Aalto University

Reproducibility, replicability, QRPs ...and possible cures

Enrico Glerean, DSc, Staff scientist (Aalto School of Science), @eglerean, www.glerean.com

28/04/2021

30 seconds about me

- **Data agent for Aalto RES** technical expert on personal data for research (GDPR), data anonymisation, ethics, open science **https://www.aalto.fi/en/services/rdm-training**
- Core member of **Aalto Scientific Computing/Science-IT team https://scicomp.aalto.fi/**
- **Neuroscientist (fMRI)** and behavioural data fan **https://scholar.google.fi/citations?hl=en&user=sD90SmMAAAAJ&view_op=list_works**

Outline of today's workshop

1. **What is responsible conduct of research (RCR)?**
2. **What is research misconduct? From FFMP to QRP**
3. **Why do we care? And why do researchers engage with unethical research practices?**
4. **How can we fix things?**

Focus is on researchers at organizations following TENK guidelines. Similar considerations affect students, teachers, policy makers, company researchers, etc…

This is work in progress, let's improve it together!

There is more material than I can cover, but you have ALL the links!

Let’s start with the references

Where to read and learn

From TENK:

https://www.tenk.fi/en/responsible-conduct-of-research

Reproducibility (quantitative methods):

https://www.slideshare.net/deevybishop/what-is-the-reproducibility-crisis-in-science-and-what-can-we-do-about-it

Reproducibility (qualitative methods):

https://openworking.wordpress.com/2019/02/11/what-does-reproducibility-mean-for-qualitative-research/

Aalto Open Science and Research policy:

https://www.aalto.fi/en/open-science-and-research/aalto-university-open-science-and-research-policy

Review:

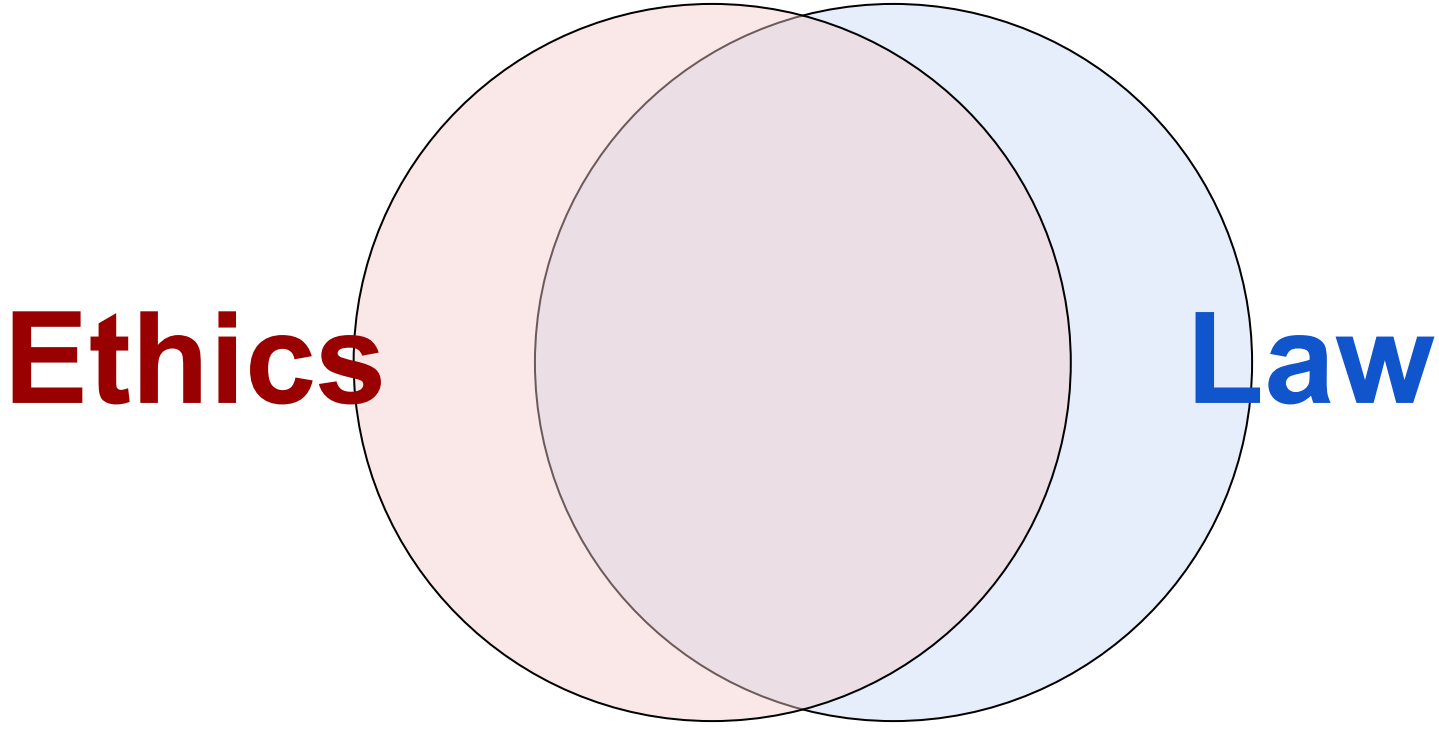
https://osf.io/preprints/metaarxiv/9hg3j

1. What is responsible conduct of research?

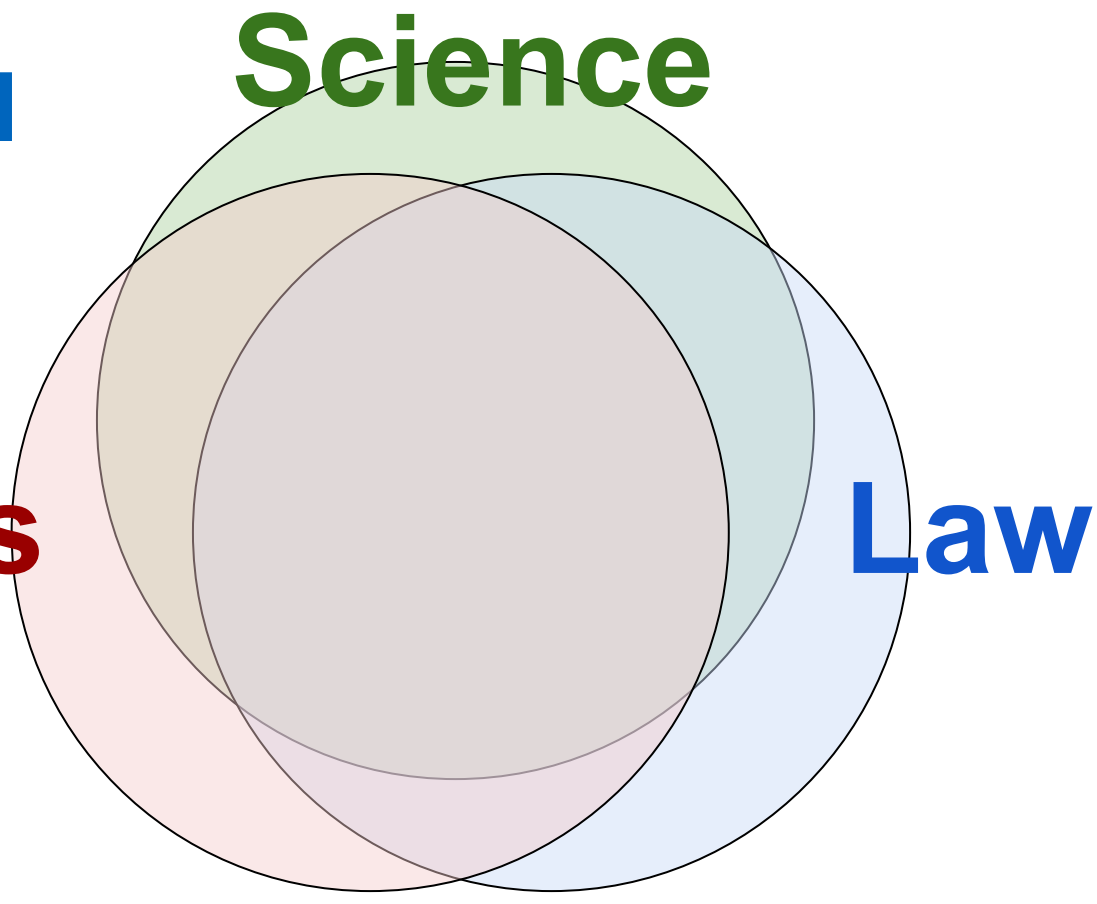
Responsible Conduct of Research

- **RCR touches ethics, law, and philosophy of science.**

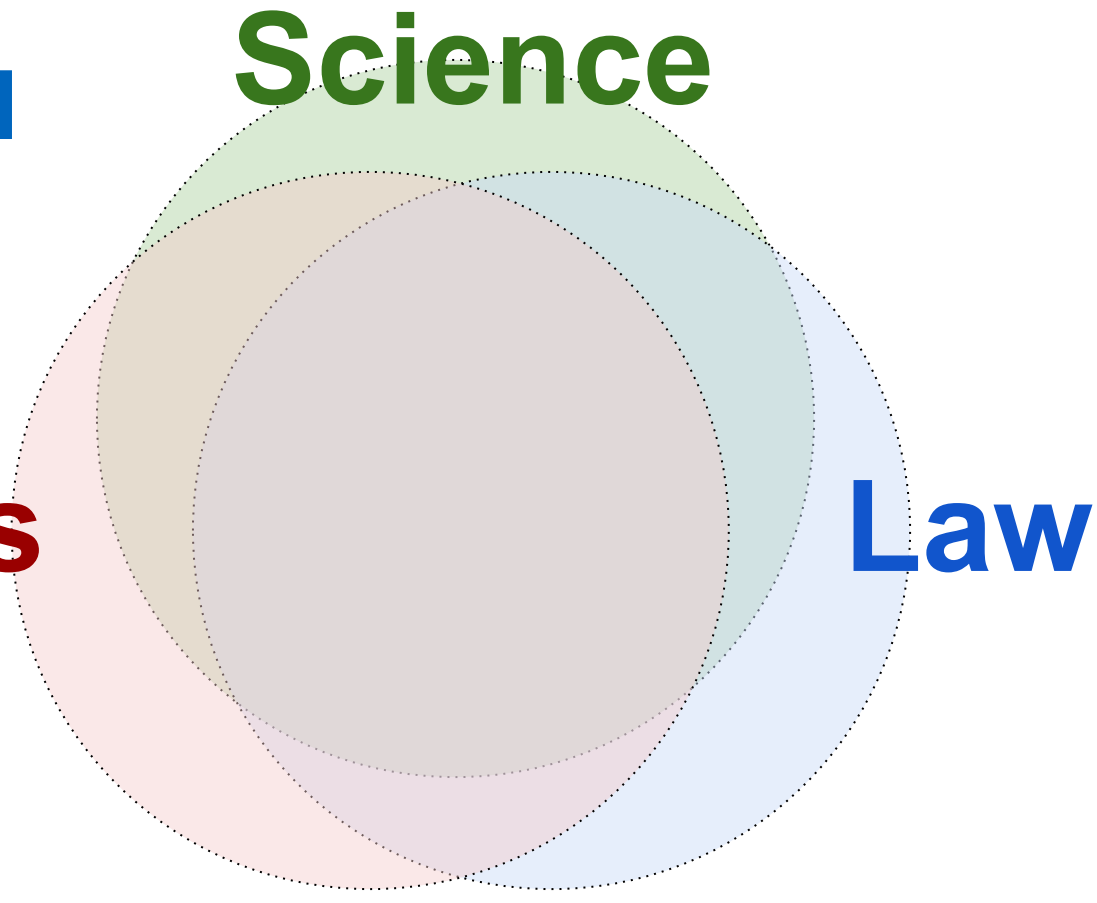
Ethics is not Law



Ethics, Law, and Science



Ethics, Law, and Science



Responsible Conduct of Research

- **RCR touches ethics, law, and philosophy of science.**
- **I find it challenging to define RCR (“***Do unto others as you would have them do unto you”***), it is easier to define by stating what it is not**
- **Research misconduct and questionable research practices:** once we all agree on what is deemed as research misconduct, we can identify ways to fix it, prevent it, and incentivise researchers towards practices that are against it.

