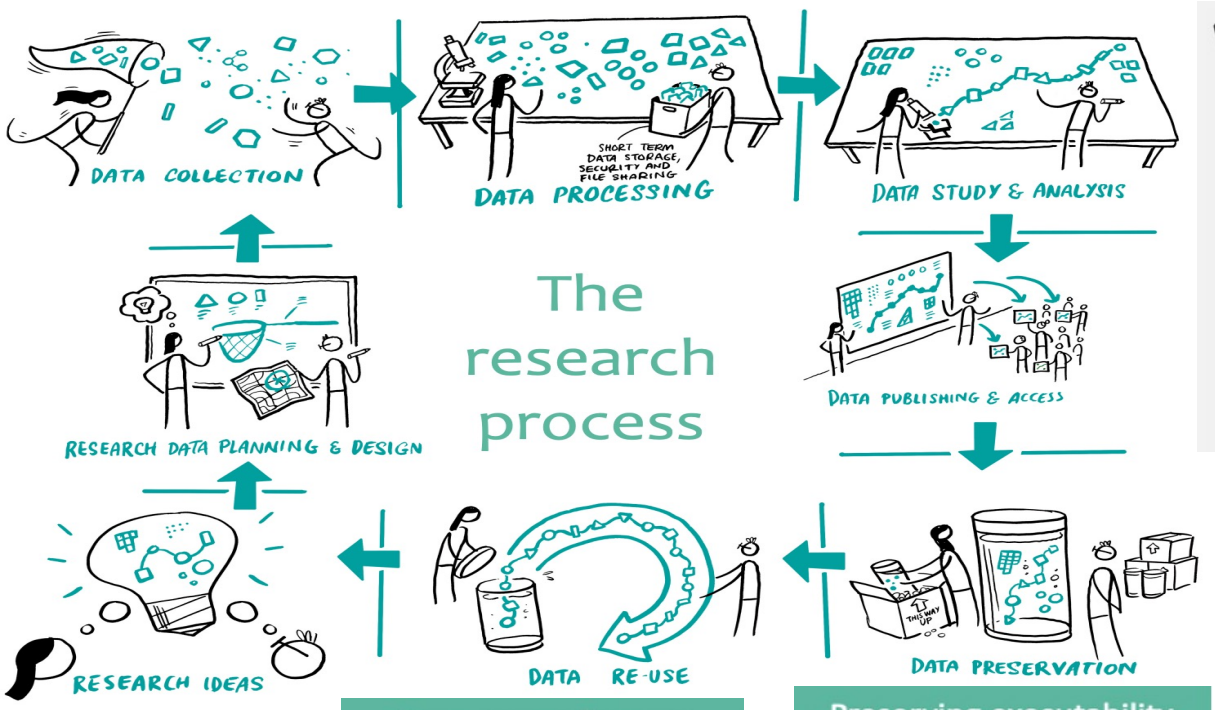
Presentation outline

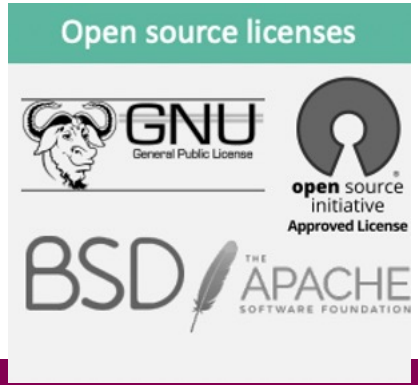
- Reproducibility – what, where & when?
- Reproducibility crisis
- Knowledge Exchange report on reproducible research
  - Changing the research practices
  - Benefits & challenges
  - Tools & infrastructures
  - Key take-away messages
- EOSC Service Portfolio & services supporting reproducibility

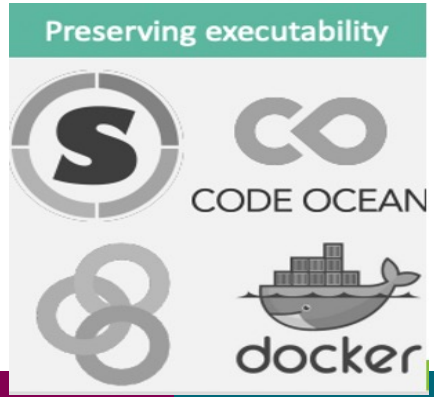
What reproducibility means



Source: The Turing Way Community., & Scriberia. (2021). Illustrations from the Turing Way book dashes. Zenodo.https://doi.org/10.5281/ZENODO.4906004





