

# Reproducibility

Silvio Peroni

[silvio.peroni@unibo.it](mailto:silvio.peroni@unibo.it) – <https://orcid.org/0000-0003-0530-4305> – [@essepuntato@scholar.social](https://www.essepuntato.org/@essepuntato)

Open Science (A.Y. 2024/2025)

Second Cycle Degree in Digital Humanities and Digital Knowledge

Alma Mater Studiorum - Università di Bologna

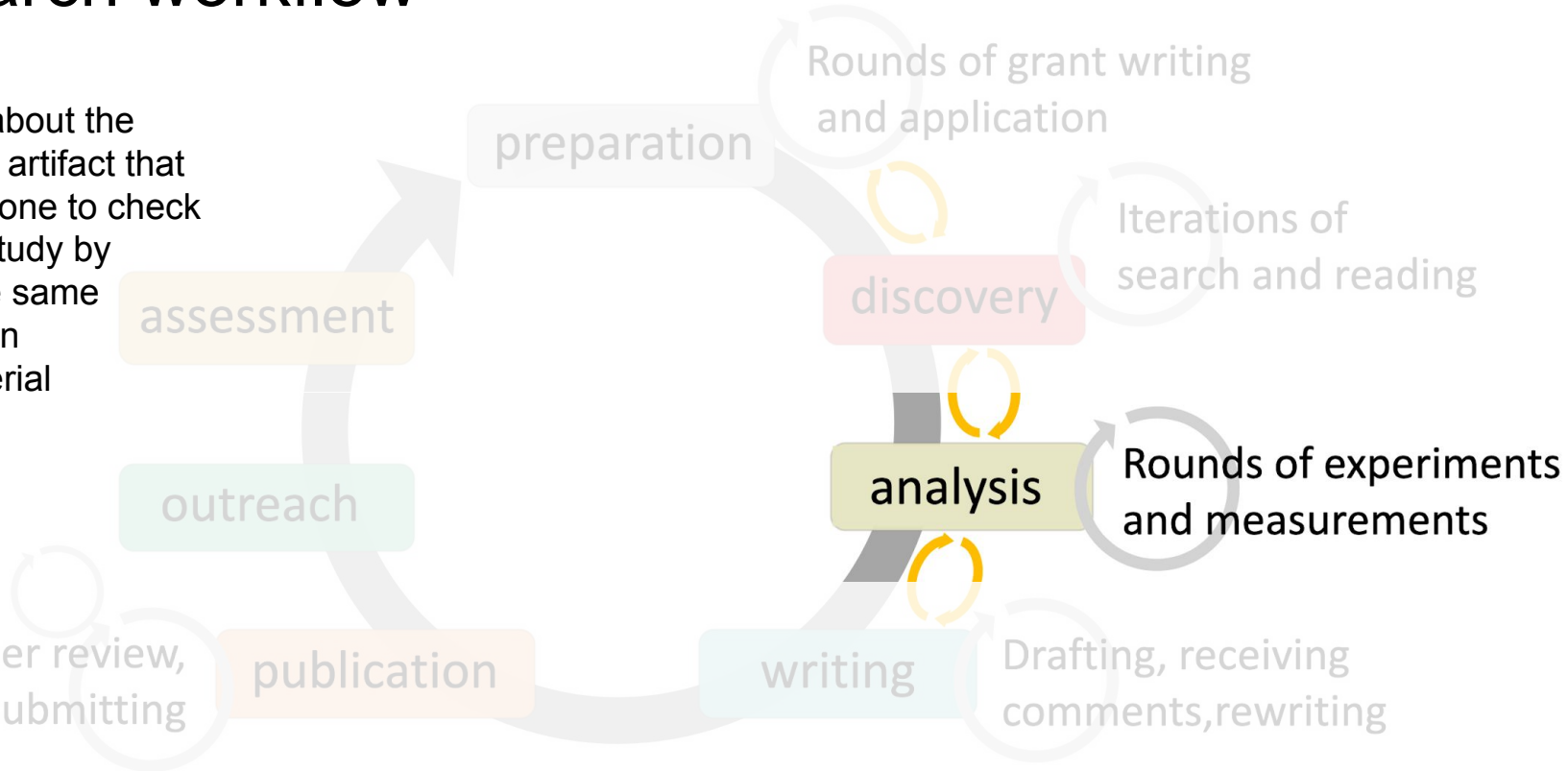


# Discussion

Meng, X.-L. (2020). Reproducibility, Replicability, and Reliability. Harvard Data Science Review, 2(4). <https://doi.org/10.1162/99608f92.dbfce7f9>