2. What is research misconduct?

Research misconduct

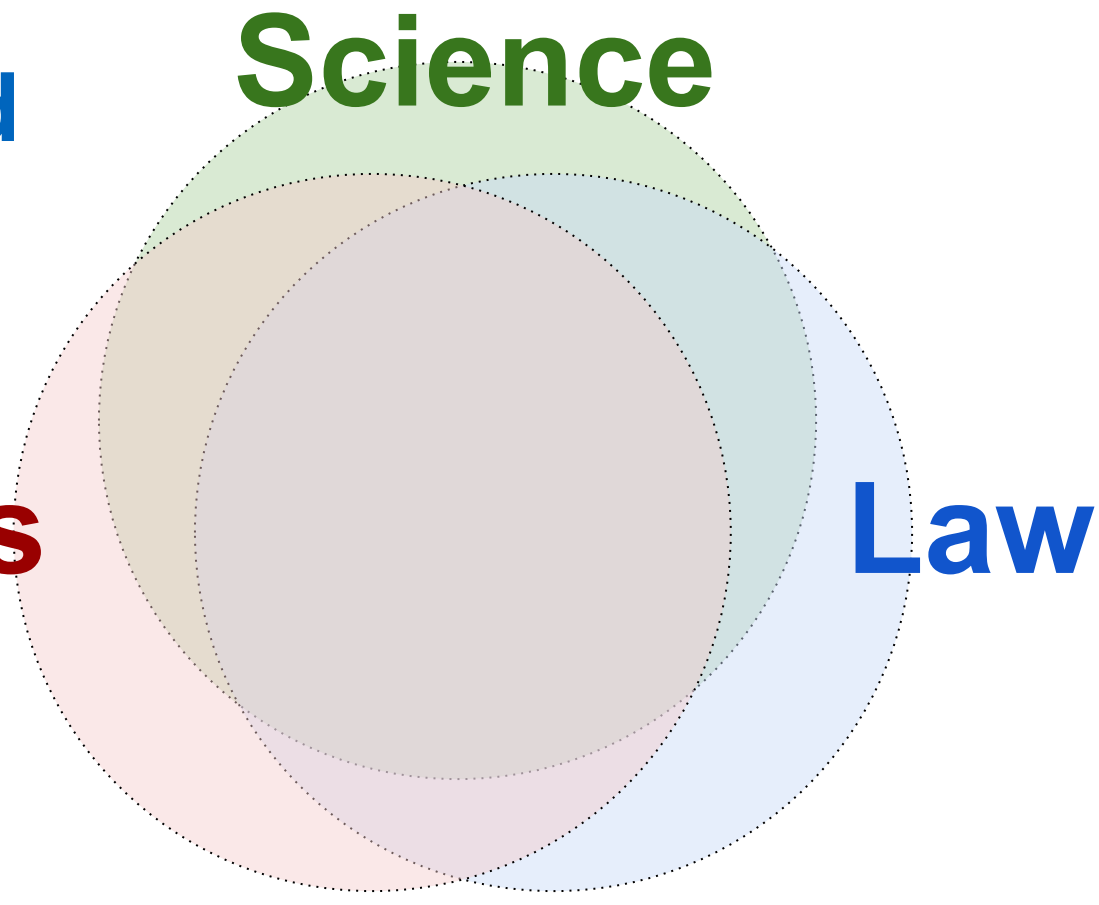
- **According to TENK 2012 guidelines**

1. **Fabrication** (false data)
2. **Falsification** (false results)
3. **Plagiarism** (stealing of other's materials)
4. **Misappropriation** (scooping)

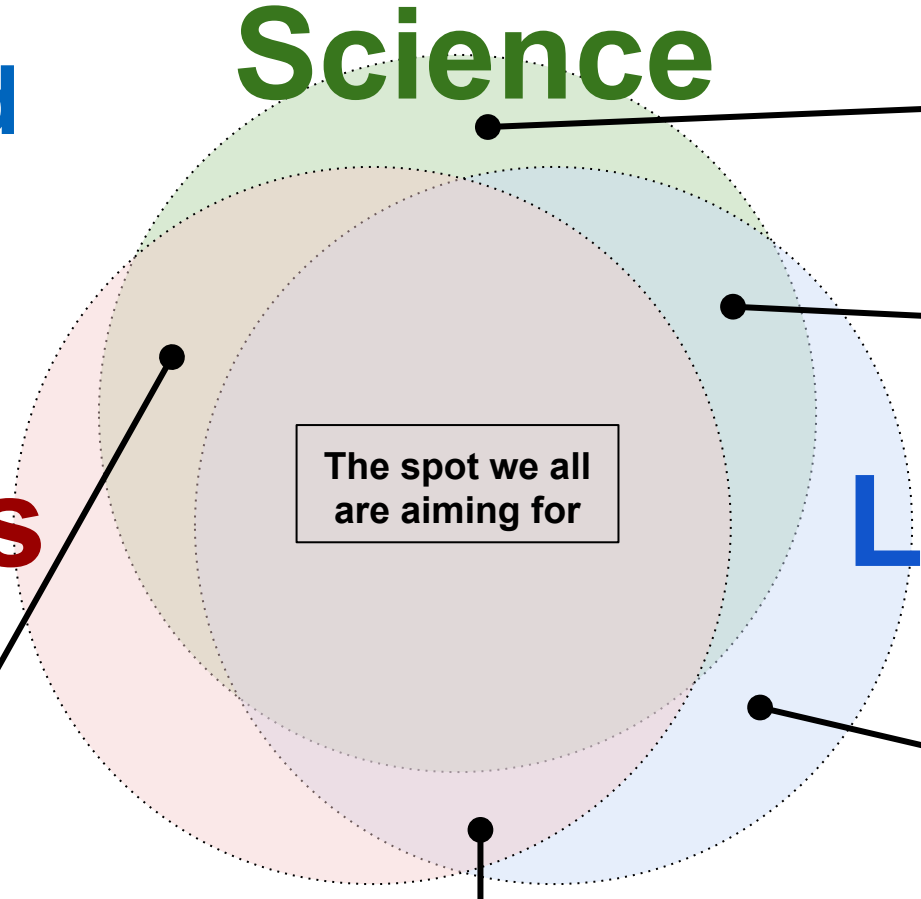
Research misconduct as disregarding RCR

- Harming others' works for profit (ignoring literature, unethical peer review, manipulating citation metrics, conflicts of interest)
- Misleading the general public (false findings excessively advertised)
- Questionable research practices (p-hacking, harking, publication bias, analysis bias)

Ethics, Law, and Science



Ethics, Law, and Science



Scientific misconduct FFPM

Questionable research practices (p-Hacking)

Sharing your findings breaking journal's copyright

Not sharing data or code that **could** be shared

Not sharing data or code that **should** be shared (ethical towards subjects but not towards science)

Research misconduct exercise

1. **Fabrication** (false data)
2. **Falsification** (false results)
3. **Plagiarism** (stealing of other's materials)
4. **Misappropriation** (scooping)

Exercise: Which one is the worst? How can you detect them?

FFPM

- **Falsification, Fabrication, Plagiarism, Misappropriation**
- **They can be detected with current technologies** although tools and other researchers can be also tricked.
- ***I think Falsification and Fabrication are the worst***
- **Solution: don't do it**

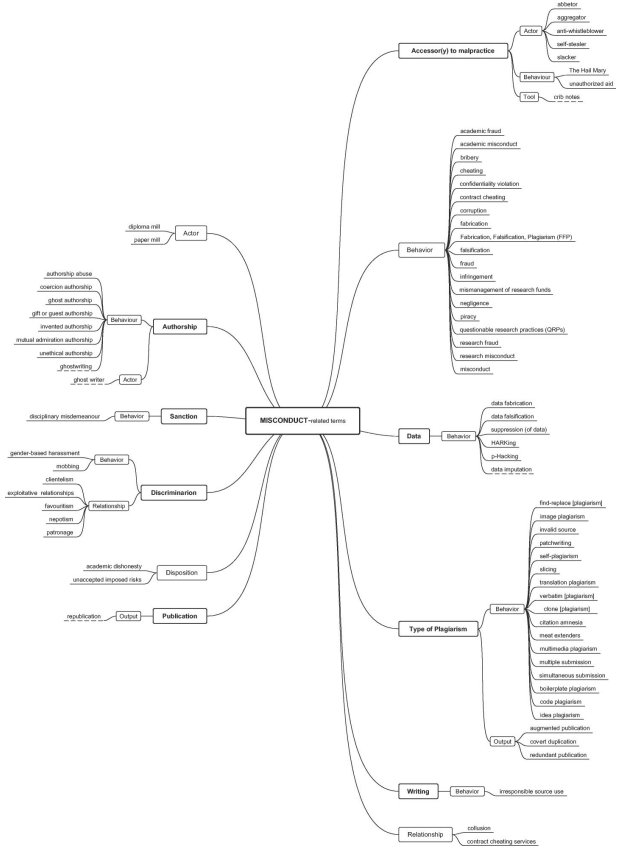
Funny recent plagiarism example from machine learning:
https://www.reddit.com/r/learnmachinelearning/comments/dh38x9/siraj_raval_has_a_new_paper_the_neural_qubit_its/

E.g. the GRIM test:
https://medium.com/@jamesheathers/the-grim-test-a-method-for-evaluating-published-research-9a4e5f05e870

More about Brown and Heathers:
https://www.sciencemag.org/news/2018/02/meet-data-thugs-out-expose-shoddy-and-questionable-research