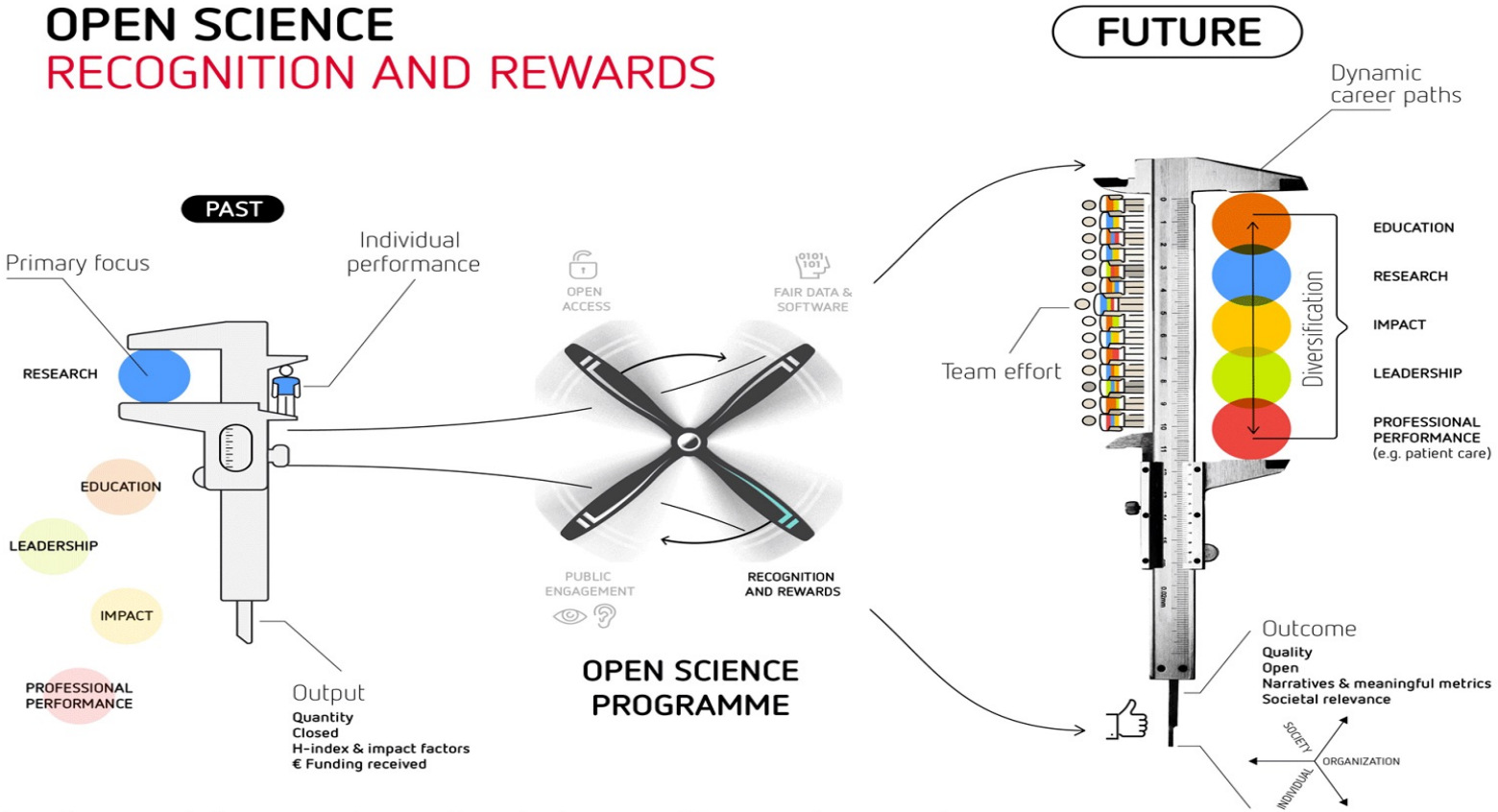


Source: Barbara Vreede and Jacques Flores. Drawn images by Scriberia

Incentivising and enabling reproducible research practices

Utrecht University

SKETCH AUTUMN 2020



https://www.uu.nl/en/research/open-science/tracks/recognition-and-rewards

The reproducibility crisis

- A reproducibility crisis? 90 % of the researchers think so
- Unable to reproduce: >70 % of another scientist's experiments, >50 % of their own experiments
- Insignificant interest to publish replicated positive studies (24 %) and journals rarely ever publish negative replication results (13 %)
- >60 % felt pressure to publish and report only selectively, which hindered reproducibility. Also competition for grants and growing bureaucracy played a part