# Research workflow

We discuss about the availability of artifact that enable someone to check results of a study by run again the same experiment on existing material



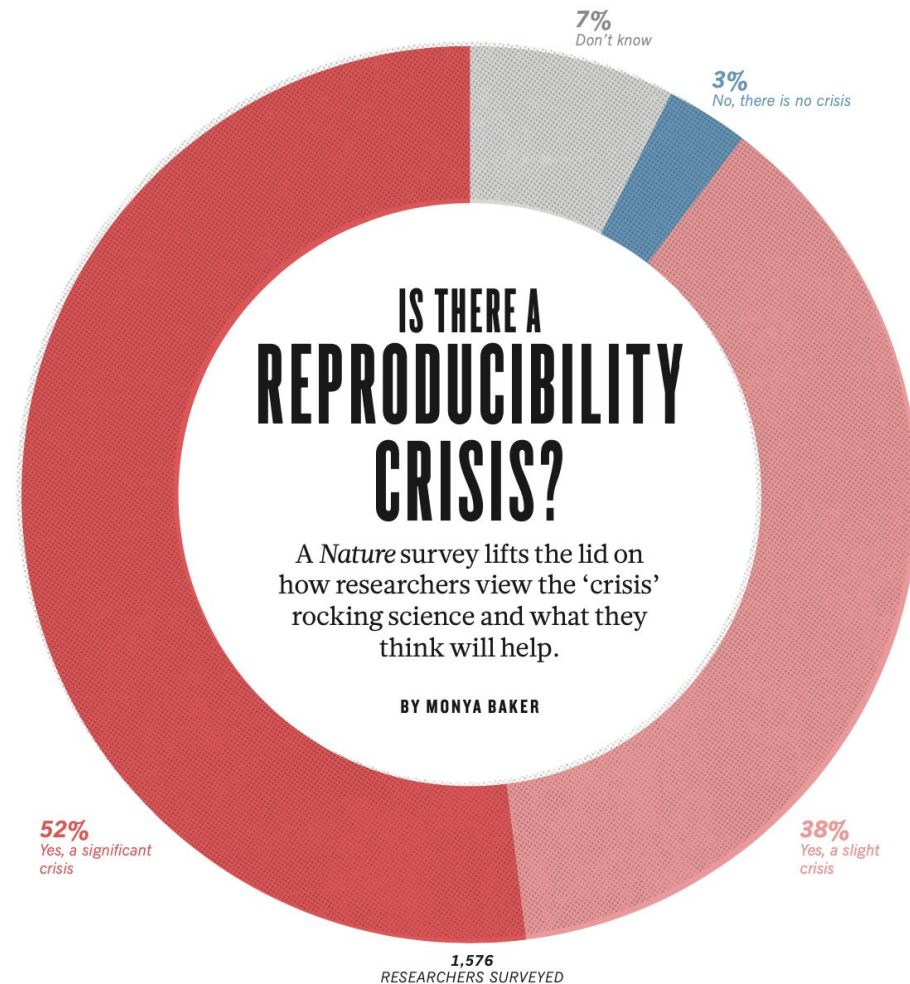
# Reproducibility crisis

>70% have tried and failed to reproduce another scientist's experiments

>50% have failed to reproduce their own experiments

<31% think that failure to reproduce means that the result is probably wrong (they trust the literature)

