

# FAIR\_bioinfo: open science & FAIR principles

---

How to make a bioinformatics project more reproducible

# The 2022 team



Hélène Chiapello\*, Céline Hernandez\*\*, Thomas Denecker\*, Julien Seiler\*\*,  
Gildas Le Corguillé\*, Claire Toffano-Nioche\*\*

Helpers: Hugo Arbes\*\*, Emilie Drouineau\*\*, Pauline François\*\*\*

\* Institut Français de Bioinformatique, \*\* Institut de Biologie Intégrative,

\*\*\* Centre International de Recherche en Infectiologie

# Promote learning



## Objective:

FAIR raw data  
+  
FAIR scripts  
=  
FAIR processed data

## Courses:

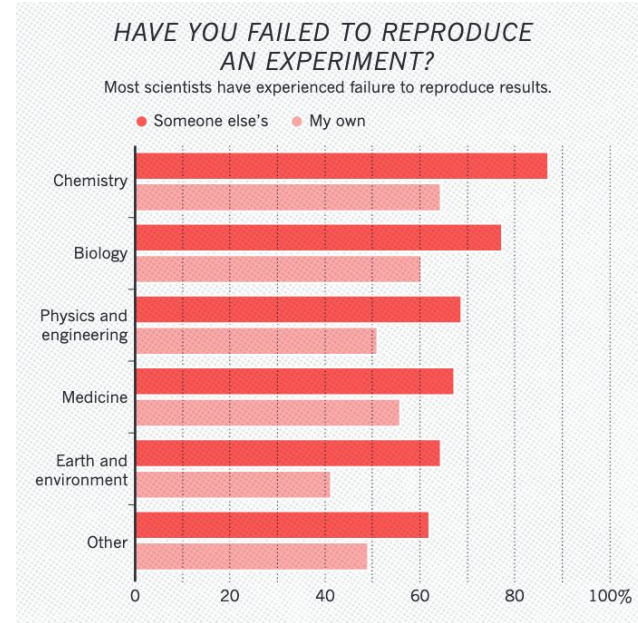
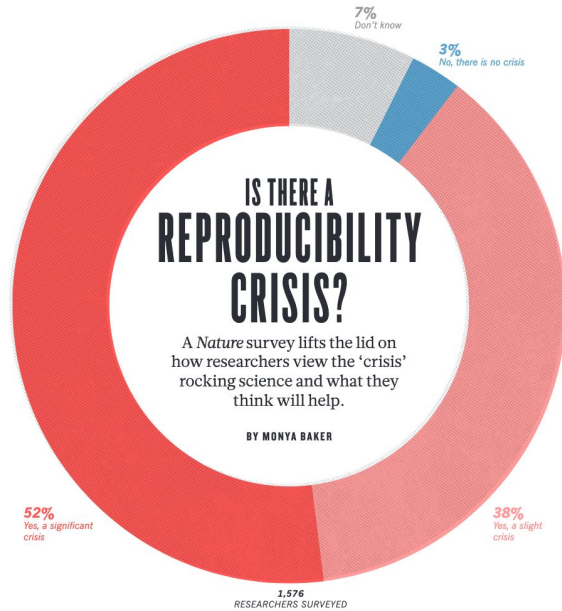
Take your first steps with several companion tools to gain in reproducibility

## Example based:

Only the beginning of an NGS analysis (the full analysis is given as a bonus)