- How to overcome the reproducibility crisis? ~ 90 % felt that a more robust experimental design and improved statistics and mentorship are needed. Also enhanced practices and journal checklists of high importance
- 80 % thinks that research funders and publishers should step up*
- One analysis estimates that 85% of biomedical research efforts are wasted**

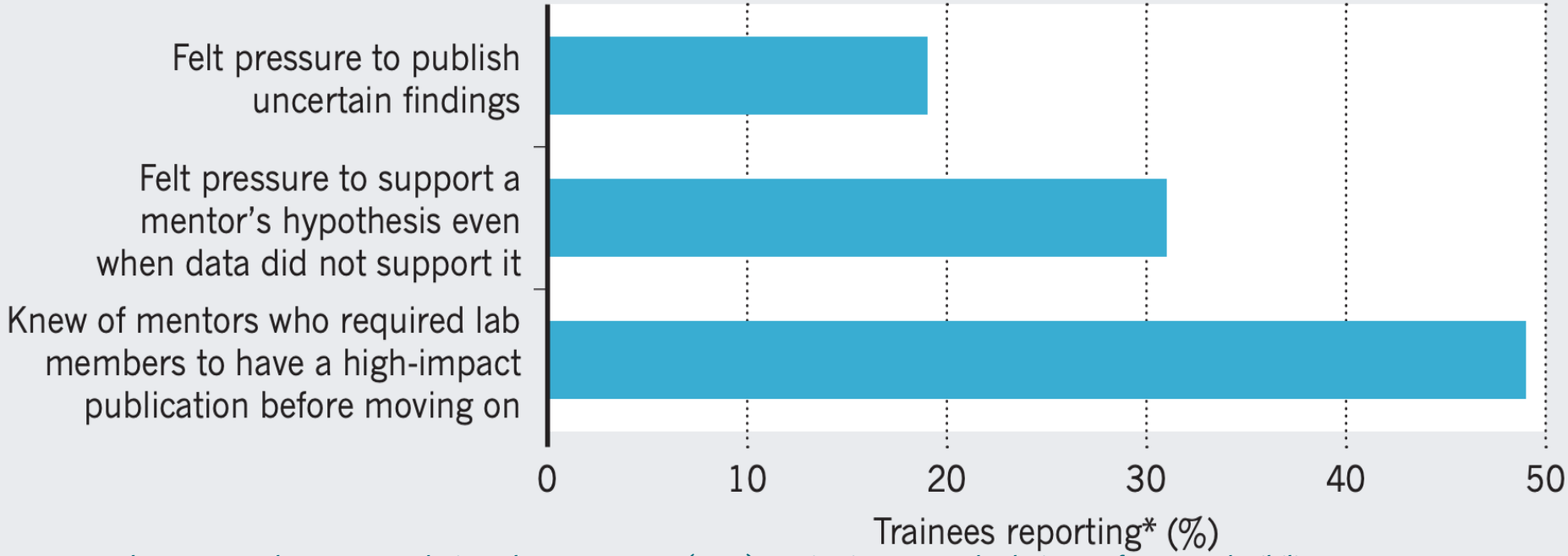
*Baker, M. *Nature* **533** (2016), 452-454.

** Macleod, M. R. et al. Biomedical research: increasing value, reducing waste. Lancet 383, 101–104 (2014)

Pressure resulting in deoriented results and unreliability

PRESSURED FINDINGS

A survey of US biomedical trainees suggests that the push to publish spurs unreliable results.

