

Introduction to open and reproducible research

Agata Bochynska, PhD

Open Research and Digital Scholarship Center
University of Oslo Library

@AgataBochynska

agata.bochynska@ub.uio.no

Roadmap

- Why open research?
- Open research plans, materials, data, analyses, publishing: why, how and where?
- Open research and reproducibility
- Take-aways
- Q&A time!

COMMENTARY

Open Access

Open science saves lives: lessons from the COVID-19 pandemic



Lonni Besaçon^{1,2*} , Nathan Peiffer-Smadja^{3,4}, Corentin Segalas⁵, Haiting Jiang⁶, Paola Masuzzo⁷, Cooper Smout⁷, Eric Billy⁸, Maxime Deforet⁹ and Clémence Leyrat^{5,10}

Abstract

In the last decade Open Science principles have been successfully advocated for and are being slowly adopted in different research communities. In response to the COVID-19 pandemic many publishers and researchers have sped up their adoption of Open Science practices, sometimes embracing them fully and sometimes partially or in a sub-optimal manner. In this article, we express concerns about the violation of some of the Open Science principles and its potential impact on the quality of research output. We provide evidence of the misuses of these principles at different stages of the scientific process. We call for a wider adoption of Open Science practices in the hope that this work will encourage a broader endorsement of Open Science principles and serve as a reminder that science should always be a rigorous process, reliable and transparent, especially in the context of a pandemic where research findings are being translated into practice even more rapidly. We provide all data and scripts at <https://osf.io/renxy/>.

Keywords: Open science, Peer review, Methodology, COVID-19

UNIVERSITY
OF OSLO

TECHNOLOGY FEATURE | 24 April 2020

Open science takes on the coronavirus pandemic

Data sharing, open-source designs for medical equipment, and hobbyists are all being harnessed to combat COVID-19.

[Mark Zastrow](#)



A student in Warsaw assembles 3D-printed protective masks. Credit: Jaap Arriens/NurPhoto/Getty



Kaitlyn M. Werner, PhD

@kaitlynmwerner



Open science truly is beautiful. Someone recommended a paper w. open data relevant to this question. Within minutes I was able to analyze my question because the data/code was so beautifully and efficiently organized -- the best I've seen! Major props to [@russpoldrack](#) and team.



Kaitlyn M. Werner, PhD @kaitlynmwerner · May 1

I have been thinking a lot about socioeconomic status and self-control/self-regulation. I'm starting to plan an esm+diary study where I can start digging into this topic in more detail, but in the meantime I'm curious: what are the interesting papers you've read in this space?

[Show this thread](#)

11:52 PM · May 9, 2022 · Twitter Web App

Open science means transparency and **knowledge-sharing** in research processes to make knowledge **accessible** across academic groups, sectors and national boundaries. The concept of open science encompasses the entire research process [...].

- The Research Council of Norway. Policy for open science 2020

The Research Council Policy for Open Science

In effect from 2020





“**Open Science** is becoming the modus operandi for carrying out research and innovation by **sharing knowledge, data and tools** as early as possible, in open collaboration with all relevant knowledge actors and society.”

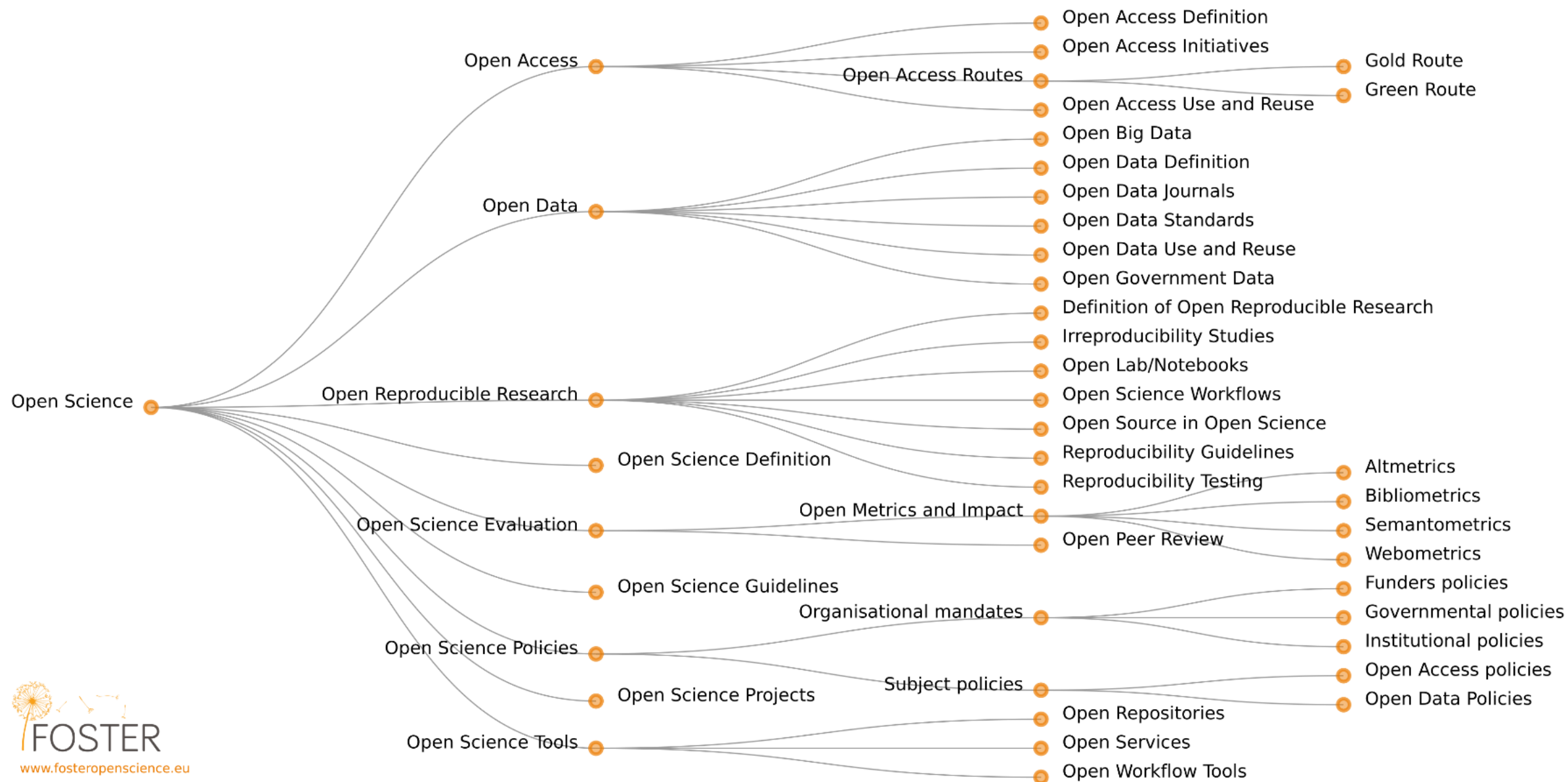
“**Open Science** has the potential of making the scientific process more **transparent, inclusive** and **democratic**. It is (...) a true game changer in bridging the science, technology and innovation gaps and fulfilling the **human right to science**.”