Taxonomy of the types of misconduct

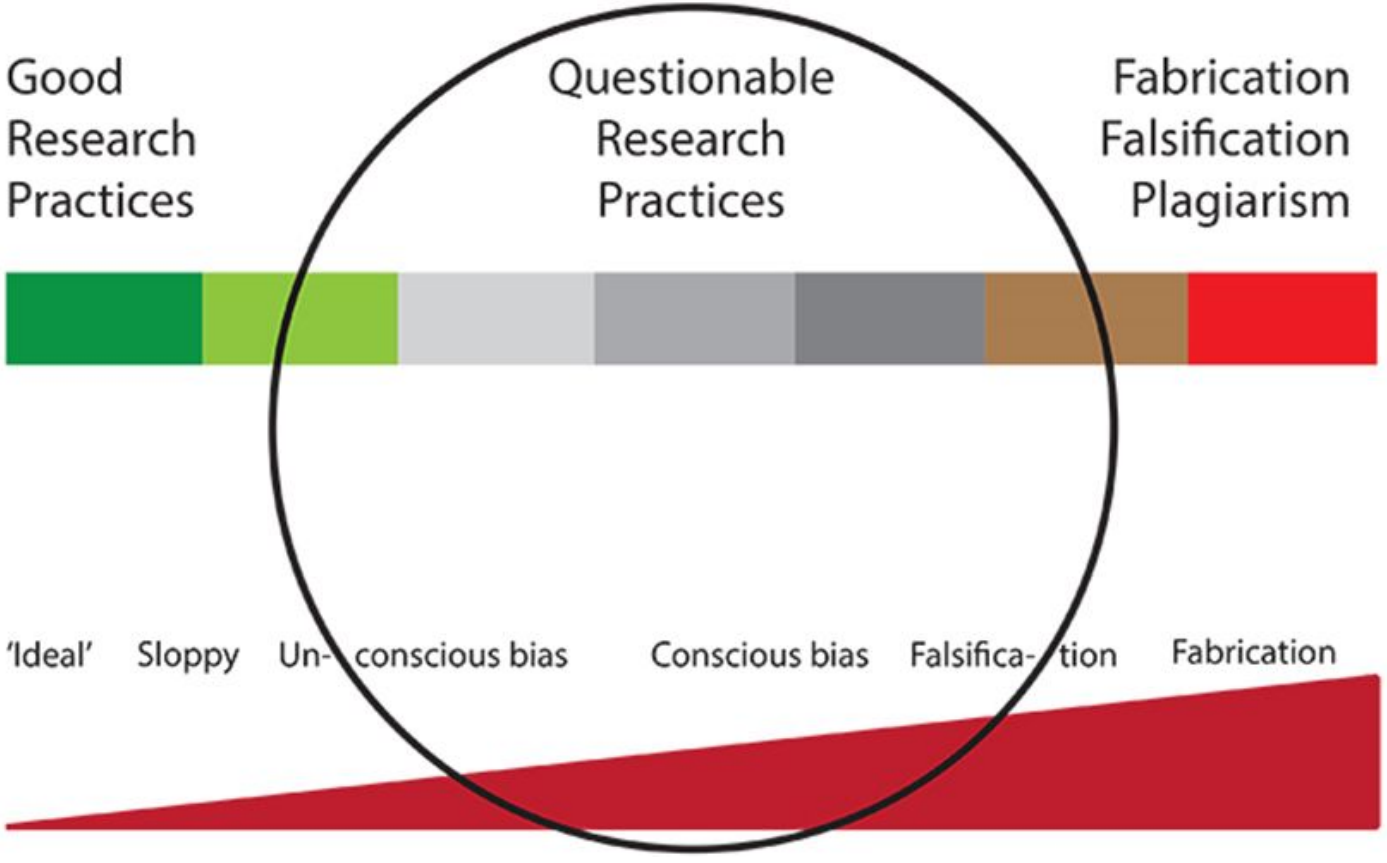
- Research misconduct is more than FFPM
- It touches all aspects of research activities
- We can define a continuum of good/bad practices



https://link.springer.com/article/10.1007/s10805-019-09342-4

2.b What should be considered research misconduct in 2021?

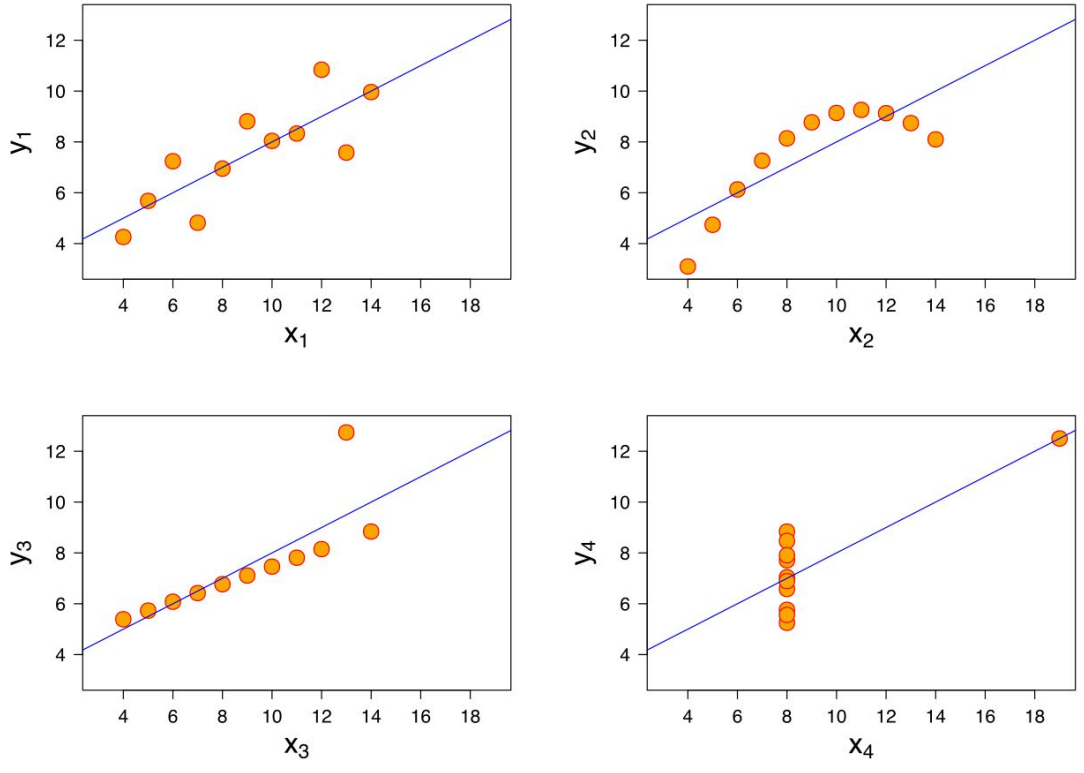
From integrity to misconduct



René Bekkers https://slideplayer.com/slide/7787128/

Examples: Honest mistakes

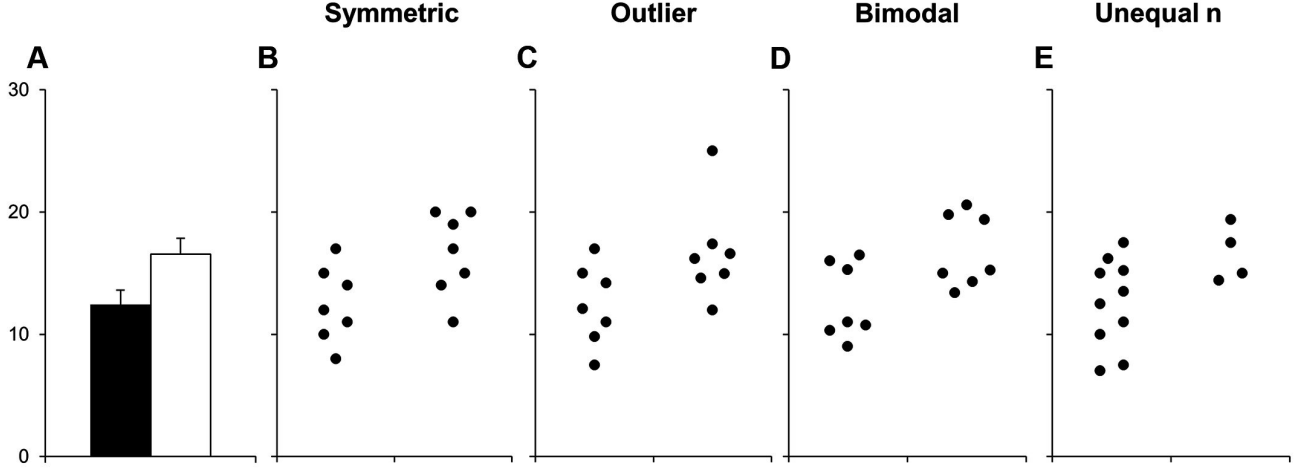
- **Reporting *r* = *0.816* while...**



https://en.wikipedia.org/wiki/Anscombe%27s_quartet

Example: Honest mistakes

- **Reporting a significant group difference while...**

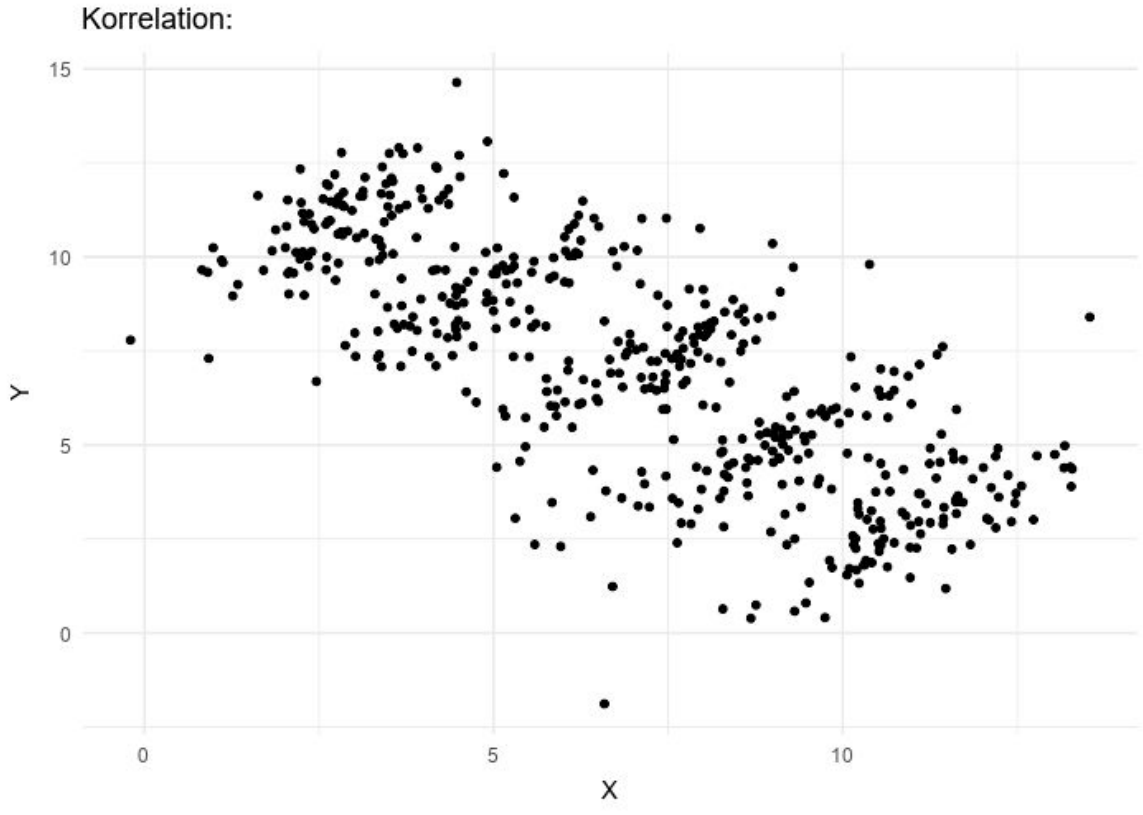


Test	p value			
T-test: Equal var.	0.035	0.050	0.026	0.063
T-test: Unequal var.	0.035	0.050	0.026	0.035
Wilcoxon	0.054	0.073	0.128	0.103