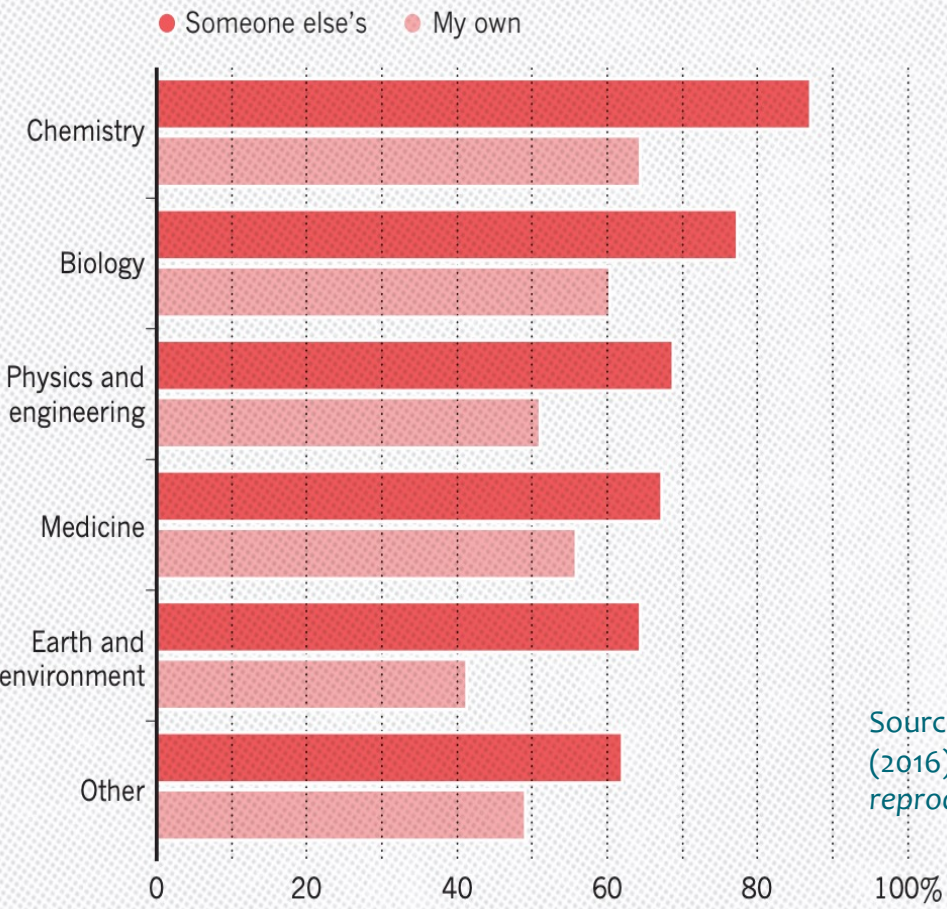


Source: Begley C. G., Buchan A. M. and Dirnagl U. Nature **525** (2015), Institutions must do their part for reproducibility

*Online survey of ~140 trainees at the MD Anderson Cancer Center in Houston, Texas.