UNESCO Recommendation on Open Science

https://youtu.be/I3Wkvx_ZaFo
<https://www.unesco.org/en/natural-sciences/open-science>

Open Science Taxonomy



Open research plans



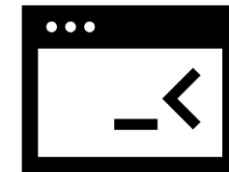
Open materials



Open data



Open analyses



Open publishing



Open research plans



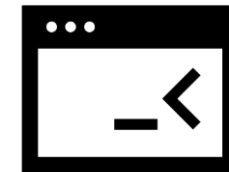
Open materials



Open data



Open analyses



Open publishing



Open research plans: why?



- More **visibility** to research ideas and plans early in the process as well as null results
- Clearer **distinction** between what was planned and what was not planned ahead (e.g. your *post hoc* analyses)
- Reduced researcher **biases** (less data/statistics manipulation)

Research Report



Power Posing: Brief Nonverbal Displays Affect Neuroendocrine Levels and Risk Tolerance

Dana R. Carney¹, Amy J.C. Cuddy², and Andy J. Yap¹

¹Columbia University and ²Harvard University

Psychological Science

21(10) 1363–1368

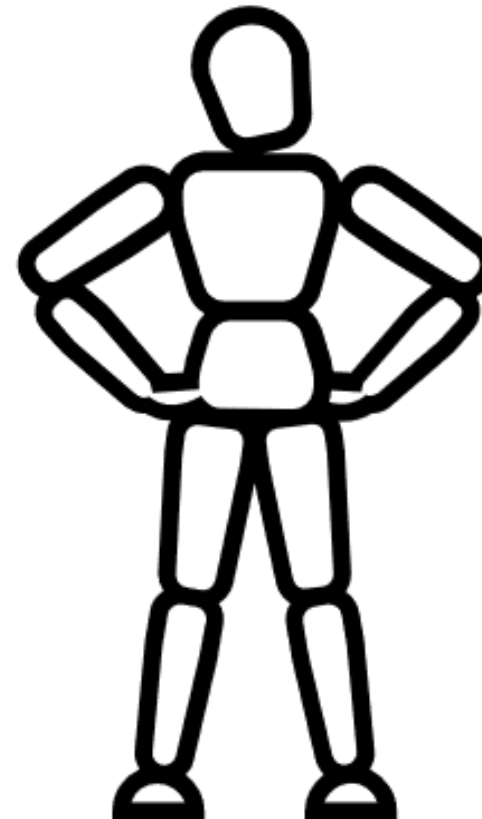
© The Author(s) 2010

Reprints and permission:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0956797610383437

<http://pss.sagepub.com>



SEPT. 26, 2016

'Power Posing' Co-author: 'I Do Not Believe That 'Power Pose' Effects Are Real'

By Jesse Singal



Amy Cuddy. Photo: Craig Barritt/Getty Images for WME Live

"We ran subjects in [chunks] and **checked the effect along the way**. It was something like 25 subjects run, then 10, then 7, then 5. Back then this did not seem like **p-hacking**. It seemed like saving money (...)"

<https://www.thecut.com/2016/09/power-poses-co-author-i-dont-think-power-poses-are-real.html>

Open research plans: why?



- More **visibility** to research ideas and plans early in the process as well as null results
- Clearer **distinction** between what was planned and what was not planned ahead (e.g. your *post hoc* analyses)
- Reduced researcher **biases** (less data/statistics manipulation)
- You will **think more deeply** about your research design and planned analyses
- You can claim credit for your **research ideas**!

Open research plans: how?



Preregister!

Specify the **research design**, **hypotheses**, and/or **analysis plan** prior to observing the outcomes of a study.

Typically in a form of a time-stamped, frozen document made available on an online platform.

Open research plans: where?



Focus	Type of research	Platform	Template
Discipline-specific	Clinical research	clinicaltrials.gov	Generic
	Animal research	animalstudyregistry.org	Generic
	Economics/Social sciences	socialscienceregistry.org	Generic
Discipline-general	Basic research	aspredicted.org	Generic
		osf.io/registries	Structured, Unstructured, Qualitative research, Replications, etc.

Open research plans



Preregistration of research studies

Time and place: May 13, 2022 9:00 AM–10:30 AM, [Add to calendar](#) [Zoom](#)

Learn about what preregistration is and how to preregister your own studies.

Preregistration on Open Science Framework (OSF)

Time and place: May 19, 2022 12:00 PM–1:30 PM, [Add to calendar](#) [Zoom](#)

Learn about how to preregister your study on Open Science Framework (OSF) and how to navigate the platform.

Need help with preregistration?

Get in touch with:

agata.bochynska@ub.uio.no

Open research plans



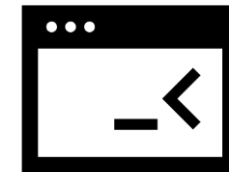
Open materials



Open data



Open analyses



Open publishing



Open research plans



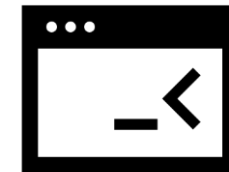
Open materials



Open data



Open analyses



Open publishing



Open materials: why?



- Enable **re-use** of different materials used in research (and reduce resources for development of tools, instruments, stimuli, etc)
- Make it easier for others to see **how (exactly) your study looked**
- Allow for **validating** your research
- Allow for **replicating** your research by other investigators or research groups

Open materials: how?



- Make sure that the materials are not **copyrighted** (when you make them fully open)
- Provide all materials with clear, consistent **names**
- Create comprehensive **documentation**
- Provide detailed **instructions** on how to use the materials

Open materials: where?



- Upload on your personal/lab/institutional website
- Upload in an online repository



Open research plans



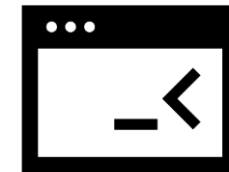
Open materials



Open data



Open analyses



Open publishing



Open research plans



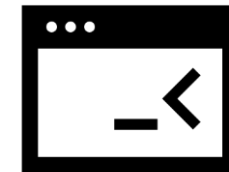
Open materials



Open data



Open analyses



Open publishing



Open data: why?



External Factors

- Funder requirements
- Publisher requirements

Career Benefits

- Increased visibility
- More data reuse
- New collaborations
- Increased citations

Scientific Progress

- More robust research
- Enables verification of results
- Enables new collaborations across disciplines and borders
- Opens up for new uses of data
- Avoids duplication
- Easier to use data in teaching

Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end

Lisa L. Shu^a, Nina Mazar^{b,1}, Francesca Gino^c, Dan Ariely^d, and Max H. Bazerman^c

^aKellogg School of Management, Northwestern University, Evanston, IL 60208; ^bRotman School of Management, University of Toronto, Toronto, ON, Canada M5S 3E6; ^cHarvard Business School, Harvard University, Boston, MA 02163; and ^dFuqua School of Business, Duke University, Durham, NC 27708

Edited* by Daniel Kahneman, Princeton University, Princeton, NJ, and approved July 23, 2012 (received for review June 11, 2012)