https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002128

Example: Honest mistakes

- **Reporting a negative correlation while...**
- **Solution: visualize all data & share visualization code**



https://en.wikipedia.org/wiki/Simpson%27s_paradox

2.c Four types of questionable research practices

Problems caused by researchers 1

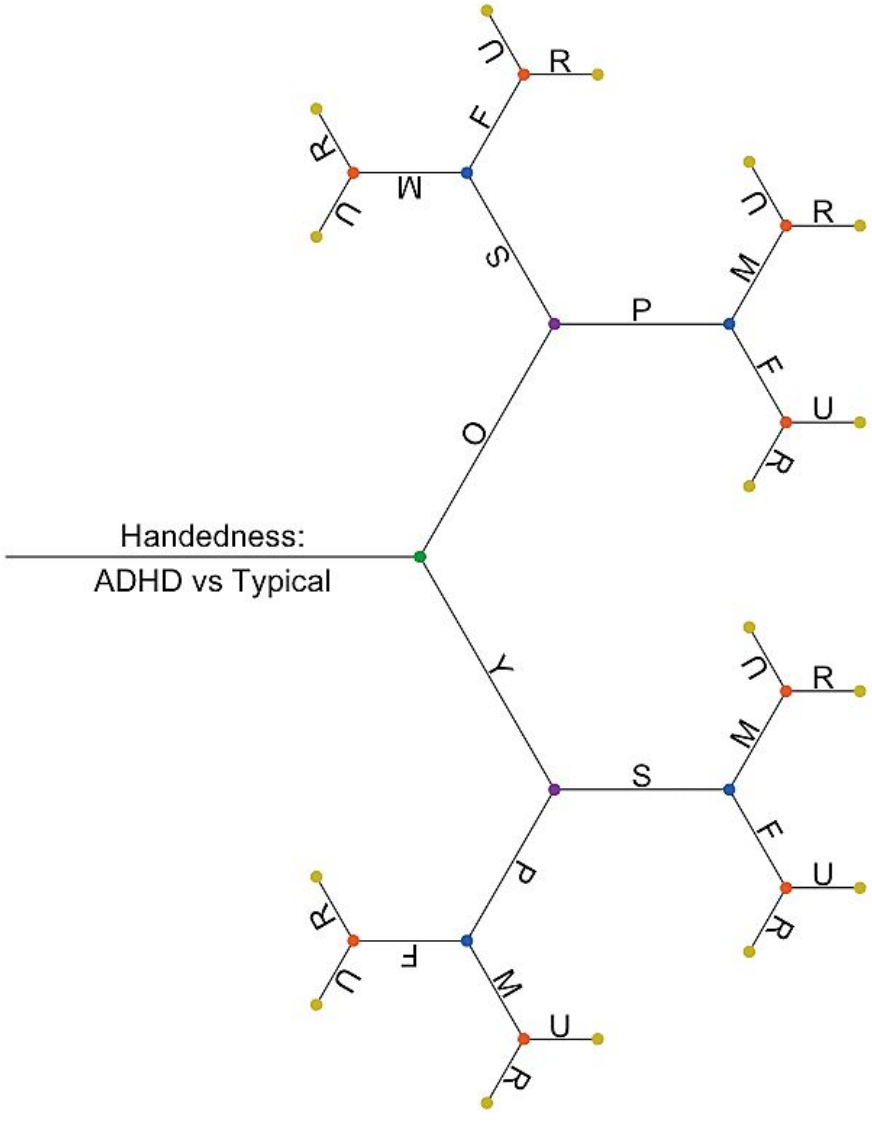
- Unconscious bias / Confirmatory bias / Seeing patterns that are not there
- … which leads to irreproducible findings or to the file drawer effect
- Solution: eradicate the subjectivity from methods, automate data collection and analysis, make data and methods open

https://www.slideshare.net/deevybishop/what-is-the-reproducibility-crisis-in-science-and-what-can-we-do-about-it

Problems caused by researchers 2

- **Failure to understand statistics**
- **… which leads to p-hacking (trying multiple analysis and report only those that reached statistical significance): huge bias for false positives (irreproducibility) and consequent publication bias (HARKing, file drawer effect)**

https://www.slideshare.net/deevybishop/what-is-the-reproducibility-crisis-in-science-and-what-can-we-do-about-it



The garden of forking paths

by Dorothy V M Bishop

An illustration of how opportunities for false positives can mount up when one has a large dataset and a flexible approach to analysis.

https://figshare.com/articles/The_Garden_of_Forking_Paths/2100379

Multiple comparisons (many statisticians)

File-drawer problem (*29*)

Pseudoreplication (*32*)

Significance questing (*33*)

Data mining, dredging, torturing (*34*)

Hypothesizing after the results are known (HARKing) (*30*)

Data snooping (*35*)

Selective outcome reporting (*36*)

Silent multiplicity (*37*)

Specification searching (*38*)

P-hacking (*31*)

https://stm.sciencemag.org/content/8/341/341ps12

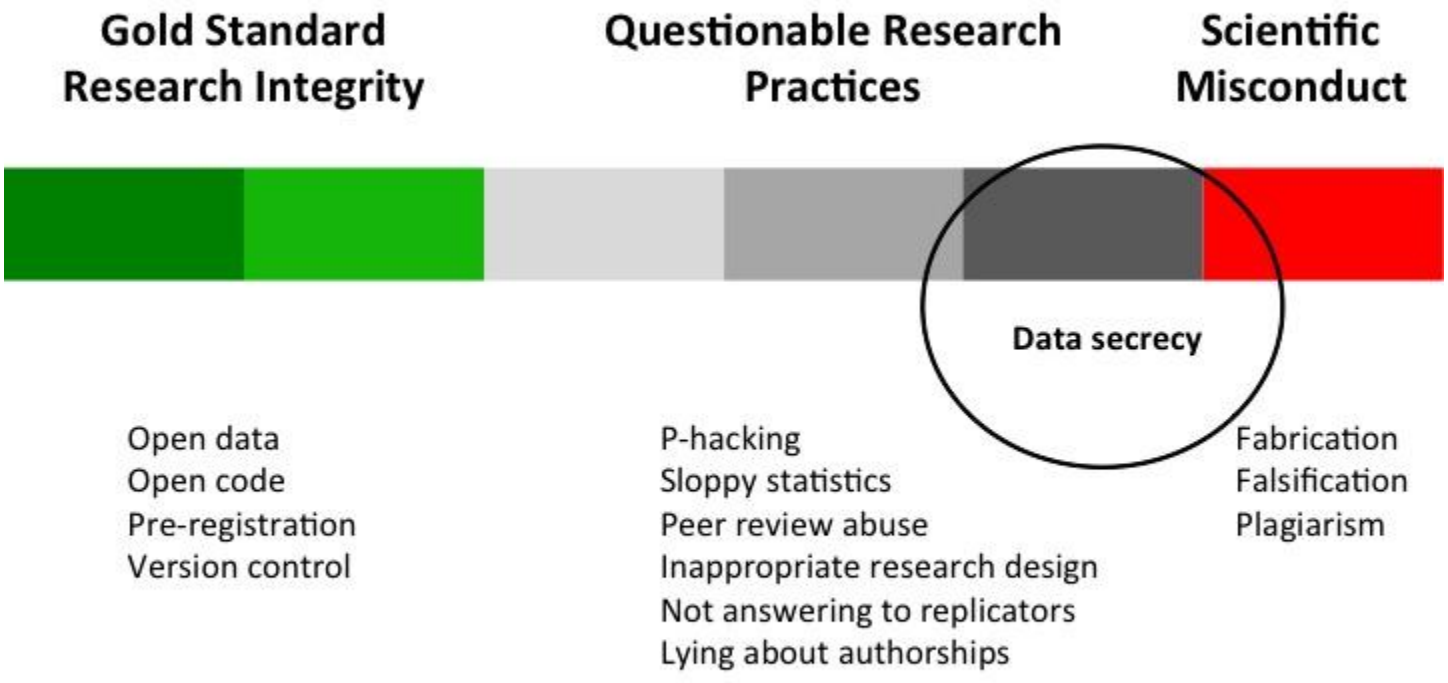
Problems caused by researchers 2

- Failure to understand statistics
- … which leads to p-hacking (trying multiple analysis and report only those that reached statistical significance): huge bias for false positives (irreproducibility) and consequent publication bias (HARKing, file drawer effect)
- Solutions: larger N, simulated data, separate replication dataset, blind analysis with masked data, pre-registration of analysis, registered reports, more stringent statistical thresholds

https://www.slideshare.net/deevybishop/what-is-the-reproducibility-crisis-in-science-and-what-can-we-do-about-it

Problems caused by researchers 3

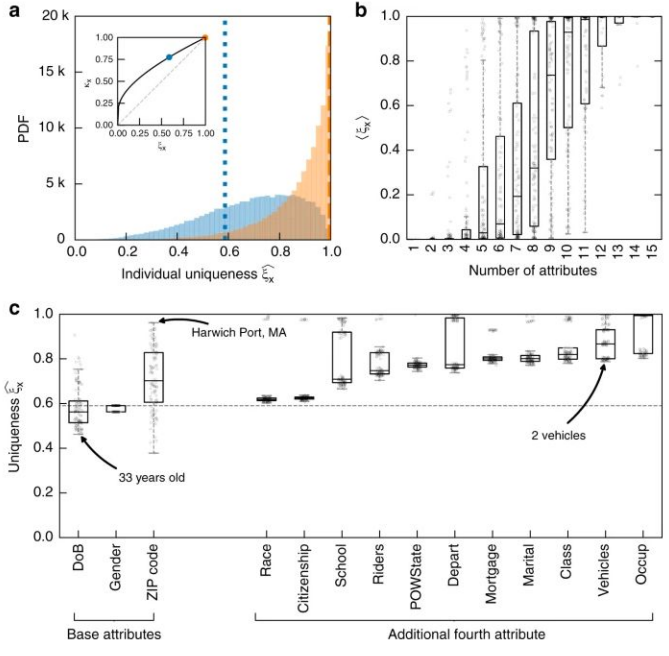
- **Data secrecy**



https://www.slideshare.net/deevybishop/what-is-the-reproducibility-crisis-in-science-and-what-can-we-do-about-it & https://blogs.lse.ac.uk/impactofsocialsciences/2015/07/03/data-secrecy-bad-science-or-scientific-misconduct/

Problems caused by researchers 3

- **When working with human research data, it is rarely possible to fully anonymise it**



- With ~15 pieces of information you can uniquely identify 99.98% of the American population.