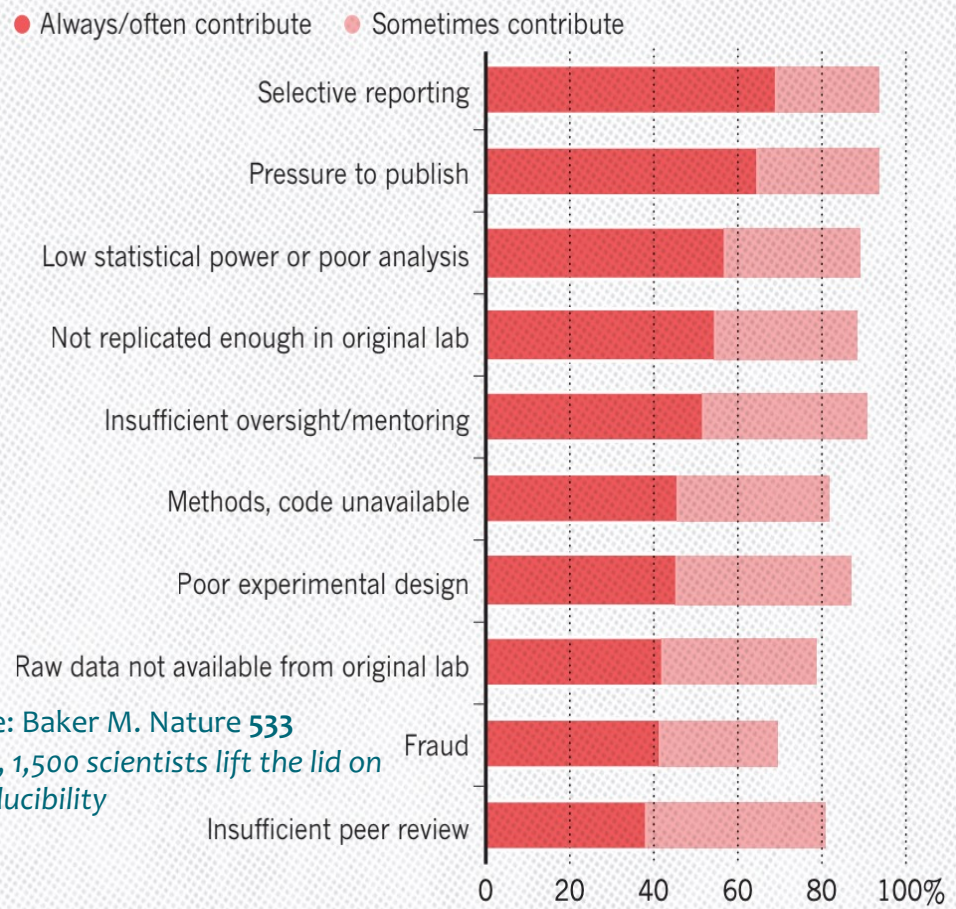
HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.



Source: Baker M. Nature **533** (2016), *1,500 scientists lift the lid on reproducibility*

The art of publishing reproducible research outputs

Knowledge Exchange

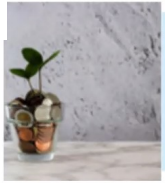
KE report on reproducible research outputs

- Approach of qualitative study:
  - Literature review
  - Interviews with selected stakeholders
- KE commissioned a study to explore the barriers and drivers of publishing reproducible research outputs from Research Consulting with the help of the KE Task & Finish group
- Final report published in late 2021

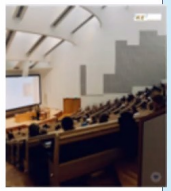
https://doi.org/10.5281/zenodo.5521076

Reproducibility and a call for changed research practices

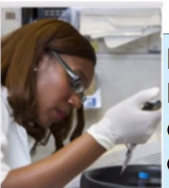
- Foster and apply reproducible workflows, including gathering data and code and curation
- Share appropriate research objects (digital and physical) alongside publications



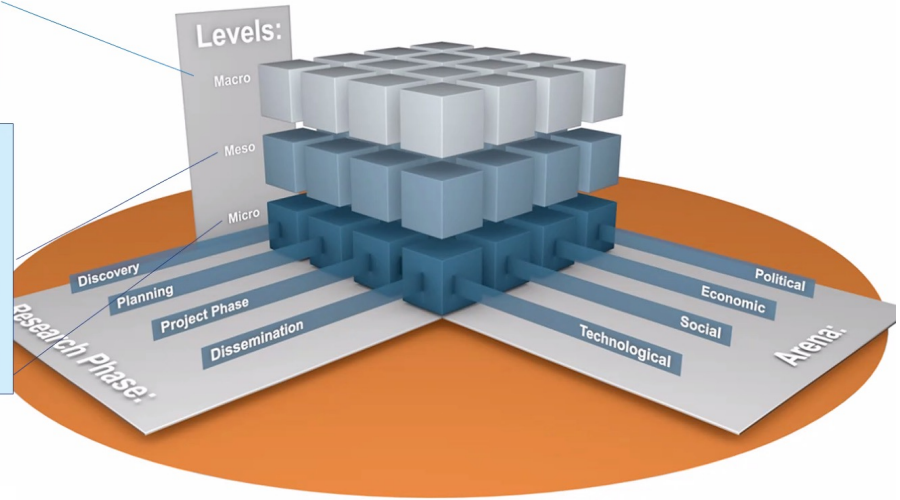
Macro level:
Research funding organisations see reproducibility as part of a broader discussion