SCIENCE

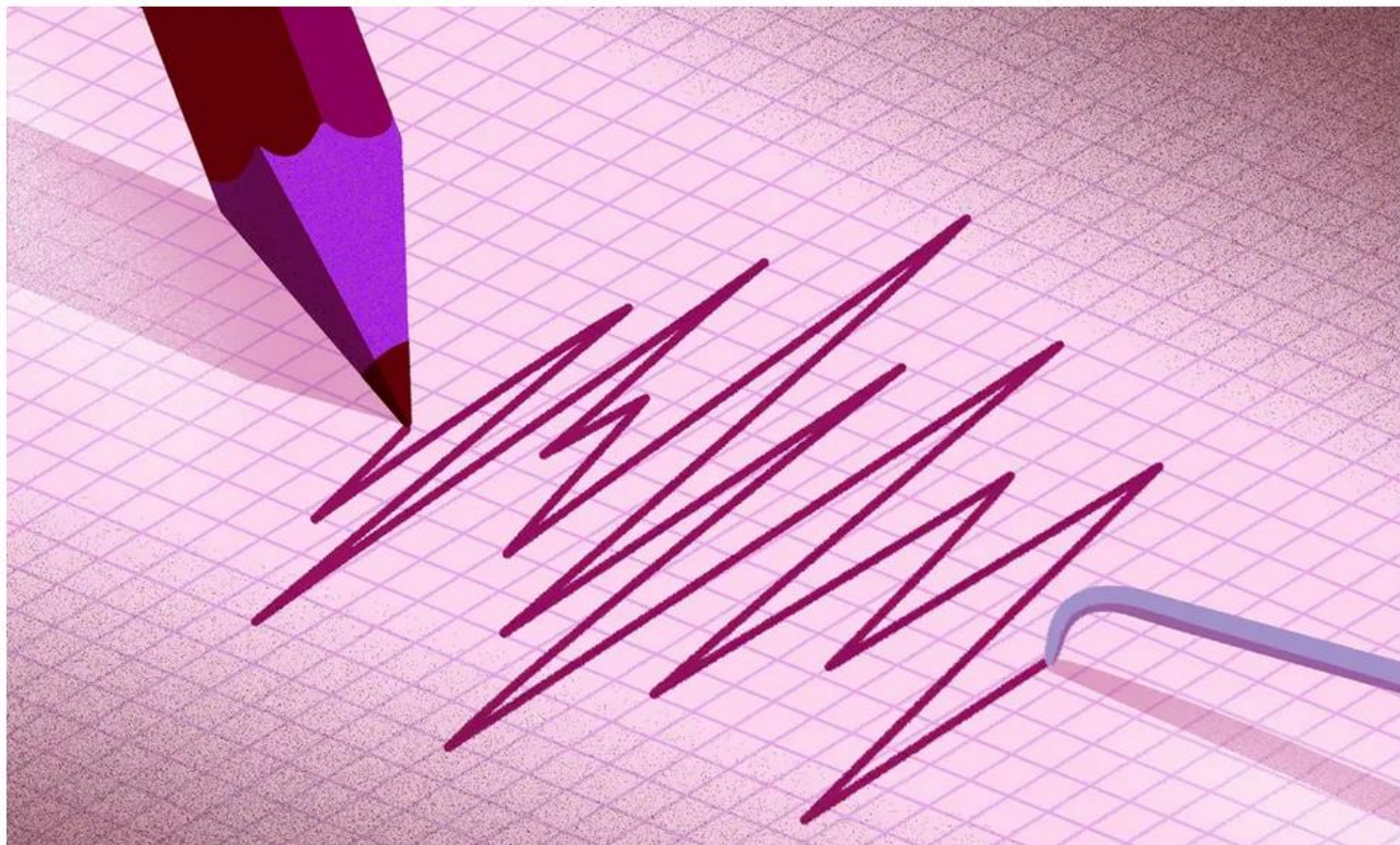
A Famous Honesty Researcher Is Retracting A Study Over Fake Data

Renowned psychologist Dan Ariely literally wrote the book on dishonesty. Now some are questioning whether the scientist himself is being dishonest.



Stephanie M. Lee
BuzzFeed News Reporter

Posted on August 20, 2021, at 2:40 p.m. ET



"When the researchers published their 2020 update, they posted the data from their 2012 paper for the first time. **Publicly sharing data was once a rarity** in science but is slowly becoming more commonplace amid calls for greater transparency."

Research Data Policy at UiO



1. Research data shall be made **openly available for further usage**
2. Research data should be **made available at an early stage**
3. Research data shall be **provided with a data management plan**
4. Research data shall **have metadata and be documented**
5. Research data must be **securely archived**
6. Research data shall be provided with **licenses for access, reuse and redistribution**
7. Research data should be **made freely available, but the actual distribution cost should be covered**

Open data: how?



MadScientist
@MadS100tist



"Data will be available upon request"



Open data: how?



- Data availability statements (“data available upon request”) do not work - deposit data in a **data repository** (data archive)
- Apply consistent, meaningful, and compatible **file naming**
- Choose accessible, patent-free, and open **file formats**
- Make sure you have the necessary **documentation** (and metadata)
- **Reduce complexity** by grouping large groups of similar files in zip bundles (and check any size limitations of the repository)

Open data: where?



- General purpose repositories
- Discipline-specific repositories
- Data papers
- Supplementary materials to an article

General-purpose data repositories

The Zenodo logo consists of the word "zenodo" in a white, lowercase, sans-serif font, centered within a solid blue rectangular background.

Open Science Framework



<https://zenodo.org/>
<https://osf.io/>
<https://figshare.com/>

National or institutional data repositories



NIRD RESEARCH DATA ARCHIVE

<https://dataverse.no/>

<https://sikt.no/>

<https://www.nsd.no/en/archiving-research-data>

<https://archive.norstore.no/>

Discipline-specific repositories

<https://www.re3data.org/>

Search

Browse ▾

Suggest

Resources ▾

Contact



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Search...

🔍 Search

<https://www.re3data.org/>

Research data

- Time and place

Research data

Good research data management plays an essential role in Open Science, which is increasingly important for scholars around the globe. The University Library provides courses on key concepts as a part of the Skills Development for Research Data project.

We can provide training on the following topics:

- Introduction to research data management (RDM)
- Current data management policies (UiO, Research Council of Norway, EU)
- Data management planning
- Data organization
- Metadata and documentation
- Data classification and storage
- Sharing and archiving research data
- Copyright and licensing
- Data discovery and reuse

Need help with research data?

Send an email to
research-data@uio.no

Open research plans



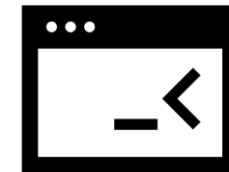
Open materials



Open data



Open analyses



Open publishing



Open research plans



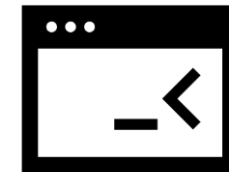
Open materials



Open data



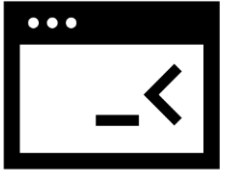
Open analyses



Open publishing

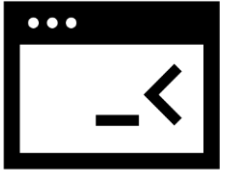


Open analyses: why?