"Estimating the success of re-identifications in incomplete datasets using generative models"
Rocher et al (2019)
https://www.nature.com/articles/s41467-019-10933-3

https://youtu.be/ILXeA4fx3cI
https://drive.google.com/file/d/1f9FSGTK3djJCynVm2UiipbHE2bwY71OL/view?usp=sharing

Problems caused by researchers 3

- **We need a process for opening personal data in a legal and ethical way**
  - From biobanks to databanks https://braindata.fi/
  - … but we should have easier faster and ways to share personal data, but how?
- **Workshop "Opening and sharing of personal data: ethical and legal issues and solutions" for the ENRIO 2021 Congress on Research Integrity Practice.**
- **Work in progress: https://psyarxiv.com/f6mnp/**

Aalto University

The Open Brain Consent: Informing research participants and obtaining consent to share brain imaging data

AUTHORS
Elise Bannier, Gareth Barker, Valentina Borghesani, Nils Broeckx, Patricia Clement, Maria de la Iglesia Vaya, Kyrre E. Emblem, Satrajit Ghosh, Enrico Glerean, Krzysztof Gorgolewski, Marko Havu, Yaroslav Halchenko, Peer Herholz, Anne Hespel, Stephan Heunis, Yue Hu, Hu Chuan-Peng, Dorien Huijser, Radim Jancalek, Vasileios Katsaros, Marie-Luise Kieseler, Camille MAUMET, Clara Moreau, Henk Mutsaerts, Robert Oostenveld, Esin Ozturk Isik, Nicolas Pascual Leone Espinosa, John Pellman, Cyril Pernet, Francesca Pizzini, Amira Trbalić, Paule-Joanne Toussaint, Matteo Visconti di Oleggio Castello, Fengjuan Wang, Cheng Wang, Hua Zhu

Problems caused by researchers 4

- **Questionable measurement practices!**
- **https://psyarxiv.com/hs7wm/ (Flake & Fried)**
- **https://www.youtube.com/watch?v=Cq6n7AS_r8w**
- **Solution: more robust theory behind what we measure**

Bonus reading:

- **"Why hypothesis testers should spend less time testing hypotheses"**
  https://psyarxiv.com/vekpu/ *"Before testing hypotheses, researchers should spend more time forming concepts, developing valid measures, establishing the causal relationships between concepts and their functional form, and identifying boundary conditions and auxiliary assumptions".*
- **"Questionable research practices may have little effect on replicability"**
  https://elifesciences.org/articles/58237 *"The analyses indicate that the base rate of true effects is the major factor that determines the replication rate of scientific results. Specifically, for purely statistical reasons, replicability is low in research domains where true effects are rare (e.g., search for effective drugs in pharmacology)."*

Summary so far (not a full taxonomy)

QRP	Solution
Unconscious or confirmatory bias	Eradicate the subjectivity from methods, automate data collection and analysis, make data and methods open. For a list of biases https://catalogofbias.org/biases/
Methodological (statistical) issues	Larger N, simulated data, separate replication dataset, blind analysis with masked data, pre-registration of analysis, more stringent statistical thresholds
Data secrecy	Data is open or not, nothing in between. New ways of sharing personal research data in the age of GDPR
Questionable measurement practices	More robust theory, more robust measurement, focusing on phenomena with stronger effect

3. Why do we care?

...and why do researchers cheat?

What is the problem?

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

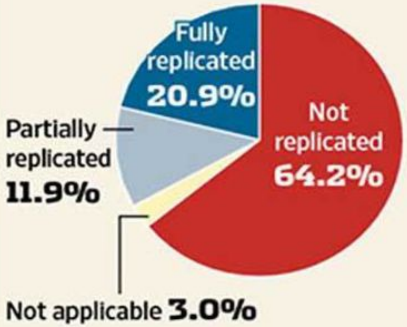
2005. PLoS Medicine, 2(8), e124. doi: 10.1371/journal.pmed.0020124

"There is increasing concern about the reliability of biomedical research, with recent articles suggesting that up to 85% of research funding is wasted."

Bustin, S. A. (2015). The reproducibility of biomedical research: Sleepers awake! *Biomolecular Detection and Quantification*

No Cure

When Bayer tried to replicate results of 67 studies published in academic journals, nearly two-thirds failed.



Source: Nature Reviews Drug Discovery

THE LANCET

Online First | Current Issue | All Issues | Special Issues | Multimedia | Information for Authors

All Content | Search | Advanced Search

Research: increasing value, reducing waste

Published: January 8, 2014

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Arch

News & Comment > News > 2015 > May > Article

NATURE | NEWS

First results from psychology's largest reproducibility test

Replicability crisis

Research misconduct and questionable research practices *are contributing* to the reproducibility crisis

Economics (2015) 22 of 67 (33%)

Experimental economics (2016) 11 of 18 (61%)

Experimental philosophy (2018) 28 of 40 (70%)

Microarray gene expression analysis (2009) 8 of 18 (44%)

Oncology & cardiovascular medicine (2011) 14 of 67 (20%)

RP: Cancer Biology (mixed results) 11%-25%

Neuroscience ~6%

Psychology (2015) ~36%

https://osf.io/ykfcq/ & https://tinyurl.com/hkujamoviworkshop

		Data	
		Same	Different
Code	Same	Reproducible	Replicable
	Different	Robust	Generalisable

https://figshare.com/articles/Publishing_a_reproducible_paper/5440621/1
See also https://stm.sciencemag.org/content/8/341/341ps12

nature

Explore our content ⌄ Journal information ⌄ Subscribe

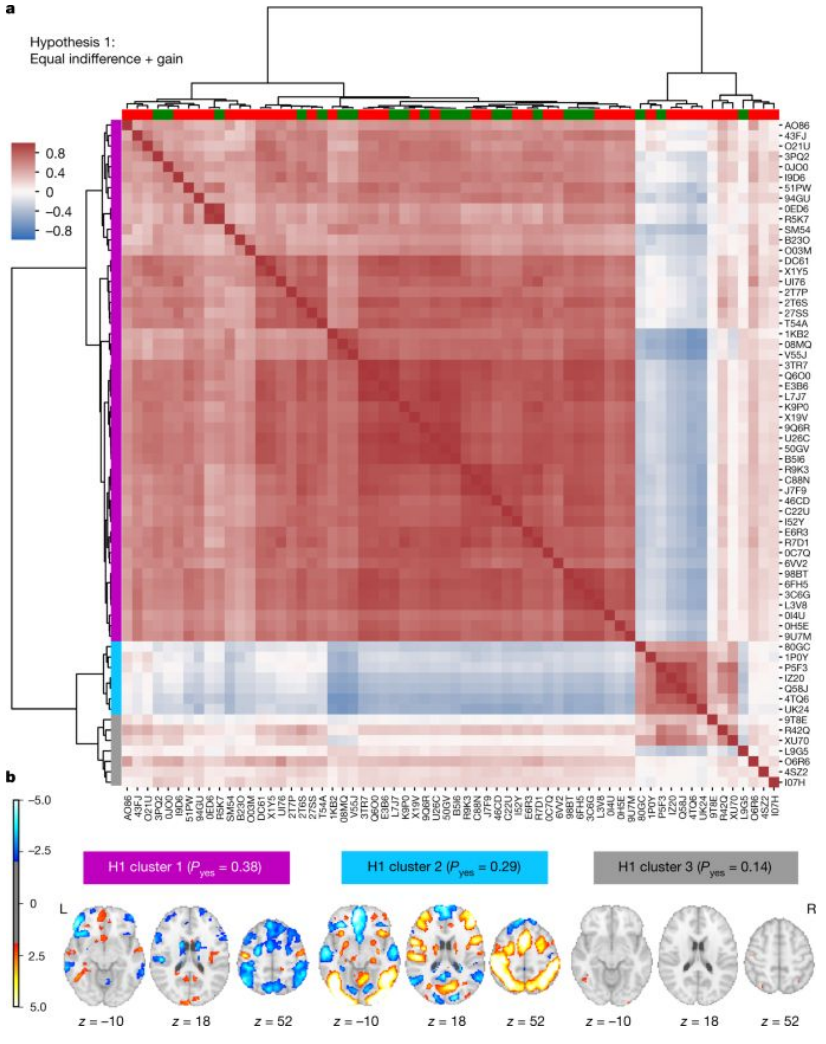
nature › articles › article

Article | Published: 20 May 2020

Variability in the analysis of a single neuroimaging dataset by many teams

Rotem Botvinik-Nezer, Felix Holzmeister, Colin F. Camerer, Anna Dreber, Juergen Huber, Magnus Johannesson, Michael Kirchler, Roni Iwanir, Jeanette A. Mumford, R. Alison Adcock, Paolo Avesani, Blazej M. Baczkowski, Aahana Bajracharya, Leah Bakst, Sheryl Ball, Marco Barilari, Nadège Bault, Derek Beaton, Julia Beitner, Roland G. Benoit, Ruud M. W. J. Berkers, Jamil P. Bhanji, Bharat B. Biswal, Sebastian Bobadilla-Suarez, Tiago Bortolini, Katherine L. Bottenhorn, Alexander Bowring, Senne Braem, Hayley R. Brooks, Emily G. Brudner, Cristian B. Calderon, Julia A. Camilleri, Jaime J. Castrellon, Luca Cecchetti, Edna C. Cieslik, Zachary J. Cole, Olivier Collignon, Robert W. Cox, William A. Cunningham, Stefan Czoschke, Kamalaker Dadi, Charles P. Davis, Alberto De Luca, Mauricio R. Delgado, Lysia Demetriou, Jeffrey B. Dennison, Xin Di, Erin W. Dickie, Ekaterina Dobryakova, Claire L. Donnat, Juergen Dukart, Niall W. Duncan, Joke Durnez, Amr Eed, Simon B. Eickhoff, Andrew Erhart, Laura Fontanesi, G. Matthew Fricke, Shiguang Fu, Adriana Galván, Remi Gau, Sarah Genon, Tristan Glatard, Enrico Glerean, Jelle J. Goeman, Sergej A. E. Golowin, Carlos González-García, Krzysztof J. Gorgolewski, Cheryl L.