Meso level:
- Disciplines should communicate their requirements, and publishers should implement
- Research performing organisations do not tend to mandate reproducible publication practices



Micro level:
Researchers and research groups have direct control over everyday practices



Knowledge Exchange

Framing the research reproducibility discourse

- Key **benefits** of reproducible research include:
  - **Increased confidence** in findings and results
  - An ability to continue one's (or someone else's) work in the future
  - Higher transparency, openness and **trust** in science

- Key **challenges** of reproducible research include:
  - **Incentive** structures
  - **Differences** in the technical capabilities of researchers
  - **Limited connectivity** between technical solutions
  - **Inconsistent** reporting standards

"Incentives in research hang on the published article. And so, the behaviour, the routine, over the last hundred years has been just to focus on getting that published article and then to move on. As long as that's allowed, I think our attempts to encourage reproducibility or to really enable and realise reproducibility across a huge segment of the literature is going to be limited."
Publisher

"I think if people are taught how to set up workflows that are reproducible, it also benefits them in the first place. If I have to touch the same project three years in the future, I might have forgotten what specifically I did, and if there is good documentation, everything is there and I save a lot of time."
Researcher

Knowledge Exchange

Tools & services

General purpose (e.g. open scholarship, data sharing)

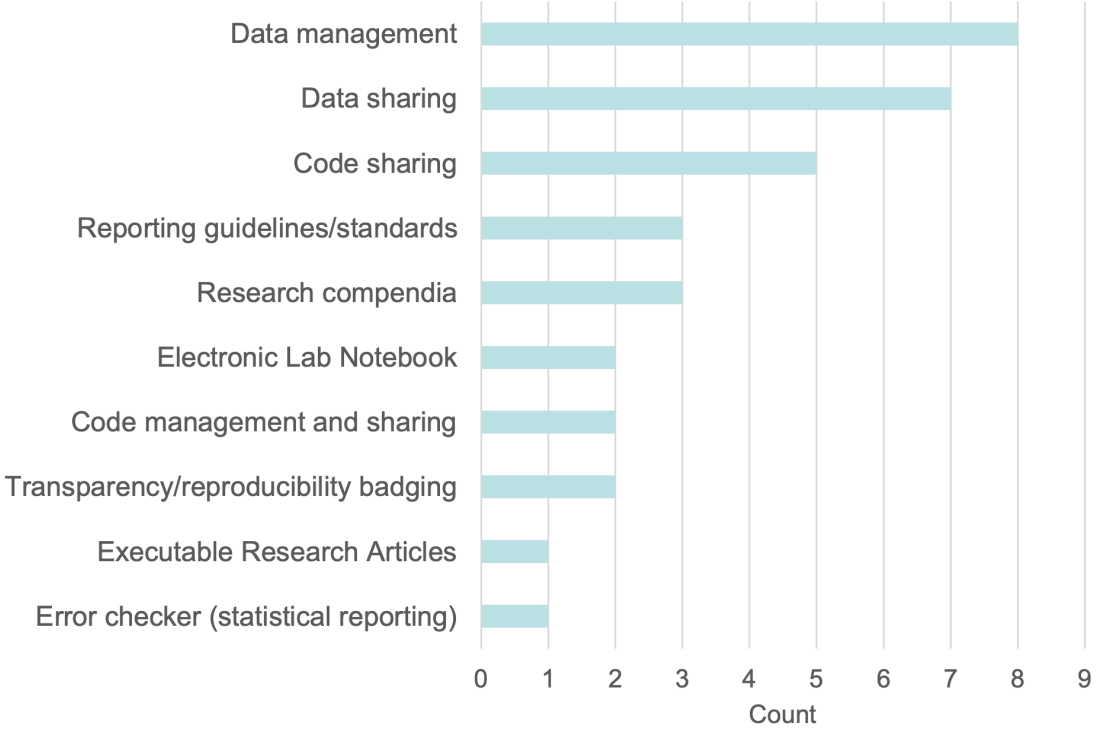
zenodo · EUROPEAN OPEN SCIENCE CLOUD · jupyter · GitLab