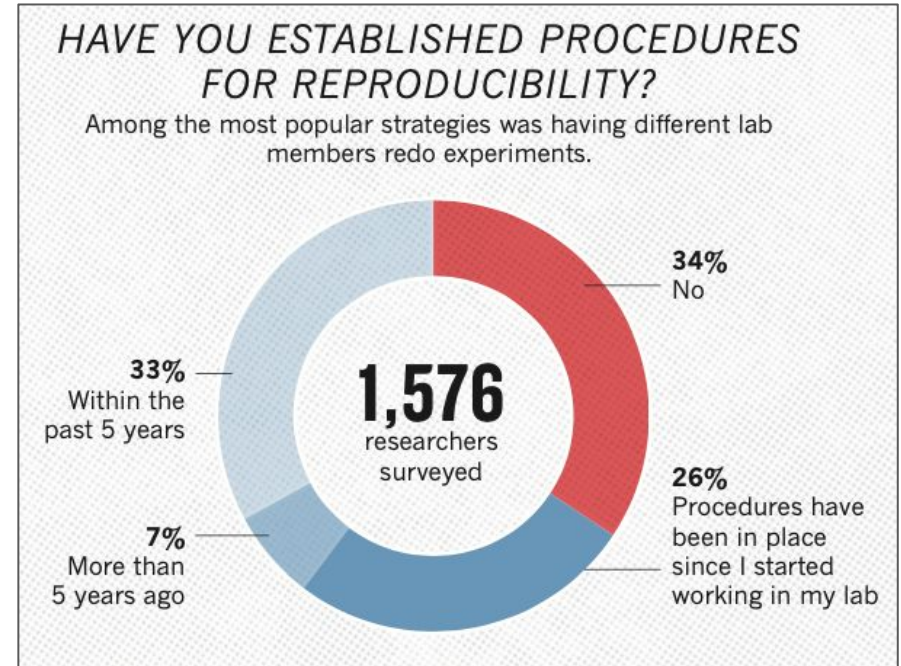
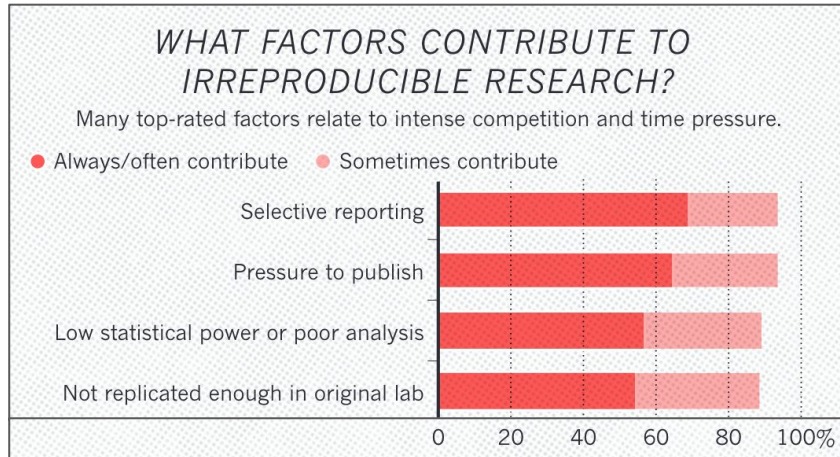
<20% contacted by another researcher unable to reproduce their work



# Some additional issues

Work does not reproduce →  
there is a perfectly valid reason

Incentives to publish positive replications are low +  
journals reluctant to publish negative findings



# Reproducibility and Open Science

Considering that more open, transparent, collaborative and inclusive scientific practices is a more efficient enterprise that improves the quality, **reproducibility** and impact of science

Open science sets a new paradigm that integrates into the scientific enterprise practices for **reproducibility**, transparency, sharing and collaboration resulting from the increased opening of scientific contents, tools and processes