- **70 teams** analysed the **same neuroimaging dataset**
- **Not a single identical analysis**
- **Consensus** showed **robustness** of some of the results



https://www.nature.com/articles/s41586-020-2314-9

Bonus materials to learn about reproducibility/replicability

- **Aalto series on RDM**
  - https://www.aalto.fi/en/services/rdm-training
  - https://youtu.be/FdMEeUqhJNw
  - https://hackmd.io/@AaltoSciComp/howToMakeYourCodeReusable
- **The Turing Way book https://the-turing-way.netlify.app/welcome**
- **CodeRefinery https://coderefinery.org/workshops/**

  (we give workshops to learn git, jupyter, reproducibility, replicability, …)

3.b ...but why do researchers engage with unethical research practices?

Researchers cut corners because of "the incentives"

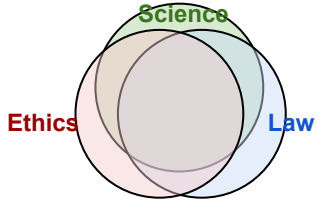
- A sensational story is more important than honest results to get it published on Nature or Science
- QRP are often justified in the "publish or perish" culture
- Null results are still not welcomed by peer review

Incentives must not justify misconduct. Nothing justifies misconduct. https://www.talyarkoni.org/blog/2018/10/02/no-its-not-the-incentives-its-you/

https://osf.io/ykfcq/ & https://tinyurl.com/hkujamoviworkshop

4. How can we fix things?

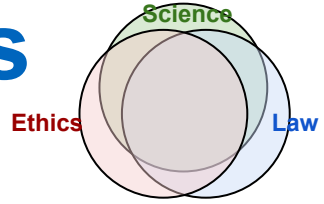
Transparency in science



Transparency is the principle that should lead us towards the sweet spot of ethical, lawful, reproducible science

- **Transparency towards data subjects**
  - **What** data are collected; **why** they should consent on data reuse; what are the **risks** of re-identification vs **benefits** for society
- **Transparency towards other scientists**
  - Sharing **data**, **processes**, **code**, **results**
- **Transparency towards authorities**
  - **GDPR** is here to help us and **WE** decide the **best practices!**

Transparency in science needs to be rewarded (My opinions)



Transparency is the principle that should lead us towards the sweet spot of ethical, lawful, reproducible science.

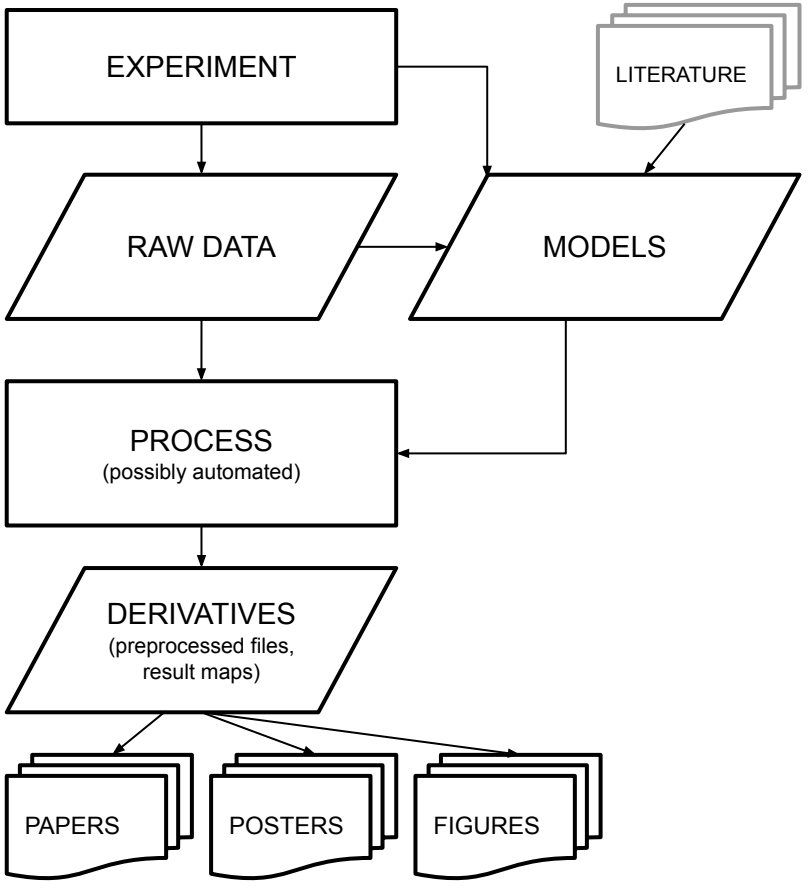
Openness and transparency in research should be the most important metric that research and funding institutions should use to evaluate the work of researchers.

However transparency comes with high costs, it should not become a new source of burnout. A strong peer community is necessary, always start from little steps.

4.b Achieve transparency of all bits related to a study (if there’s time)

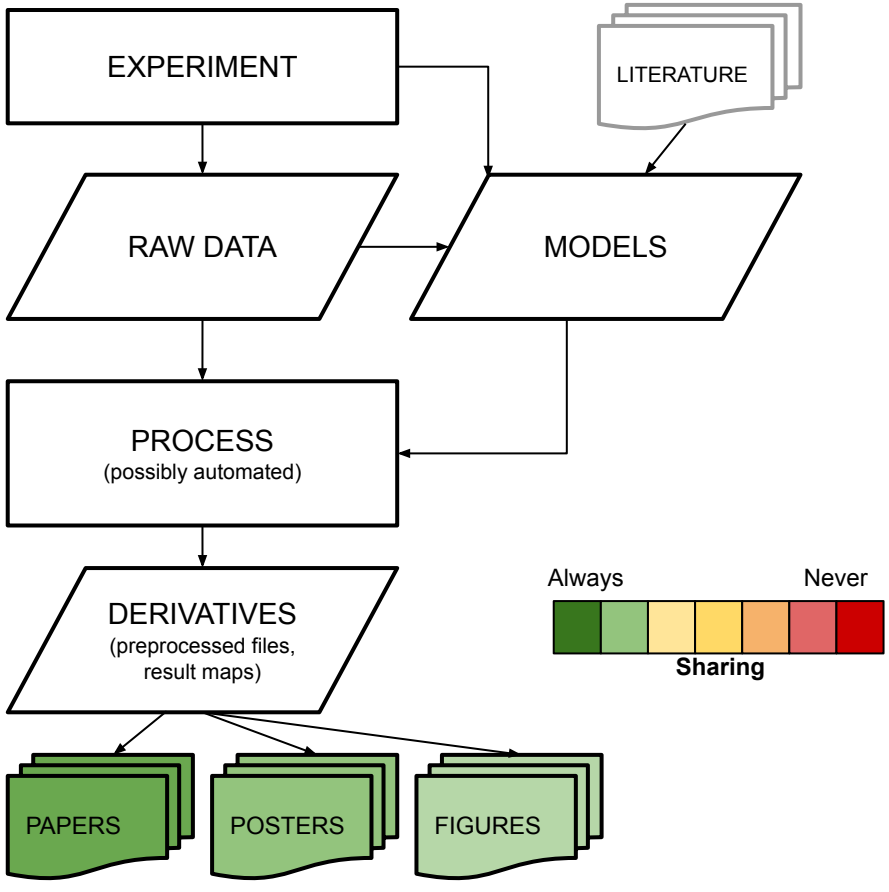
The (brain) experiment pipeline

- **The most simple and generic pipeline of an experimental work**
- **From very rich data formats** (M/EEG, fMRI, behaviour) **to documents containing 2D colourful pictures, tables and text**



Sharing is fundamental in science

- **Some bits are always shared** (research output)

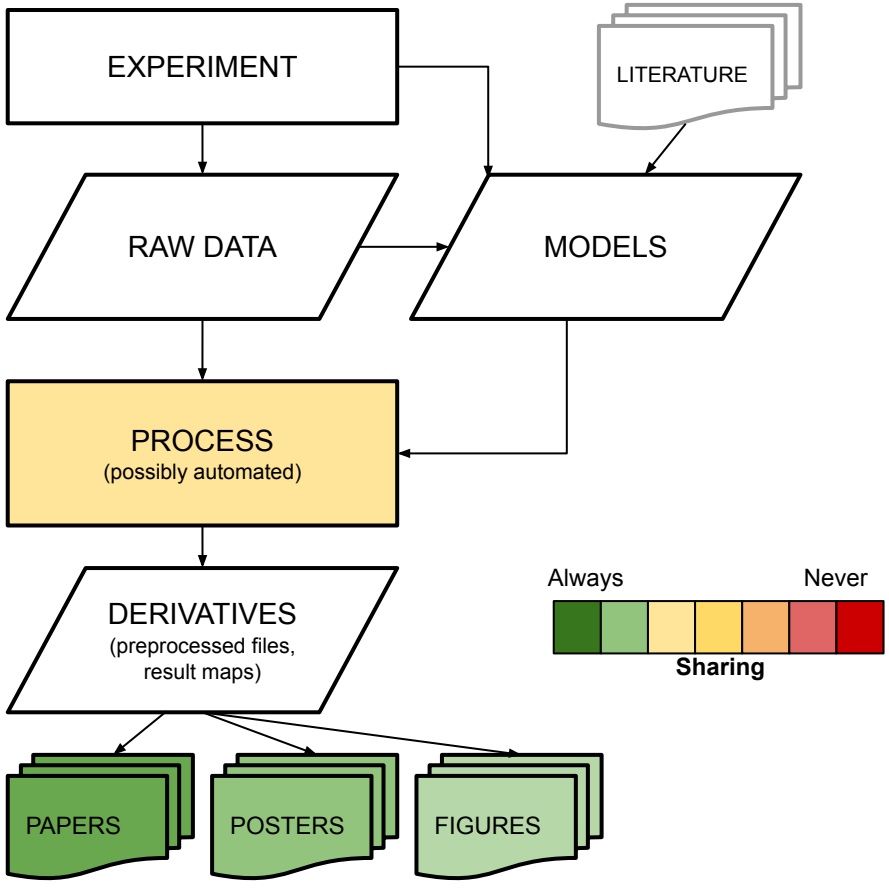


Should we just trust other scientists for all the other bits?

Younger me: yes!
Current me: no!