Specific purpose

ReproZip The Reproducibility Packer! · WHOLE TALE · cascad · eLife · RC

Tool or service	URL	Focus
cascad	https://bit.ly/3B0gtKc	Reproducibility checking
CODECHECK[89]	https://bit.ly/3nwsAtR	Independent execution of computations underlying research articles
Docker	https://dockr.ly/3aW723R	Software containerisation
eLife ERA	https://bit.ly/3oGCWcz	Executable research articles
ISA Framework	https://bit.ly/3pkCqRW	Whole research process (Life sciences)
Jupyter Notebooks	https://bit.ly/2Z8Ay4b	Code sharing and documentation
Binder	https://bit.ly/2Yw4SoX	Interactive and reproducible environments
Code Ocean	https://bit.ly/3vAU3hA	Computational research platform
Gigantum	https://bit.ly/3GiRyFF	Computational research platform
LabArchives	https://bit.ly/3G7jlse	Electronic Lab Notebook
LabFolder	https://bit.ly/3jolLsW	Electronic Lab Notebook
Octopus (in development)	https://bit.ly/3C6nESu	Whole research process (Sharing and documentation of research objects as research is developed)
Observable[90]	https://bit.ly/3ixJypO	Computational research platform
ReproZip	https://bit.ly/3piFrSO	Reproducibility packaging/bundling solution (research compendia)
Whole Tale	https://bit.ly/3m2Ftfs	Reproducibility packaging/bundling solution (research compendia)
R Markdown and R Notebooks	https://bit.ly/2Z7J6rF https://bit.ly/3BdU8Jj	Code sharing and documentation
R-Squared	https://bit.ly/2YqDojW	Reproducibility checking
Renku (See also case study on page 41)	https://bit.ly/3DWuf24	Whole research process (Collaborative Data Science)
Stencila	https://bit.ly/3m3lnS8	Executable research articles and Executable document pipelines

Infrastructure & tools – Technical and social dimensions



34 entries showing the most relevant types of infrastructures enabling reproducible research

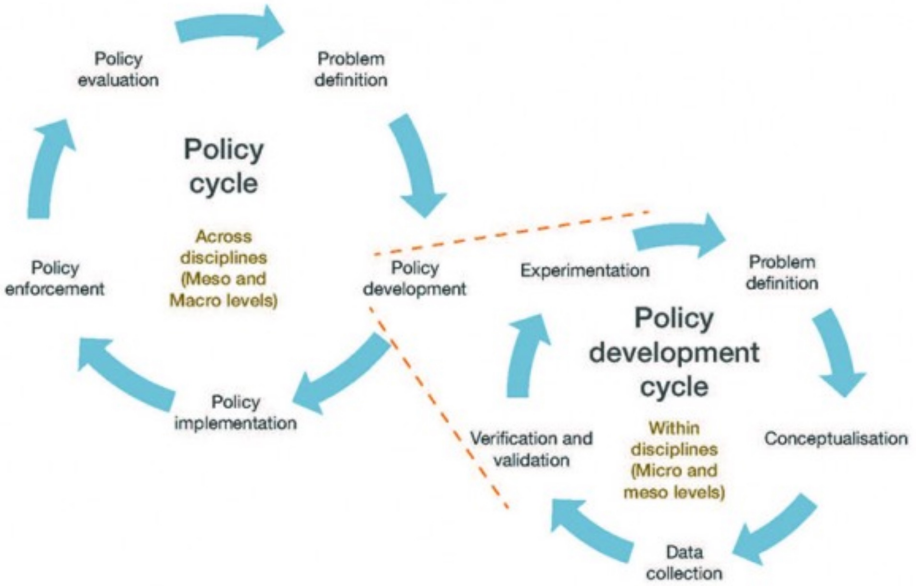
Key take-away messages – *Part 1*

- Blog post - 'Five things you need to know to support reproducible publication practices':

https://knowledge-exchange.info/news/articles/04-11-2021

1. **Reproducibility is part of the long-term vision for open science**
   - Coherent concepts and terminologies needed

2. **Disciplinary requirements for reproducible publications need to emerge from the bottom up**
   - Shaping policies at community-level on discipline specific needs to inform policy-making



Key take-away messages – *Part 2*

3. Reproducibility efforts should be "baked into" the research process, but incentives are needed

- Researchers are faced with many dis-incentives, such as publication numbers, impact factors and time restraints, which needs to change first