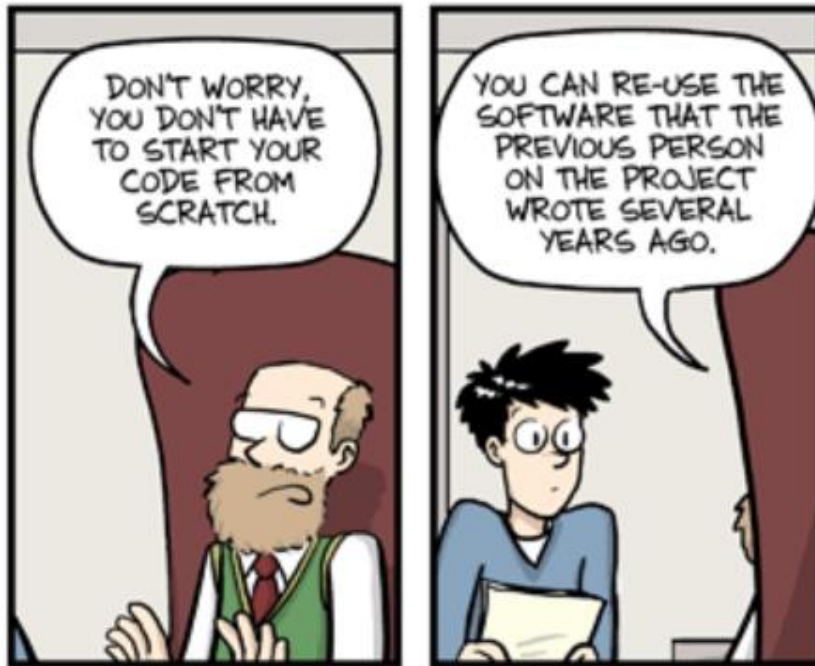


- Makes it easier for you to **re-use** your analyses/code
- Makes it possible for others to **re-run** your analyses
- ...and spot mistakes (if there are any)
- ...or use your code as an inspiration!
- Allows for **verification** of the results (e.g. during peer-review process)

Open analyses: how?

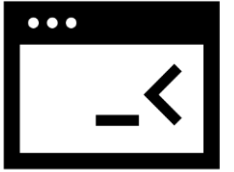


Piled Higher and Deeper by Jorge Cham



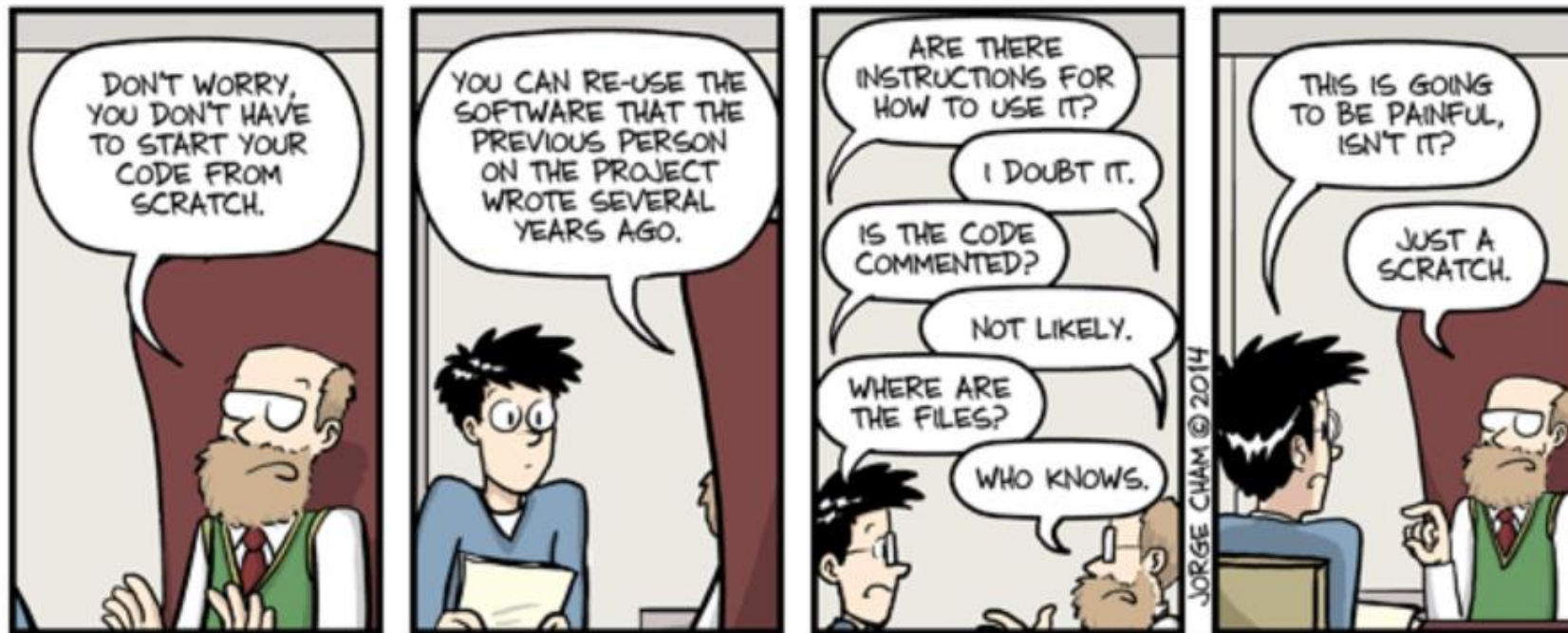
title: "Scratch" - originally published 3/12/2014

Open analyses: how?



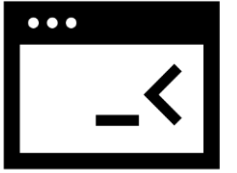
Piled Higher and Deeper by Jorge Cham

www.phdcomics.com



title: "Scratch" - originally published 3/12/2014

Open analyses: how?



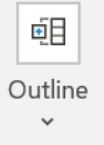
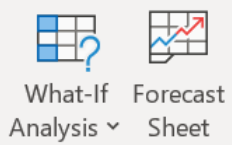
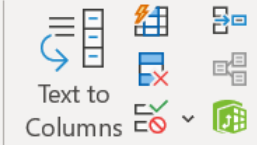
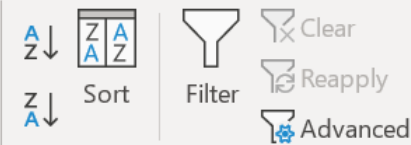
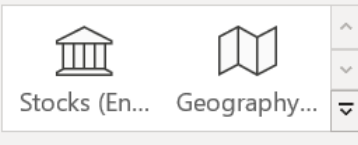
- If you use point-and-click software (Excel, SPSS) write notes that reflect your **workflow**

File Home Insert Draw Page Layout Formulas **Data** Review View Help

Share Comments



Refresh
All ▾



Get & Transform Data

Queries & Connections

Data Types

Sort & Filter

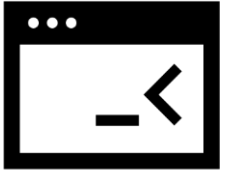
Data Tools

Forecast

A1     f_x 1

[illegible]

Open analyses: how?

