

Responsible conduct of research

Enrico Glerean, DSc, @eglerean, www.glerean.com



Outline of today's workshop

1. What is responsible conduct of research (RCR)?
2. What is research misconduct?
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...

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/>

When in doubt, ask! researchdata@aalto.fi or your dept. lawyer or data agent

...and also (not covered today)

Research ethics at Aalto

<https://www.aalto.fi/en/research-art/research-ethics-and-research-integrity>

<https://mycourses.aalto.fi/course/view.php?id=23138> (needs Aalto login)

TENK guidelines to work with human subjects:

https://www.tenk.fi/sites/tenk.fi/files/lhmistieteiden_eettisen_ennakkoarviointin_ohje_2019.pdf

How to handle personal data in research

<https://www.aalto.fi/en/services/how-to-handle-personal-data-in-research>

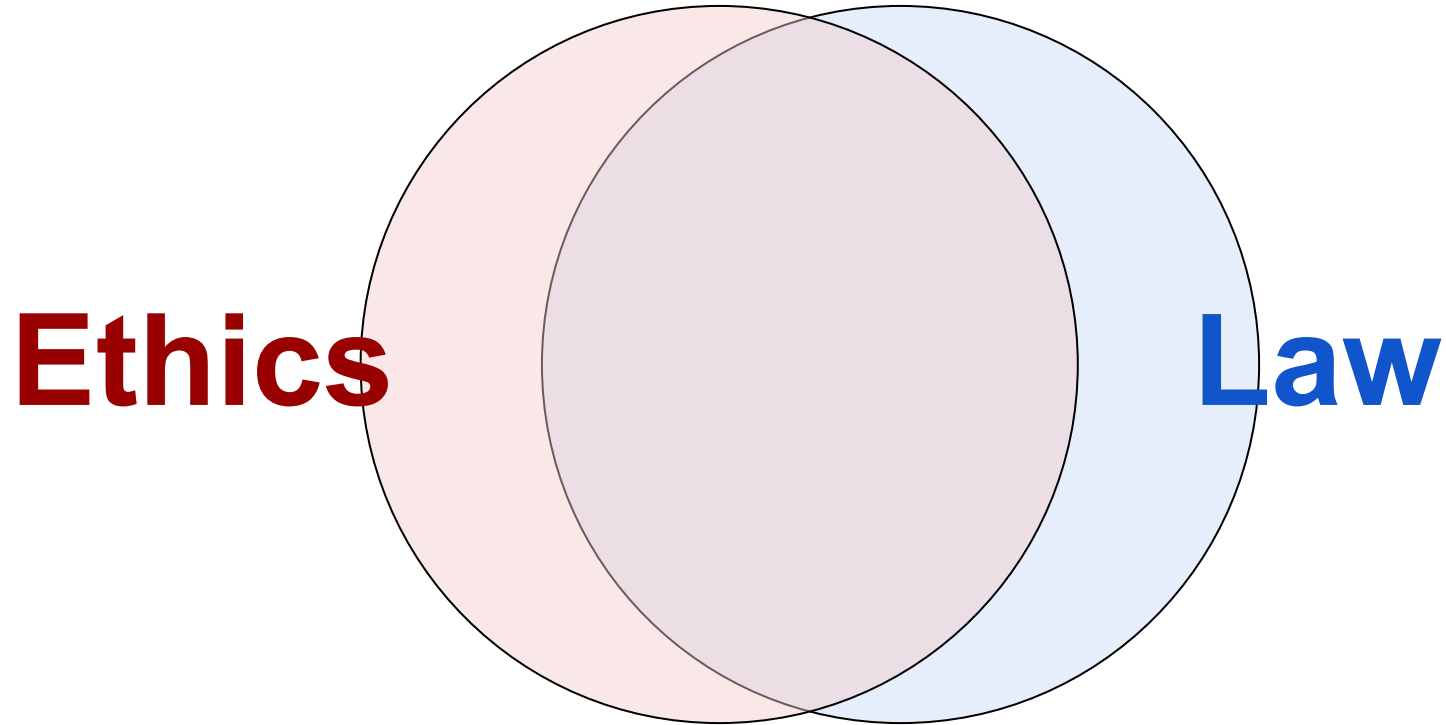
Again, when in doubt, ask! researchdata@aalto.fi or your dept. lawyer or data agent