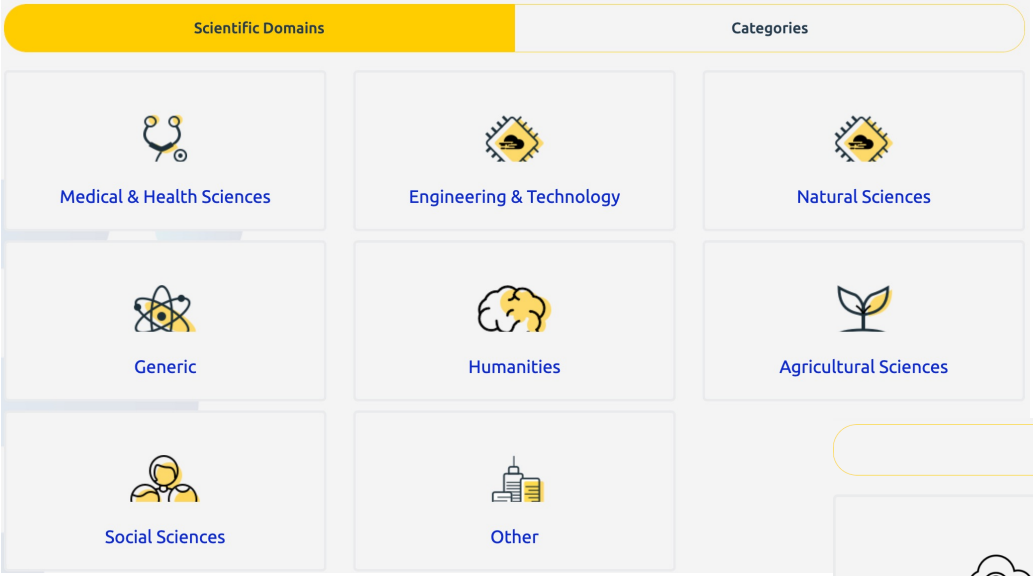
4. Good data management practices are a necessary condition for reproducible publication

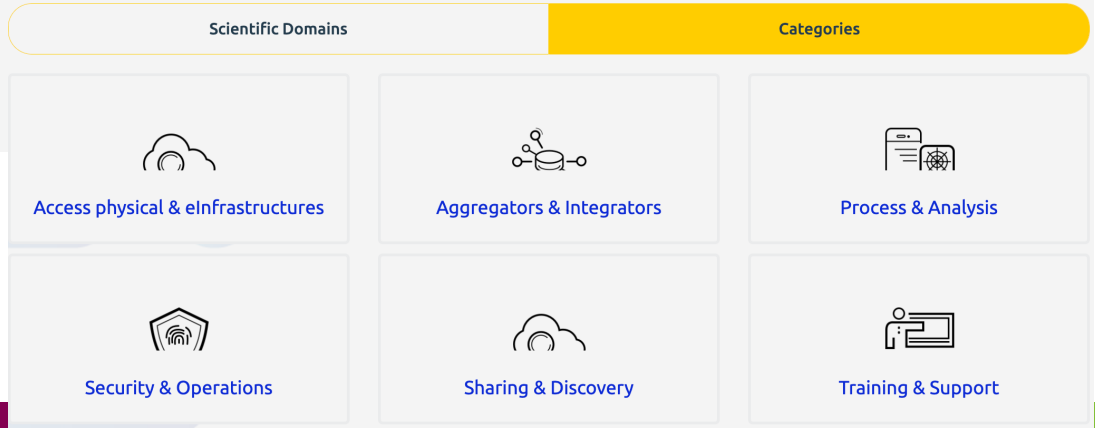
- It is key to early on in the research phase adopt good data management skills to make sure all research phases are being reproducible

5. Although digital tools and infrastructures are available, interoperability remains a gap

- Interoperability between systems still under-developed, which hinders reproducibility workflows

EOSC Service Portfolio





Services in EOSC Portal supporting reproducibility

- NOMAD Repository and NOMAD AI toolkit
- PRACE Code Vault
- Cloudferro Data related Services – EO Finder
- EcoPortal

Thank you!

Questions?


facebook.com/CSCfi


twitter.com/CSCfi


linkedin.com/company/csc---it-center-for-science


github.com/CSCfi