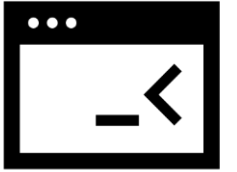


- If you use point-and-click software (Excel, SPSS) write notes that reflect your **workflow**
- If you use code, **annotate** it clearly

```
57 {r}
58 # Main Analysis Data
59
60 #load wide format data and preview
61 sum_data <- read.csv("Data/Experiment2_SumData.csv")
62 head(sum_data)
63
64 #check summary statistics for the dataset
65 describe(sum_data)
66 {r}
```

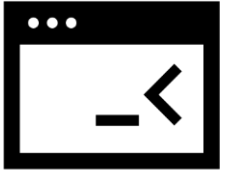
```
67
68 {r}
69 # Main Analysis
70
71 #ttest on the total proportion looking to shape change against chance (0.5)
72 t.test(sum_data$ShapeProportion, mu=0.5)
73 sd(sum_data$ShapeProportion)
74 se <- sd(sum_data$ShapeProportion)/sqrt(length(sum_data$ShapeProportion))
75 se
76
77 #compute the effect size (Cohen's D)
78 cohensD(sum_data$ShapeProportion, mu=0.5)
79
80
81 # Bayesian ttest on the total proportion looking to shape change against chance (0.5)
82 testMain <- ttestBF(sum_data$ShapeProportion, mu=0.5)
83 testMain
84 sd(sum_data$ShapeProportion)
85 se <- sd(sum_data$ShapeProportion)/sqrt(length(sum_data$ShapeProportion))
86 se
87 {r}
```

Open analyses: how?

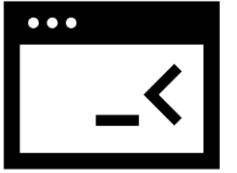


- If you use point-and-click software (Excel, SPSS) write notes that reflect your **workflow**
- If you use code, **annotate** it clearly
- Use **version control** system
- Put your analyses and/or code in **shareable notebooks or websites** (e.g. Jupyter, GitHub)
- Consider using **containers** (capture computational environment)

Open analyses: where?



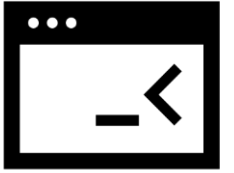
Open analyses: where?



- Git: Free and open source version control system
- GitHub: is an internet hosting service for software development and version control using Git



Open analyses: where?



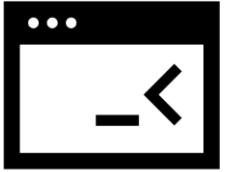
[Try](#) [Install](#) [Get Involved](#) [Documentation](#) [News](#) [Governance](#) [Security](#) [About](#)



Free software, open standards, and web services for interactive computing across all programming languages

<https://jupyter.org/>

Open analyses: where?



https://jupyter.org/try-jupyter/lab/

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/ notebooks /

Name	Last Modified
Intro.ipynb	7 days ago
Lorenz.ipynb	7 days ago
sqlite.ipynb	7 days ago

Intro.ipynb

See Also: For a more in-depth tour of Jupyter Notebooks and the Classic Jupyter Notebook interface, see [the Jupyter Notebook /Python tutorial on Binder](#).

An example: visualizing data in the notebook ✨

Below is an example of a code cell. We'll visualize some simple data using two popular packages in Python. We'll use [NumPy](#) to create some random data, and [Matplotlib](#) to visualize it.

Note how the code and the results of running the code are bundled together.

```
[1]: from matplotlib import pyplot as plt
import numpy as np

# Generate 100 random data points along 3 dimensions
x, y, scale = np.random.randn(3, 100)
fig, ax = plt.subplots()

# Map each onto a scatterplot we'll create with Matplotlib
ax.scatter(x=x, y=y, c=scale, s=np.abs(scale)*500)
ax.set(title="Some random data, created with JupyterLab!")
plt.show()
```

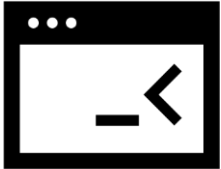
Some random data, created with JupyterLab!

Simple 0 1 Pyolite | Idle

Mode: Command Ln 1, Col 1 Intro.ipynb

<https://jupyter.org/>

Open analyses: where?



UNIVERSITY
OF OSLO

Menu

← Services and tools ← IT services

IT Services for Research

Nonwegian
version of this
page

Research Platforms →	Computations (HPC) →	Storing Research Data →
Research Statistics →	Data Collection and Analysis →	Research Data Management →
Publish, Write and Cite →	Presenting Data in GIS and Maps →	

Carpentry: Training on foundational coding and data science skills

At University of Oslo (UiO), Carpentries workshops based on Software Carpentry, Data Carpentry, and Library Carpentry lessons are offered to facilitate sharing and re-using of code and data among graduates and researchers from all faculties and units.

Carpentry: R for Reproducible Scientific Analysis (Novices)

May 31, 9:00 AM, Ole-Johan Dahls hus, seminarrom Perl

Carpentry@UiO: Plotting and Programming with Python (Novices)

June 1, 9:00 AM, Kristine Bonnevie's hus, rom 3127

Open research plans



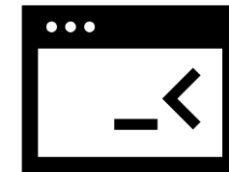
Open materials



Open data



Open analyses



Open publishing



Open research plans



Open materials



Open data



Open analyses



Open publishing



Open publishing: why?



Open publishing: how?



What requirements apply? Funds from RCN, EU, etc.? Institutional requirements?

- <https://www.ub.uio.no/english/writing-publishing/open-access/research-funders.html>

Find a publisher, check UiO's website if we have an agreement.

- <https://www.ub.uio.no/english/writing-publishing/open-access/deals-and-discounts/index.html>

If you do not publish open access (OA), check the journal's policy for self-archiving.

- <https://v2.sherpa.ac.uk/romeo/>

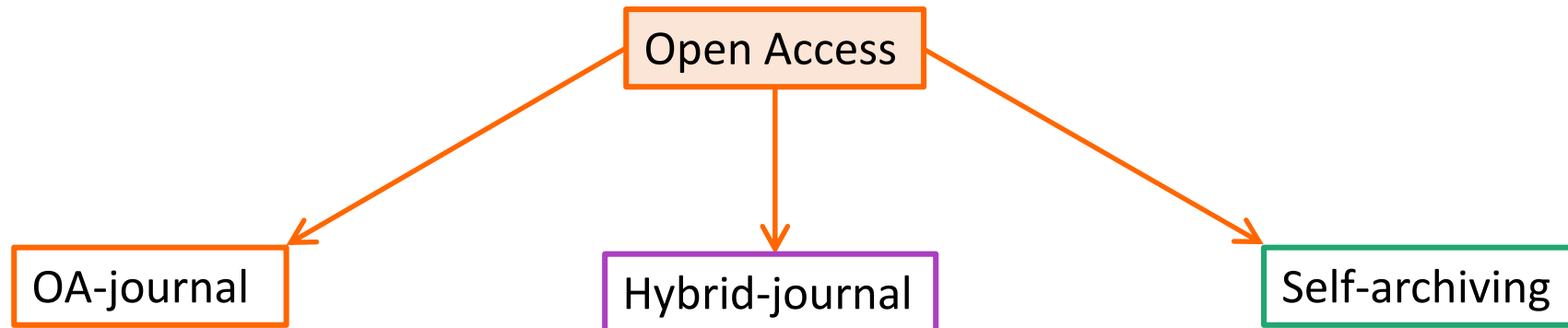
Is it a legitimate journal?

- Check [Kanalregisteret](#); if Gold OA check [DOAJ](#); Read more here: [Think. Check. Submit.](#)

Plan S (RCN funds)?

- Use [Journal Checker Tool](#) to see if the journal is compliant.

Open publishing: where?