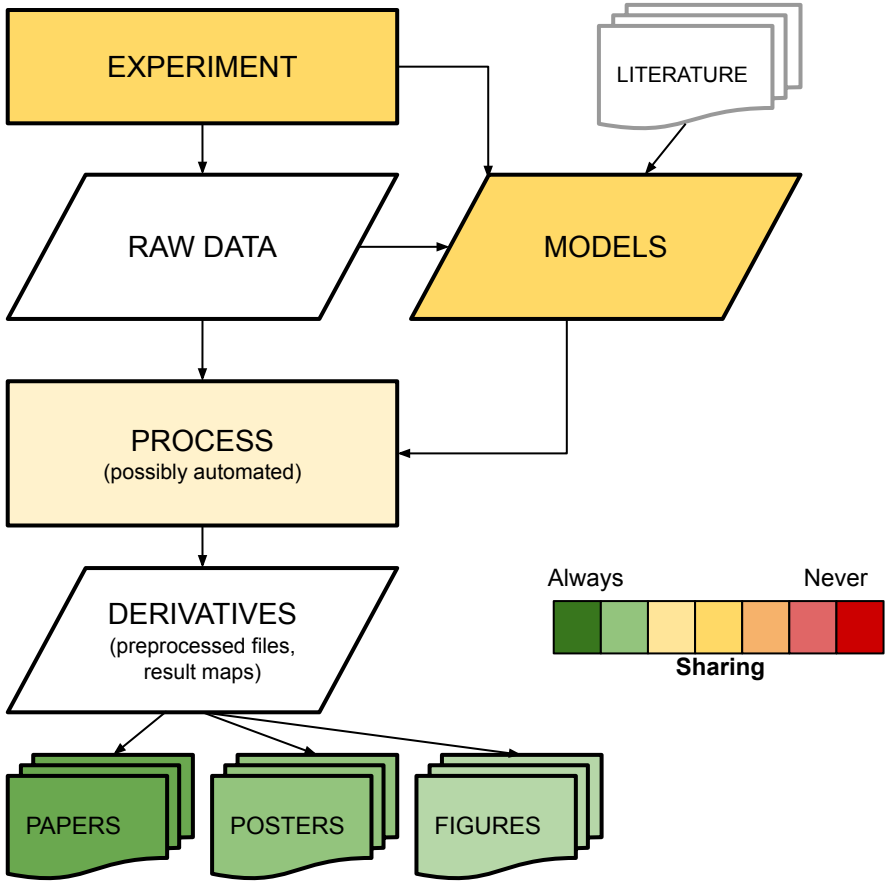
Sharing is fundamental in science

- **Some bits are always shared** (research output)
- Sometimes **code/methods** are shared (reusing methods)



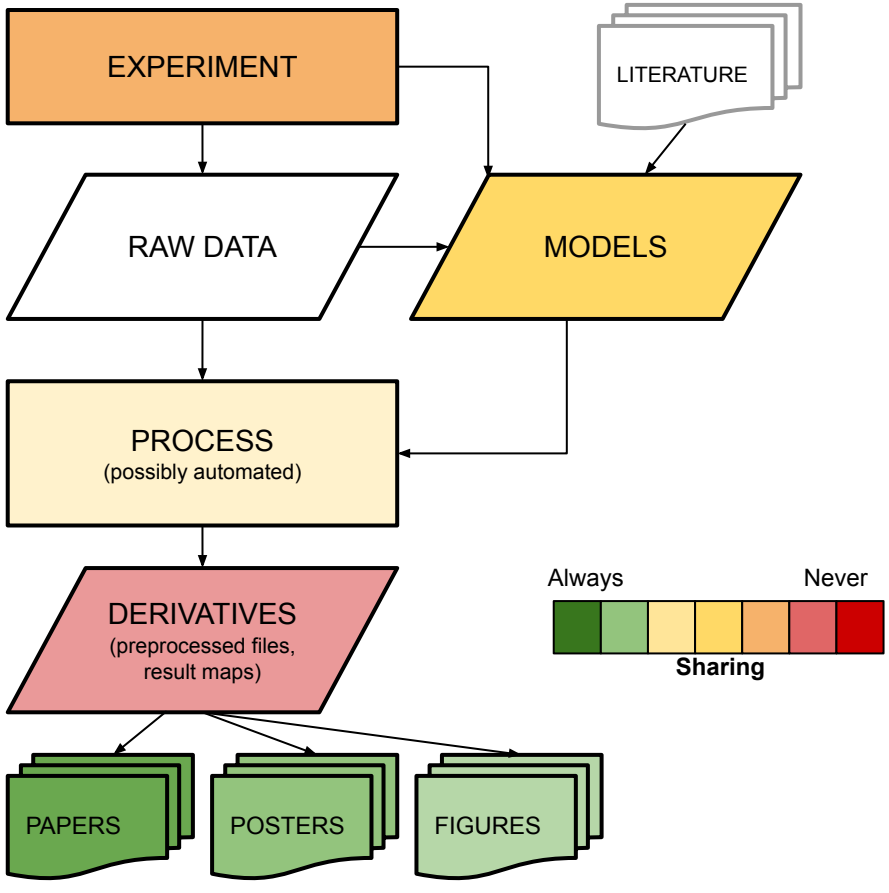
Sharing is fundamental in science

- **Some bits are always shared** (research output)
- Sometimes **code/methods** are shared (reusing methods)
- **Stimulus and models** are less frequently shared (rerunning experiments)



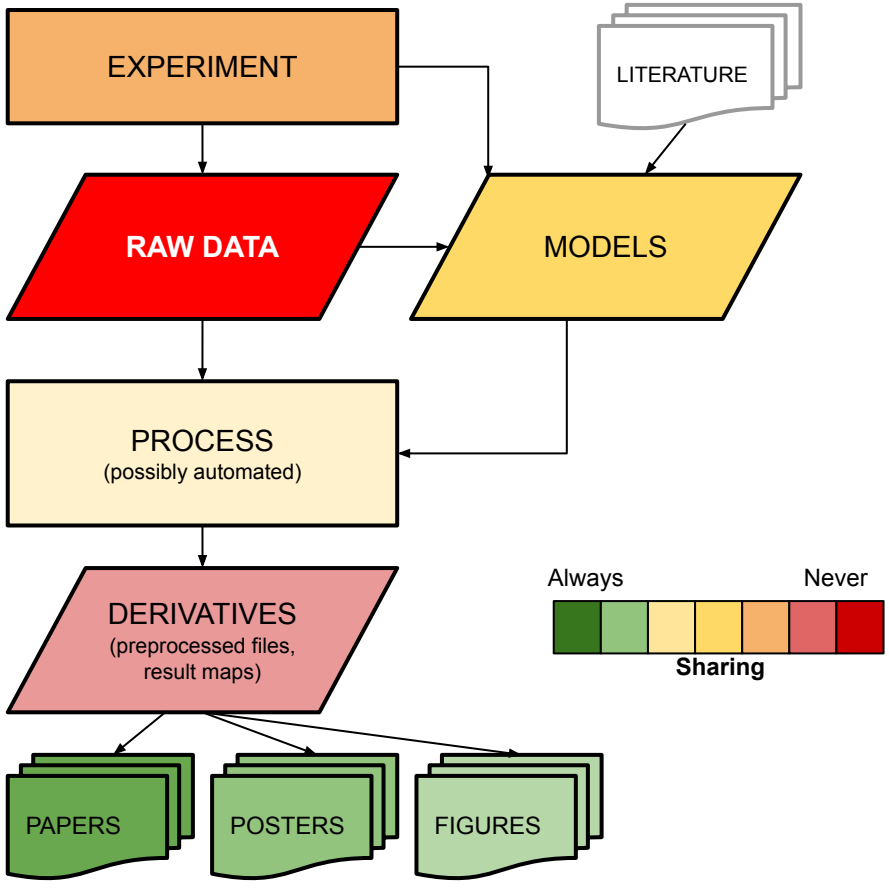
Sharing is fundamental in science

- **Some bits are always shared** (research output)
- Sometimes **code/methods** are shared
- **Stimulus and models** are less frequently shared (rerunning experiments)
- **Very rarely derivatives files are shared** (meta-analysis!)



Sharing is fundamental in science

- **Some bits are always shared** (research output)
- Sometimes **code/methods** are shared
- **Stimulus and models** are less frequently shared (rerunning experiments)
- **Very rarely derivatives files are shared** (meta-analysis!)
- **Raw brain data never shared** in Finland (full re-analysis, novel analyses)



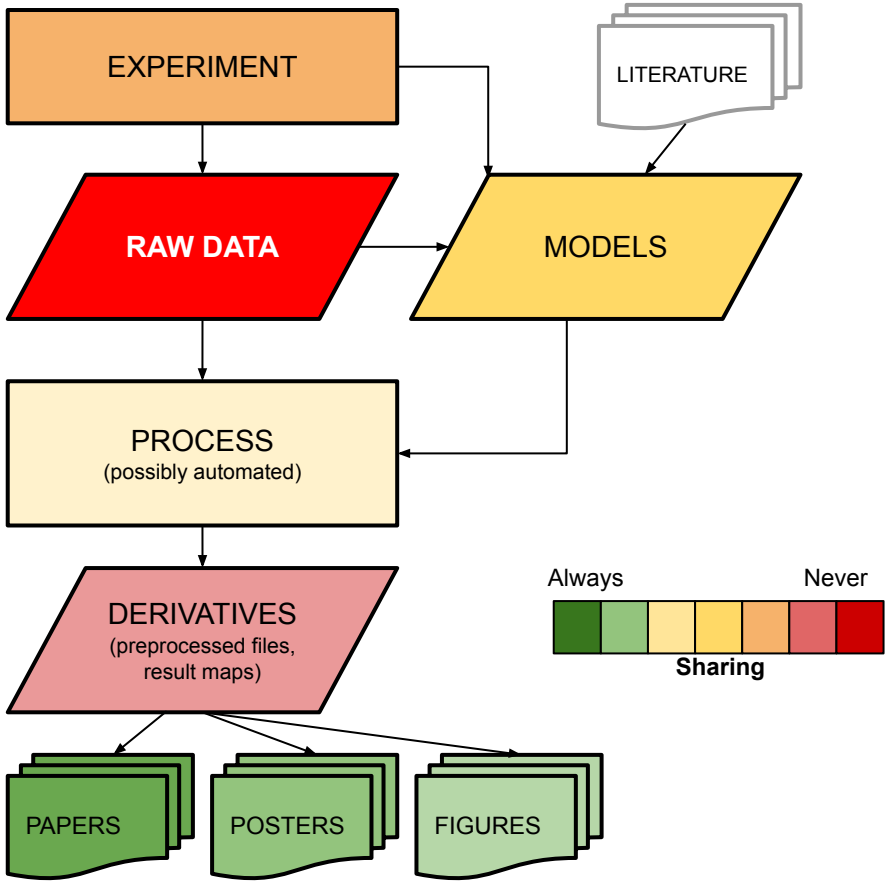
Sharing everything should be at the basis of the scientific process

Why sharing is not happening?

Understanding the causes

Why sharing is not happening?