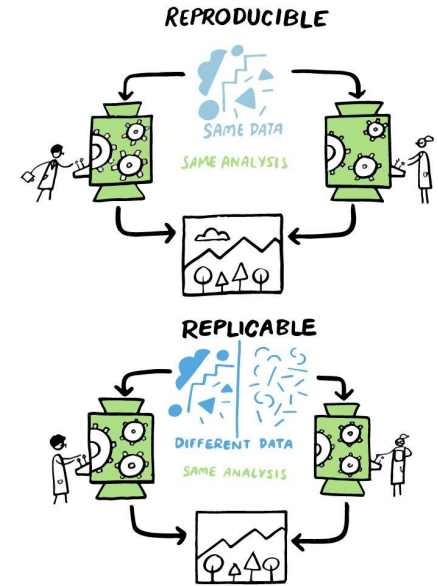
The following guiding principles for open science provide a framework for enabling conditions and practices within which the above values are upheld: transparency, scrutiny, critique and **reproducibility**

Promoting open science from the outset of the research process and extending the principles of openness in all stages of the scientific process to improve quality and **reproducibility**

# A definition from the European Commission

We consider reproducibility as a continuum based on three main research processes

- **Reproduction:** the **re-enactment of the results** of a study by a third party, using the **original set-up, data and methodology** of analysis
- **Replication:** for more general **re-enactment of the results**, using the **same analytical method**, but on different datasets
- **Re-use:** for the more loose possibility to **re-use the results beyond the original research context**, both inside and outside the original scientific discipline



# Importance of reproducibility

Enabling reproducibility allows **checking the soundness of analysis** and it is one of the key factors (but not the only one) to **build trust**

Example:

- Potti et al.'s article “Genomic signatures to guide the use of chemotherapeutics” published in Nature Medicine in 2006 (<https://doi.org/10.1038/nm1491>) – the authors claimed to have built an algorithm using genomic microarray data that predicted which cancer patients would respond to chemotherapy
- Keith Baggerly and Kevin Coombes (<https://doi.org/10.1214/09-AOAS291>) obtained the data and attempted to apply such an algorithm, and found that the data analysis conducted in the original study contained several errors that invalidated the outcomes of the study

# Problems derived from missing reproducibility

In principle, enabling reproducibility increase the chance to identify questionable research practices and misconducts such as fabricated, false, biased, and irreproducible findings

About data falsification: a survey reported that, on average, around 2% of scientists admitted to having fabricated or falsified data at least once

Ideal: the reproducibility of research methods should be expected to be 100%

Reality: the reproducibility of results and inferences is lower and to vary across subfields and methodologies

# What does it mean “to reproduce” in science

## Three types of “reproducibility”

1. **methods reproducibility**, i.e. the ability to exactly reproduce a study by using the same raw data and the same methodologies to obtain the same results
2. **results reproducibility**, i.e. the ability to obtain the same results from an independent study using the same methodologies as the original study
3. **inferential reproducibility**, i.e. coming to qualitatively similar conclusions from either an independent replication of a study or a reanalysis of the original study – it is the most common when talking about research reproducibility in the humanities – and implies either to **draw the same conclusions from different studies** and data or to come up with **different conclusions from the same original data**, sometimes even if they agree on the analytical results

# Mandatory text to read for the next lecture

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, 19, 43.

<https://doi.org/10.5334/dsj-2020-043>

# End

## Reproducibility

Silvio Peroni

[silvio.peroni@unibo.it](mailto:silvio.peroni@unibo.it) – <https://orcid.org/0000-0003-0530-4305> – [@essepuntato@scholar.social](https://twitter.com/essepuntato)

[Open Science \(A.Y. 2024/2025\)](#)

[Second Cycle Degree in Digital Humanities and Digital Knowledge](#)

[Alma Mater Studiorum - Università di Bologna](#)