- Without APC = **Diamond OA**
- With APC = **Gold OA**



<https://doaj.org/>

- Traditional journals → reseracher pays APC to open up an article → library pays access
- With APC
- Majority of them

= **Green OA**

- Publishing in a traditional journal with archiving of one version of your article in DUO via Cristin
- Always check journal's/publisher's rules about which article-version can you self-archive

Sherpa Romeo

<https://v2.sherpa.ac.uk/romeo/>

Open Access

[Norwegian
version of this
page](#)

Open Access

- [Self-archiving](#)
- [Plan S](#)
- [Publishing deals and discounts](#)
- [Publish Open Access](#)



It is the University of Oslo's goal that research results shall be openly available to individuals, the public sector, trade and industry, and the global research community.

Publishing Open Access? →

How do I publish Open Access, and how do I avoid fraud?

Requirements for Open Access →

Open Access requirements from University of Oslo, Plan S / Research Council of Norway, EU and the Government.

Open Access-agreements →

UiOs deals and discounts for Open Access publishing.

Research Data Management →

The UiO policy and guidelines. Help and advice on managing research data.

Self-archiving →

How to self-archive in DUO research archive via Cristin. Information on self archiving and other archives.

About Open Access →

What is Open Access? Which advantages does it give you? The paths to Open Access.

Need help with open access?

Send an email to
openaccess@ub.uio.no

Open research plans



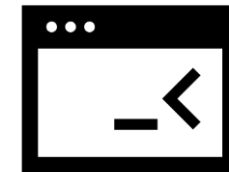
Open materials



Open data



Open analyses



Open publishing



**Open methods, data and
analyses are essential for
reproducibility**

Reproducibility means
obtaining identical results
with the same data

"For my first work-related tweet of 2020, I am totally bummed to announce that we have retracted last year's paper on enzymatic synthesis of beta-lactams. **The work has not been reproducible**"

"It is painful to admit, but important to do so. I apologize to all. I was a bit busy when this was submitted and did not do my job well."

BBC Sign in Home News Sport Reel Worklife Tr

NEWS

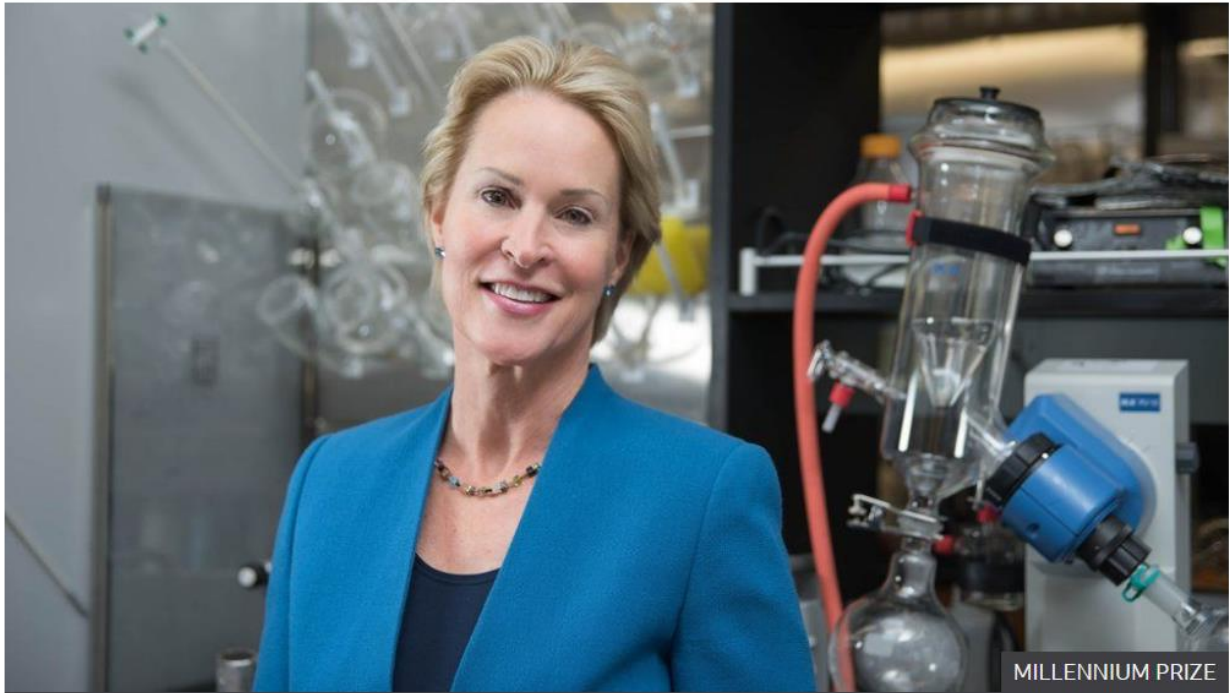
Home Coronavirus Climate Video World UK Business Tech Science Stories Entertainment & Ar