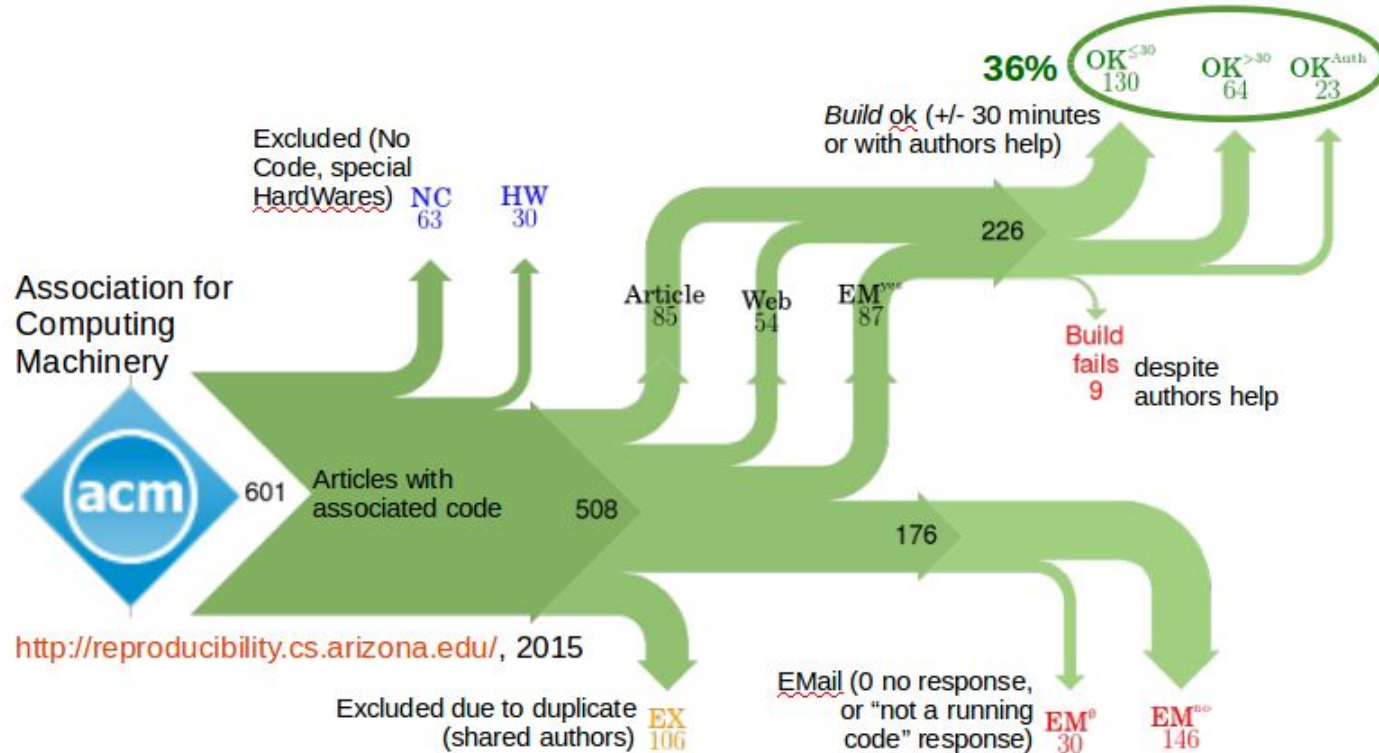
# Reproducibility crisis

2016



Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452–454 (2016). <https://doi.org/10.1038/533452a>

# Also in computer sciences

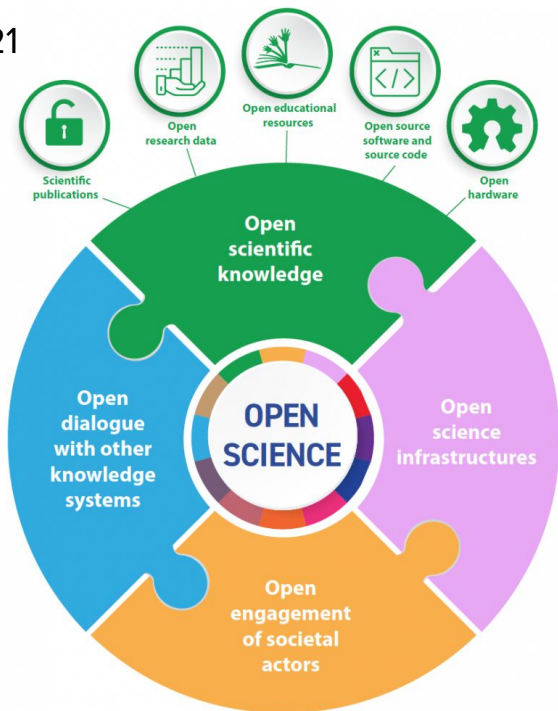


# Long term negative impact of retracted papers

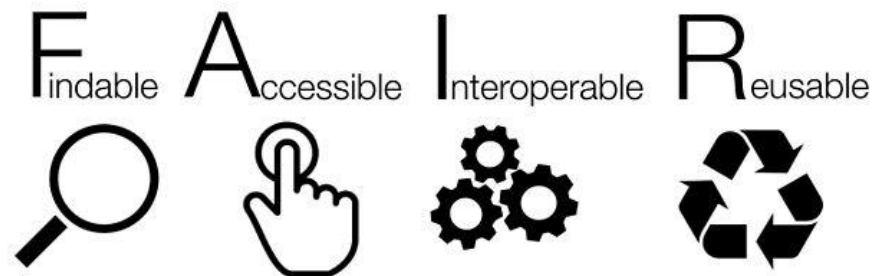
Article	Year of retraction	Citing Articles before retraction	Citing Articles after retraction	Total cites (journals indexed by Web of Science)
1. Primary Prevention of Cardiovascular Disease with a Mediterranean Diet. N ENGL J MED; APR <b>2013</b> . Estruch R, et al.	2018	1919	816	2735
2. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. LANCET; FEB 28 <b>1998</b> . Wakefield AJ, et al.	2010	642	867	1509
3. Visfatin: A protein secreted by visceral fat that mimics the effects of insulin. SCIENCE; JAN <b>2005</b> . Fukuhara A, et al.	2007	232	1192	1424
4. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. PLANT J; MAR <b>2003</b> . Voinnet O, et al.	2015	896	375	1271
5. Lysyl oxidase is essential for hypoxia-induced metastasis. NATURE; APR <b>2006</b> . Erler JT, et al.	2020	977	81	1058

