

# Lecture 25 – Open Science and Data Archiving

## Learning Objectives:

**6. Learn how to document your work and prepare scientific publications.**

**6.5 Define Open Science, Open Access, and Open Research.**

**6.6 Discuss the ethics of open science and open access to scientific data.**

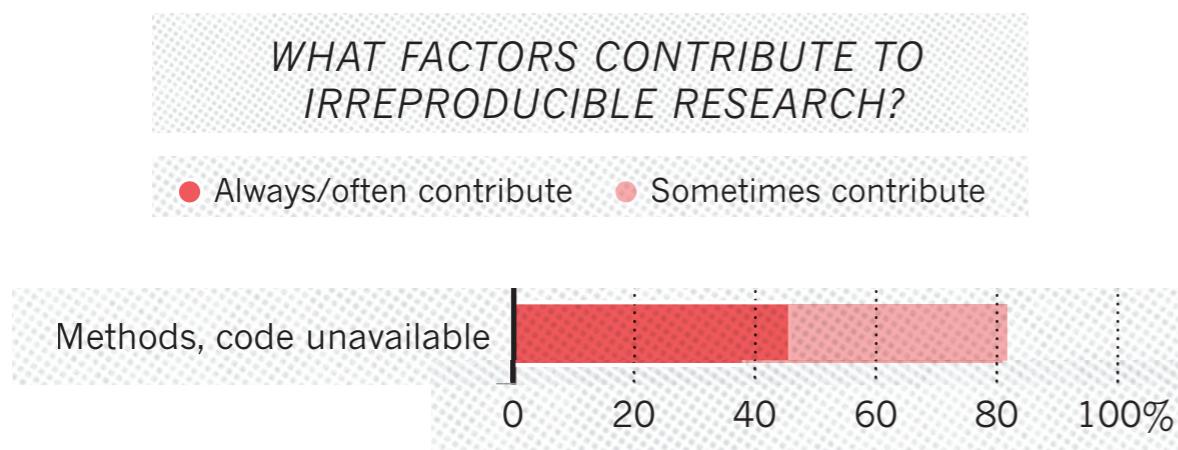
# Publication is not the end

- Publication is a fundamental unit of science. Peer-reviewed publications are used as a basic unit of evaluation, basis for promotion, and merit. Important for dissemination of research.

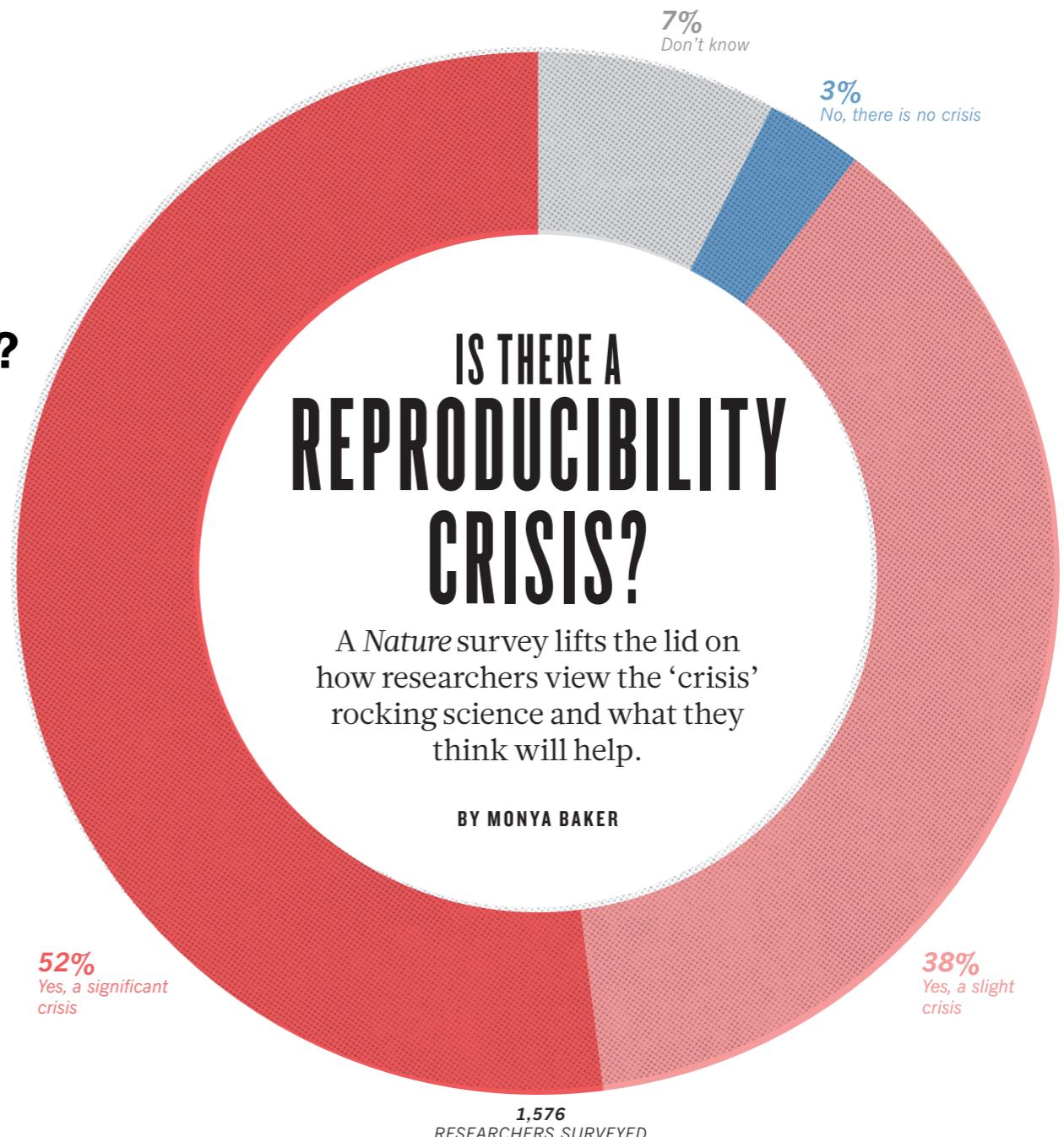
- After publication:

1. Papers are read and cited.
2. Other works are built upon results.
3. Results are replicated... sometimes?

- Reproducibility crisis in science



!!!!



Baker 2016

# Open Science and Computing

**Open Science:** movement to make all scientific research accessible to all levels of people (including laypeople and other researchers).

## Within Open Science:

- **Open research:** practice of making all scientific software free and open source.
- **Open access:** promotes research outputs disseminated online without barriers such as cost or location.
- **Open-notebook science:** practice of releasing entire primary record of scientific work publicly available online.

## For scientific computing, three components:

- **Open Access to papers**
- **Open Access to data**
- **Open source and accessible software**

# Open Science and Computing

## The five “R’s” of Benureau & Rougier 2018:

- 1. Re-run.** Code should be able to be re-run more than once with minimal explanation.
- 2. Repeat.** Code should produce the same result more than once.
- 3. Reproduce.** Code should allow researcher to reproduce results.
- 4. Reuse.** Code should be in a form that can be re-used by other researchers.
- 5. Replicate.** Code should make it possible for *others* to reproduce results of published work.

**Reproducibility and Replicability require planning!**

# Data and Code Archiving: Policies

Many journals require data to be freely accessed by anyone, not simply shared “by request.”



**DATA AND CODE  
AVAILABLE  
"BY REQUEST"**

Many funding organizations now require information about how data is to be used, shared, accessed, and formatted (e.g. NSF's Data Management Plan).



**DATA AND CODE  
POSTED IN  
REPOSITORY**

<https://fairsharing.org/> : FAIR sharing website catalogs standards and policies for reuse and access of data

<https://www.re3data.org/> : Registry of Research Data Repositories

# Data and Code Archiving: Repositories

- General Use. Issue permanent DOI's, citable.



**figshare**

[figshare.com](https://figshare.com)



[datadryad.org](https://datadryad.org)



**Open Science Framework**

<https://osf.io/>

- Pre-print servers

arXiv – <https://arxiv.org/>

bioRxiv – <https://www.biorxiv.org/>

TechRxiv – <https://www.techrxiv.org/>

- Social Networking

Researchgate –  
<https://www.researchgate.net/>

Linkedin –  
<https://www.linkedin.com/>

OSF Preprints – <https://osf.io/preprints/>

# **Open Science: is it that hard?**

- Now that you've had a chance to prepare a project for release, what are your thoughts on open science?**
- What are reasons that researchers may not want to release their code and/or data?**

# Additional Resources

**<https://www.nature.com/sdata/policies/repositories> – Recommended Data Repositories (Nature)**

**<https://fairsharing.org/> – FAIR sharing data standards and policies database**

**<https://www.re3data.org/> – Registry of Research Data Repositories**

**<https://asappbio.org/preprint-servers> – List of pre-print servers and policies**