World Africa Asia Australia Europe Latin America Middle East US & Canada

Nobel Prize-winning scientist Frances Arnold retracts paper

🕒 3 January 2020

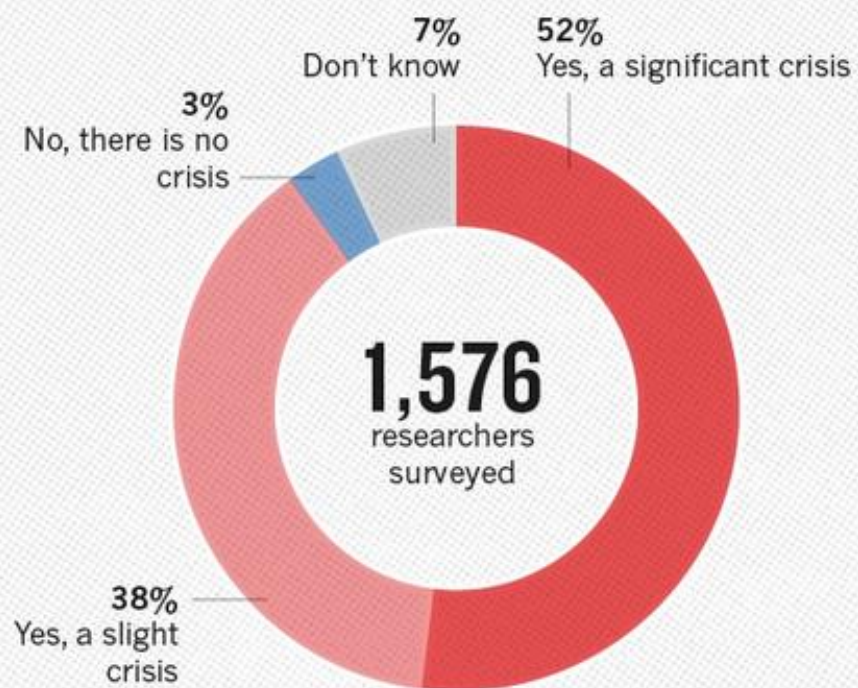
 Nobel Prize



MILLENNIUM PRIZE

Prof Arnold works in the department of chemical engineering at Caltech

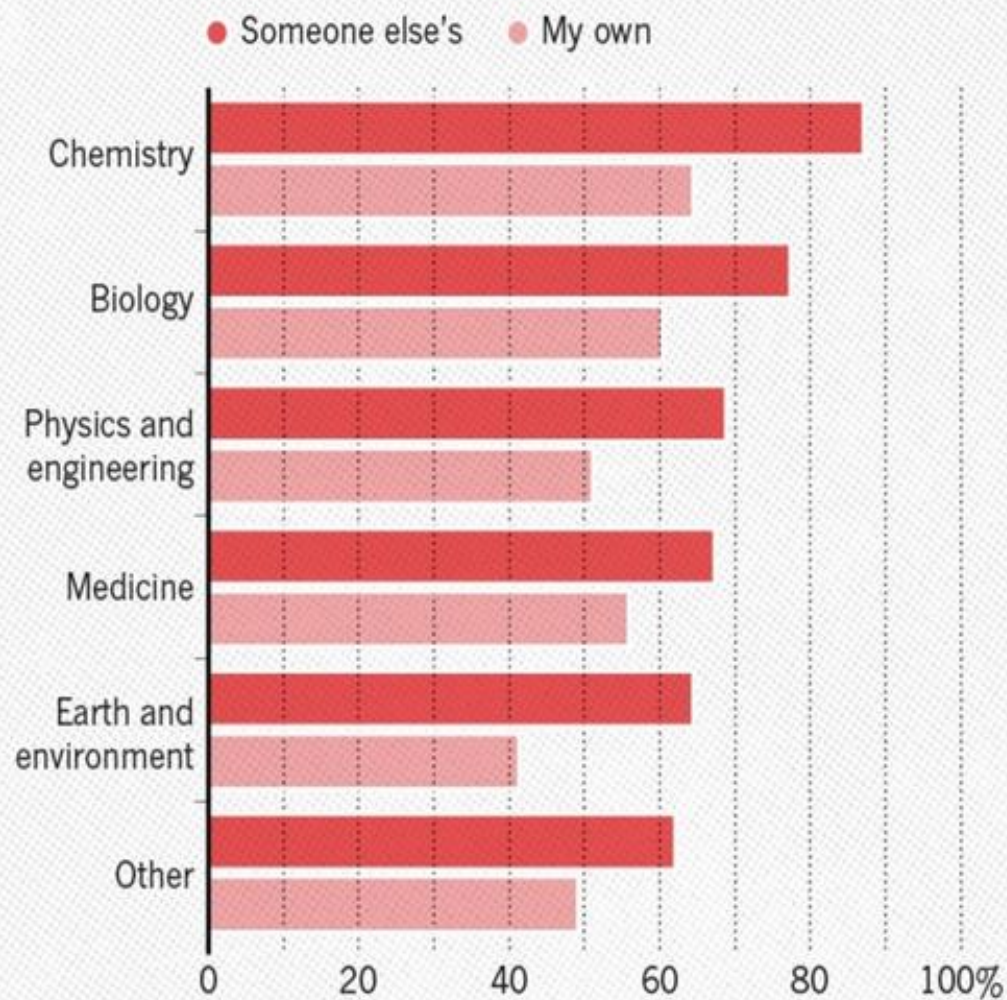
IS THERE A REPRODUCIBILITY CRISIS?



©nature

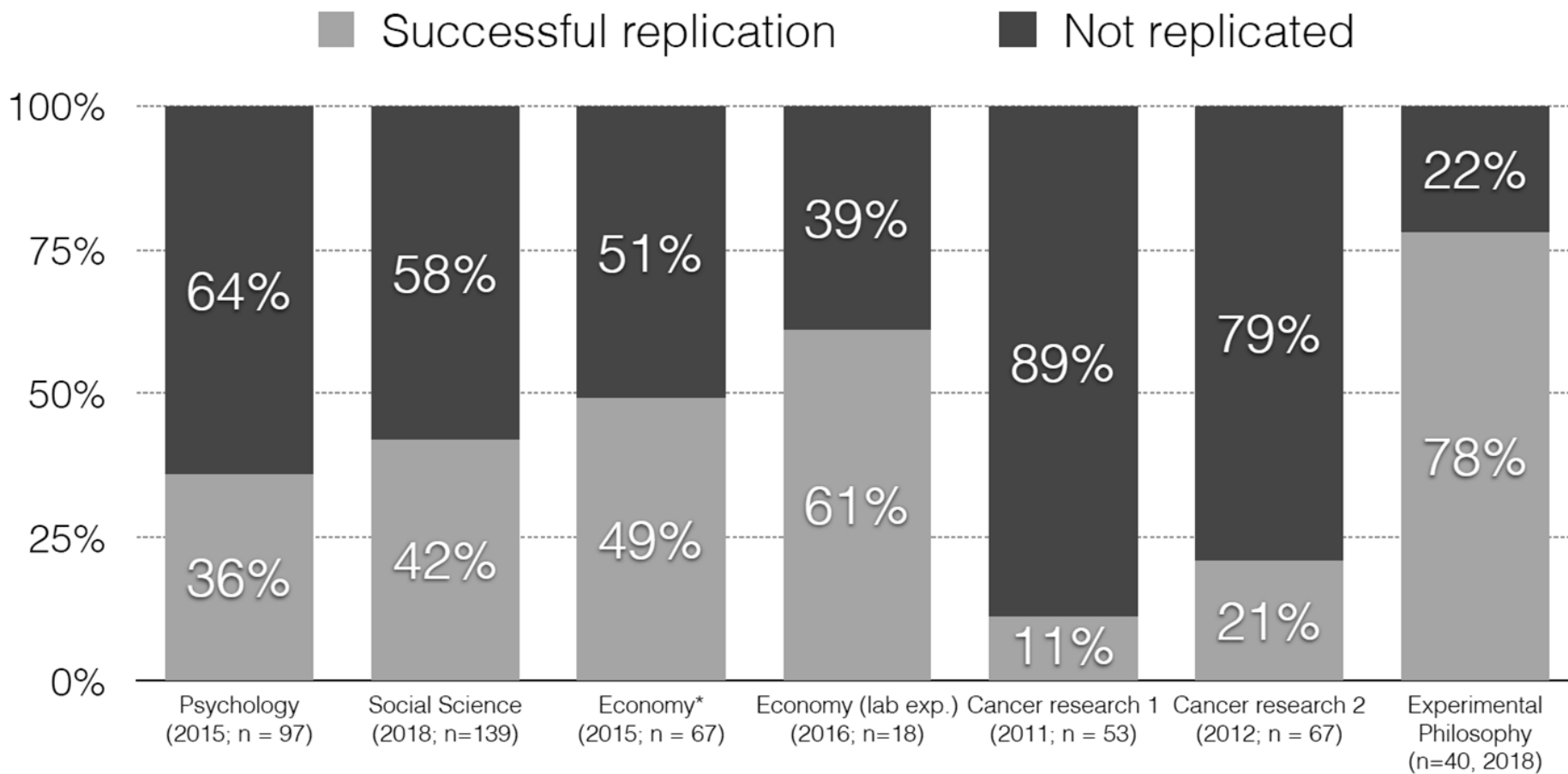
HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