# A way out: Open science and FAIR principles

2021



2016



Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

# FAIR principles

F  
indable



By 維基小霸王 - Own work, CC BY-SA 4.0,  
<https://commons.wikimedia.org/w/index.php?curid=88894774>

PID  
Repository



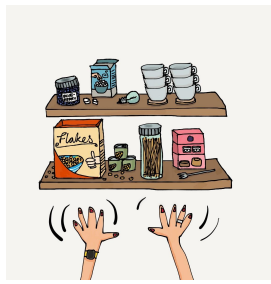
# FAIR principles

F  
indable

A  
ccessible



By 維基小霸王 - Own work, CC BY-SA 4.0,  
<https://commons.wikimedia.org/w/index.php?curid=88894774>



<https://nlsfirstworldproblems.tumblr.com/post/147555650875/i-cant-reach-the-top-shelves-of-the-kitchen>

PID  
Repository

Protocols  
(free, open, auth.)

# FAIR principles

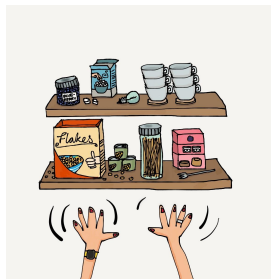
F  
indable



By 維基小霸王 - Own work, CC BY-SA 4.0,  
<https://commons.wikimedia.org/w/index.php?curid=88894774>

PID  
Repository

A  
ccessible



<https://nlsfirstworldproblems.tumblr.com/post/147555650875/i-cant-reach-the-top-shelves-of-the-kitchen>

Protocols  
(free, open, auth.)

I  
nteroperable



By Unknown author - Popular Science Monthly Volume 88,  
Public Domain,  
<https://commons.wikimedia.org/w/index.php?curid=22614407>

Standards  
(format, vocabulary)

# FAIR principles

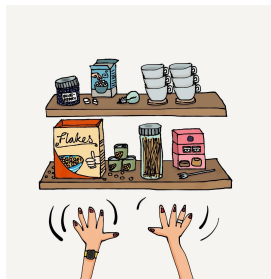
F  
Findable



By 維基小霸王 - Own work, CC BY-SA 4.0,  
<https://commons.wikimedia.org/w/index.php?curid=88894774>

PID  
Repository

A  
Accessible



<https://nlsfirstworldproblems.tumblr.com/post/147555650875/i-cant-reach-the-top-shelves-of-the-kitchen>

Protocols  
(free, open, auth.)

I  
Interoperable



By Unknown author - Popular Science Monthly Volume 88,  
Public Domain,  
<https://commons.wikimedia.org/w/index.php?curid=22614407>

Standards  
(format, vocabulary)

R  
Reusable



By Sun Ladder - Own work, CC BY-SA 3.0,  
<https://commons.wikimedia.org/w/index.php?curid=5746428>

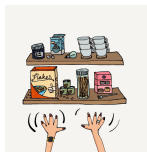
Metadata  
License  
Origin

# FAIR tools

**F**  
Findable



**A**  
Accessible



**I**  
Interoperable



**R**  
Reusable



Data



Software  
and  
analyses

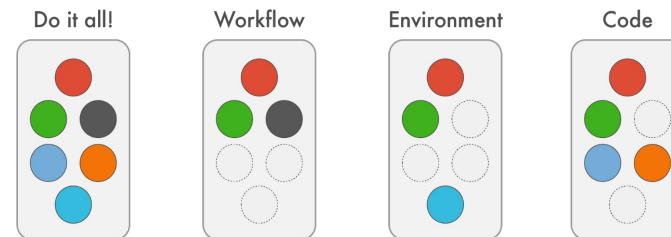


# Tools & use cases

Several tools but which ones to use and how? do some of them interact with each other?

3 use cases based on the previous sessions:

- E-labbook
- Reproducibility of running code
- Reproducibility in HPC



# Resources

- [awesome](#) a curated list of reproducible research case studies, projects, tutorials, and media
- The Role of [Metadata](#) in Reproducible Computational Research
- [Towards reproducible computational biology](#)
- A very similar sweden [courses](#) with git, conda, snakemake, jupyter, r-markdown, docker, singularity