- **Lack of incentives**
  ...actually sharing increases citations
- **Lack of requirements from journals/agencies/universities**
  ...actually this is changing
- **Lack of tools for sharing**
  ...actually there are places for sharing each part of the process
- **Lack of resources (time&money)**
  ...actually Aalto is happy to help your team to share each part of the process and streamline the process
- **Lack of training** … e.g. licensing of code
- **Ethical concerns** ...and that's why we are here
- **Fear** from impostor syndrome to fear of being "scooped"

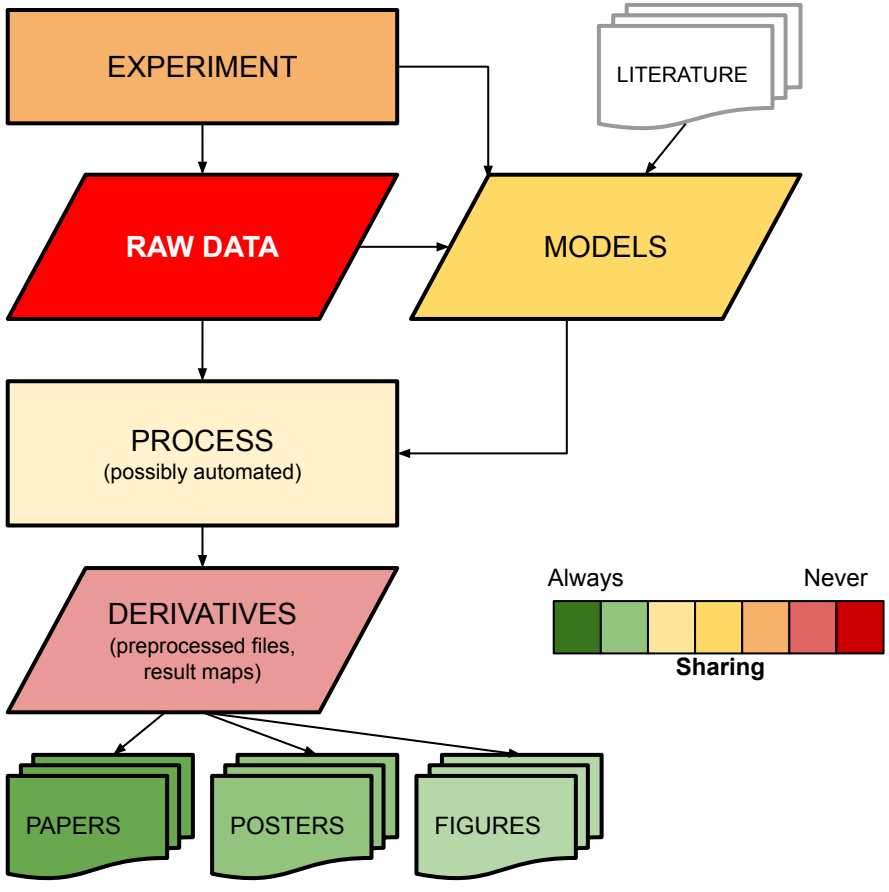


How can we share?

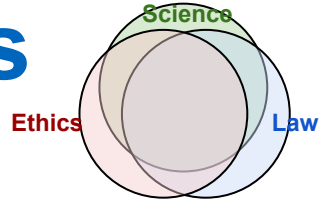
Know your tools and share all the parts

How to share and get benefits from it

- **Papers/figures/posters**
  Scientific journals, preprint servers (arXiv, biorxiv), storage services that provide a DOI (zenodo, figshare)
- **Code and process**
  GitHub and similar + zenodo for github DOI integration
- **Experiment/models**
  Zenodo, figshare, eudat
- **Derivatives**
  Zenodo, figshare, eudat
- **Raw data** sometimes you can't share raw data because of privacy, keep them safe and aim at finding an open dataset to replicate your findings



Transparency in science needs to be rewarded



Transparency is the principle that should lead us towards the sweet spot of ethical, lawful, reproducible science.

Openness and transparency in research should be the most important metric that research and funding institutions should use to evaluate the work of researchers.

Bonus

Resources for qualitative studies

Reproducibility and qualitative research methods

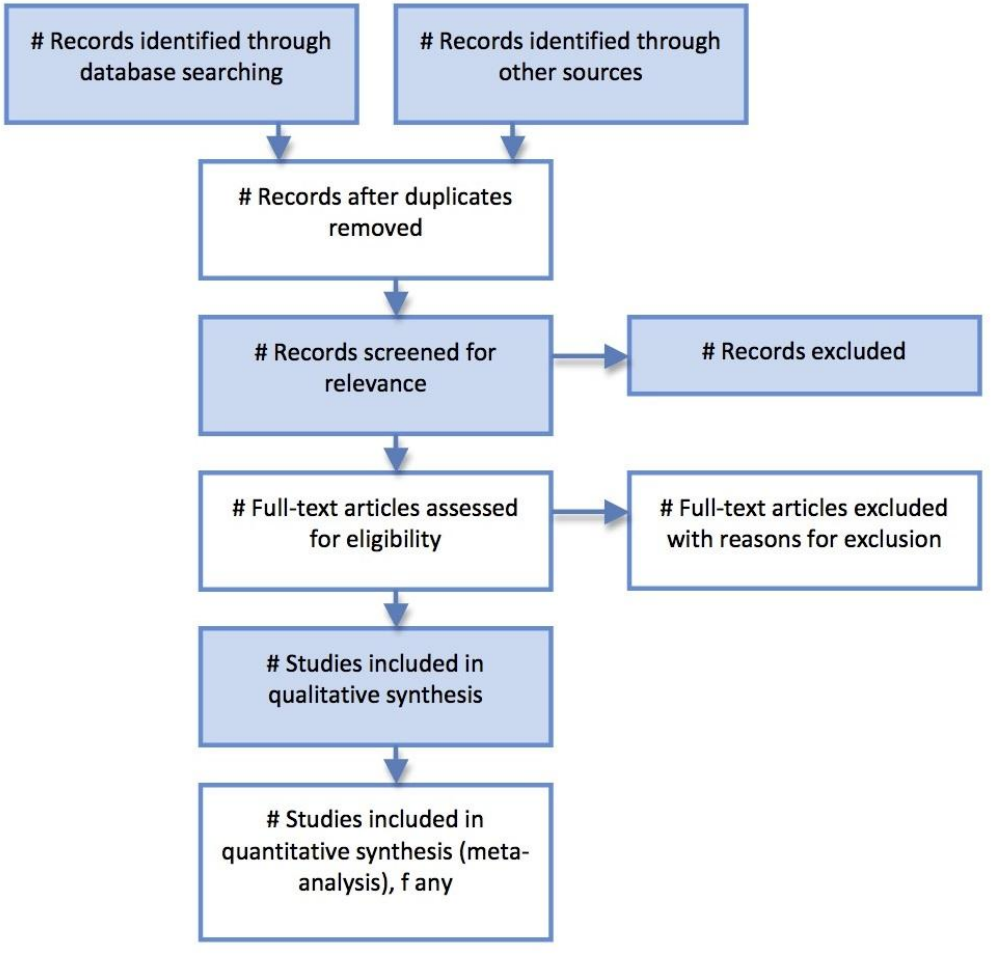
While some ideas like analysis pre-registration do not fit well with qualitative research methods, scientists should still share the full data collection protocol and analysis choices to ensure that their work is reproducible.

https://openworking.wordpress.com/2019/02/11/what-does-reproducibility-mean-for-qualitative-research/

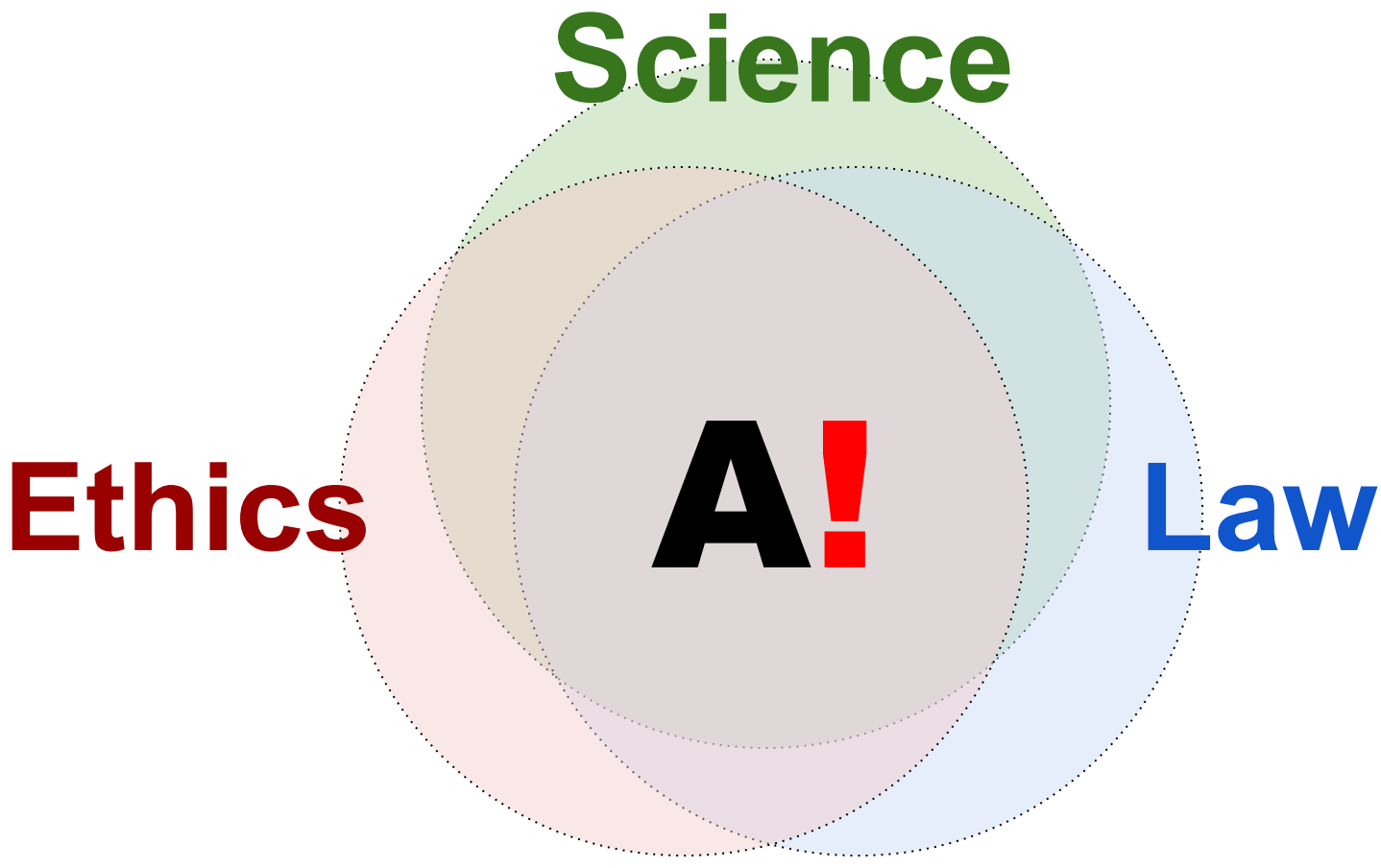
Bonus II

Make literature research reproducible

Systematic literature reviews (e.g. PRISMA guidelines)



https://en.wikipedia.org/wiki/Preferred_Reporting_Items_for_Systematic_Reviews_and_Meta-Analyses


Science
Ethics
A!
Law