Reproducibility means
obtaining identical results
with the same data

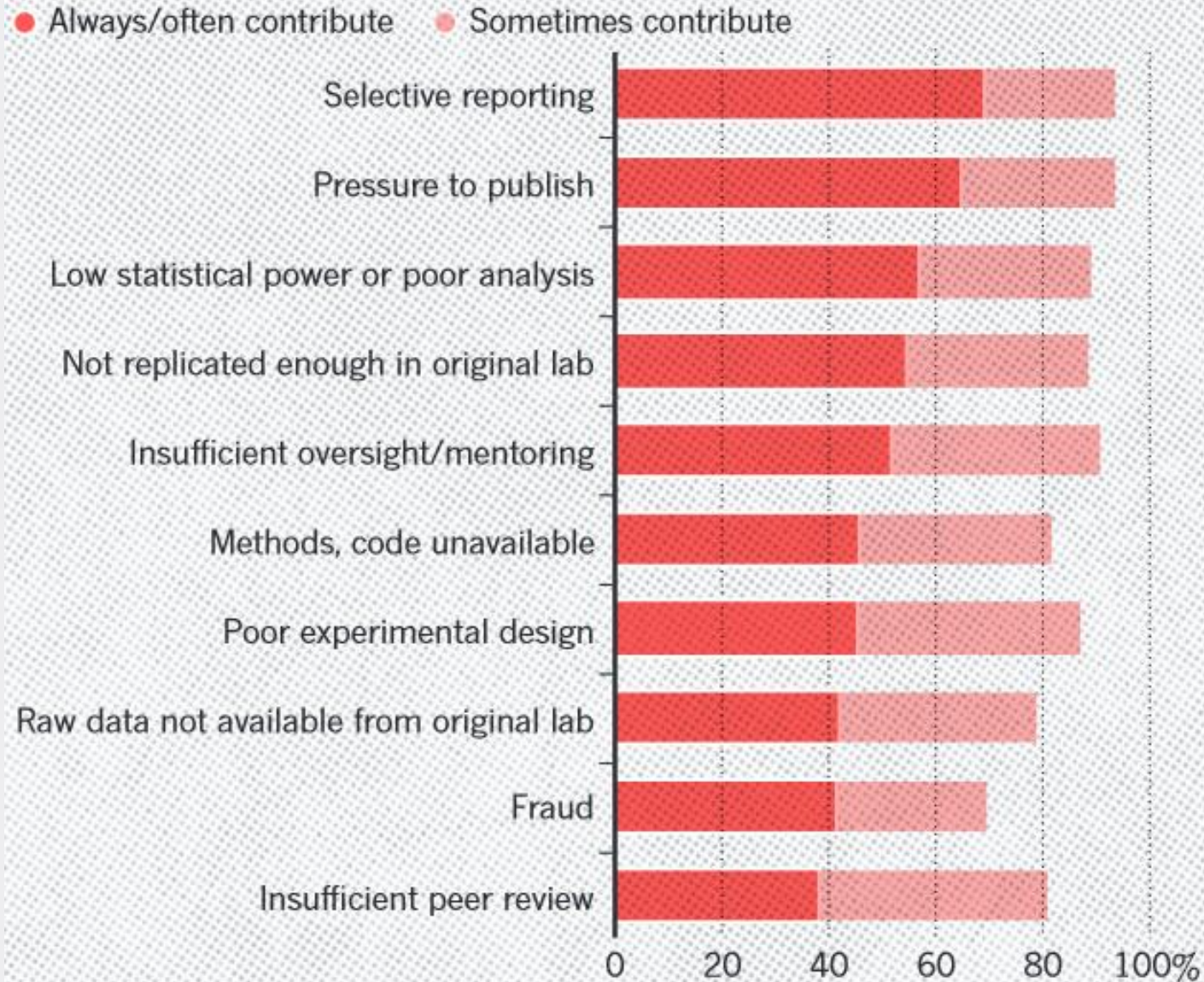
Replication means
obtaining similar results
with new data



Open Science Collaboration (2015); Social Science: Combined sample of systematically sampled projects (RPP, SSRP, EERP); Chang & Li (2015); Camerer et al (2016); Begley, C. G., & Ellis, L. M. (2012). Prinz, F., Schlange, T., & Asadullah, K. (2011); Cova et al. (2018)

WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

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



What can be done?

- Be **transparent** about your research plans, methods, procedures, and analyses (e.g. by preregistering your studies)
- Share the exact research **materials** you use (if you can)
- Share your **data** (ideally raw data if you can)
- Make sure you have good **documentation** for all stages of your research process: methods, data, analyses/code
- **Verify** your own work: try to reproduce your own results (or have others do it)

**Reproducible does not
(have to) mean fully open**

As open as possible,
as closed as necessary

Welcome to Norway's Reproducibility Network

Towards open & reproducible science

JOIN US

Our Mission

The Norwegian Reproducibility Network (NORRN) is a peer-led network that aims **to promote and enable rigorous, robust and transparent research practices in Norway**. We attempt to achieve this goal by establishing appropriate training activities, designing, and evaluating research improvement efforts, disseminating best practices, and working with stakeholders to ensure coordination of efforts across the sector. NORRN's activities span multiple levels, including researchers, librarians, institutions, and other stakeholders (e.g., funders and public authorities).



Journal Clubs

Our mailing list:

<https://sympa.uio.no/uio.no/subscribe/open-science-oslo>

University of Oslo

Welcome

Our journal club is open to both staff and students at UiO across all departments. Everyone is welcome to join us - whether you are an enthusiast, a skeptic, or a cautious explorer. Feel free to get in touch if you are interested to participate or contribute!

Our mission

- provide an informal and friendly platform for discussions about meta scientific topics
- help each other get familiarized with open science practices (e.g., pre-registrations, sharing data, sharing preprints, etc.)
- connect students and researchers from various disciplines who are interested in meta science.

Format

Before each meeting, we read an article on meta scientific topics, which we then discuss during the meeting. At each meeting, a different discussion leader will begin by providing a short overview of the paper and facilitate discussion throughout the meeting. Anyone can propose a possible paper or topic to present if we have not covered it already. Grab a cup of tea (coffee?) and join us! Due to the current pandemic, all sessions will be held online (Zoom) for the time being.

 Universitetet i Oslo

 osf.io/mvx54

Z [collection](#)

? timo.b.roettger@gmail.com

 Timo B. Roettger

 Agata Bochynska

Digital Scholarship Center

UiO's Digital Scholarship Center (DSC) can help you take advantage of digital tools and methods in your research. We offer support with data management, visualization, digital research methods, and digital skill development.



Research Data Management at UiO

Having proper Research Data Management (RDM) routines will help you save time, enables sharing and reuse of data, and promotes open science. However, RDM in a digital context can be challenging and complex. We offer guidance and consultation on RDM topics including:

- data management plans (DMP)
- classification and storage
- organization, documentation, and metadata
- legal questions (i.e. privacy, copyright, and licensing)
- archiving and sharing

Digital research methods

On these pages you will find overviews and user guides on digital research methods for researchers and students. There will be information about relevant digital research methods in many subject areas.

Currently, we have information about text mining - or digital research methods for those working with text data.

Need advice?

Visit us at the room 230, in the 2nd floor of Niels Henrik Abels hus.

Book an appointment



Or send us an email to research-data@uio.no

Digital scholarship communities

Interested in teaching or helping at workshops? Join us in:

- Carpentries@UiO
- [CodeRefinery - Get Involved](#)
- [UiO Data Managers Network](#)
- [ReproducibiliTea Journal Club](#)

Do you have any suggestions for workshops or seminars or questions about the workshops? Send an email to digitalscholarship@ub.uio.no

Thank you!

Questions?

Agata Bochynska, PhD

Open Research and Digital Scholarship Center
University of Oslo Library

@AgataBochynska

agata.bochynska@ub.uio.no