1. What is responsible conduct of research?

Responsible Conduct of Research

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

Ethics is not Law

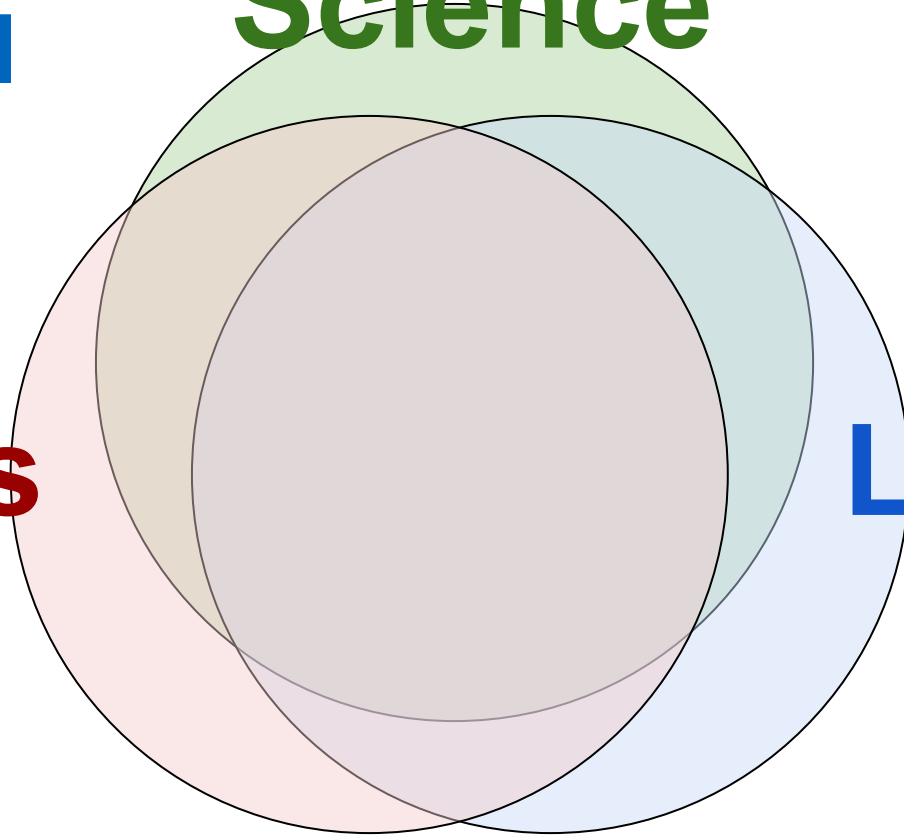


**Ethics,
Law, and
Science**

Science

Ethics

Law

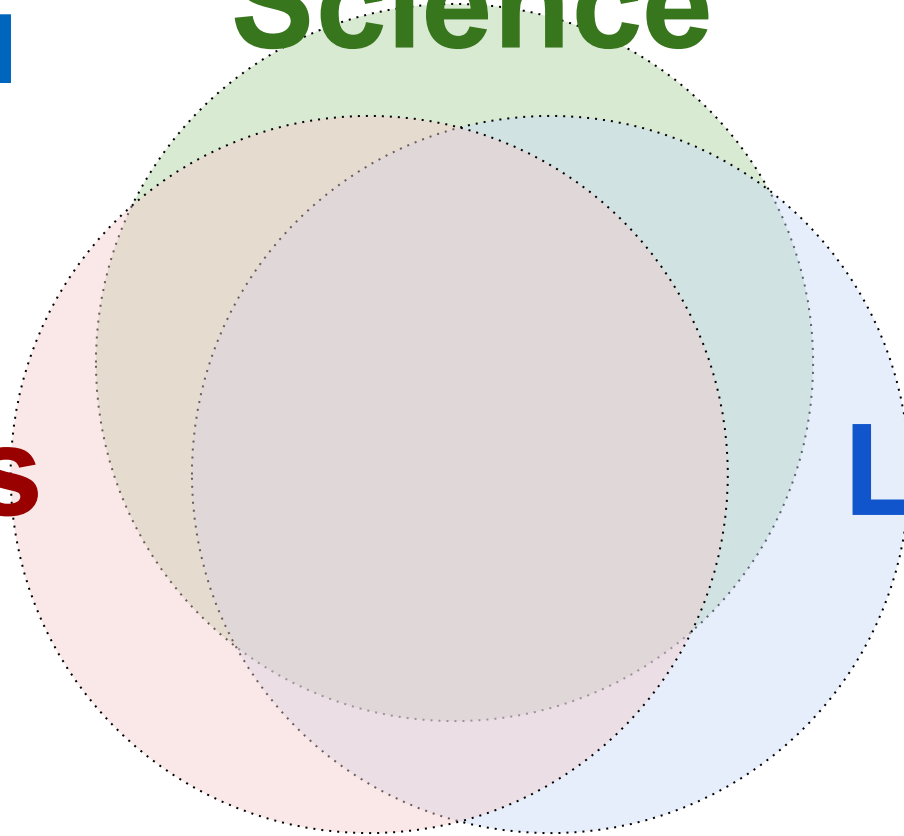


**Ethics,
Law, and
Science**

Science

Ethics

Law



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, not acknowledging work done by others)

Research misconduct as disregarding RCR

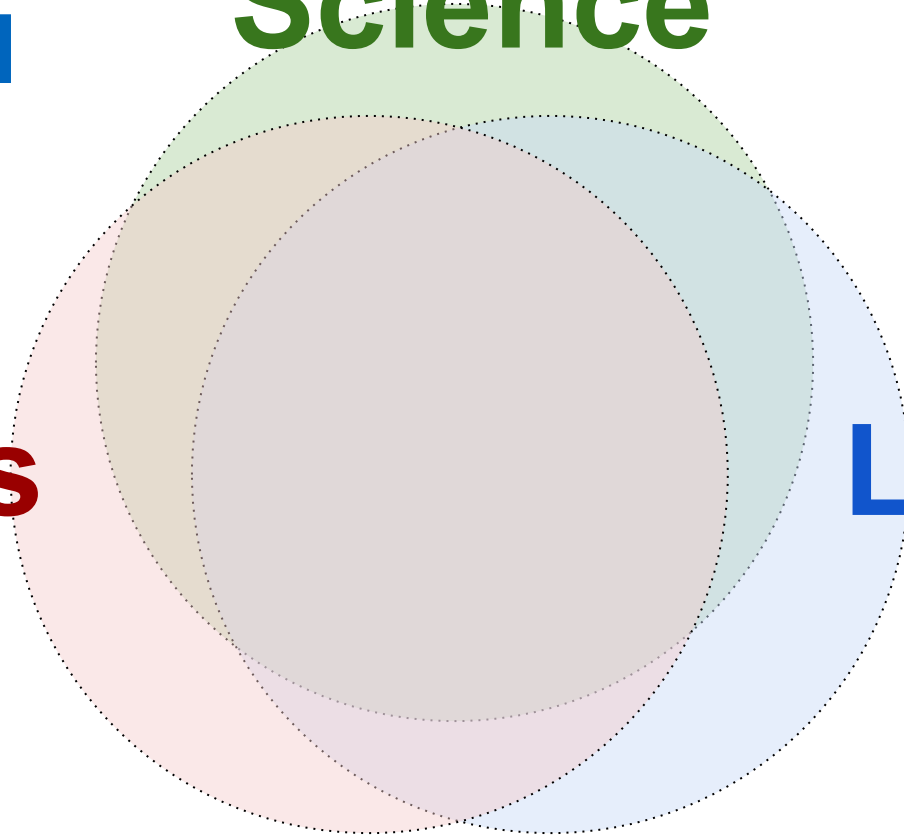
- **Harming others' works for profit (ignoring literature, unethical peer review, manipulating citation metrics, conflicts of interest)**
- **Misleading the general public (false findings excessively advertised)**
- **Questionable research practices (p-hacking, harking, publication bias, analysis bias)**

**Ethics,
Law, and
Science**

Science

Ethics

Law



Ethics, Law, and Science

Science

Scientific misconduct
FFPM

Questionable
research
practices
(p-Hacking)

Ethics

Sharing your
findings breaking
journal's
copyright

The spot we all
are aiming for

Law

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_gub_it_its/

E.g. the GRIM test:

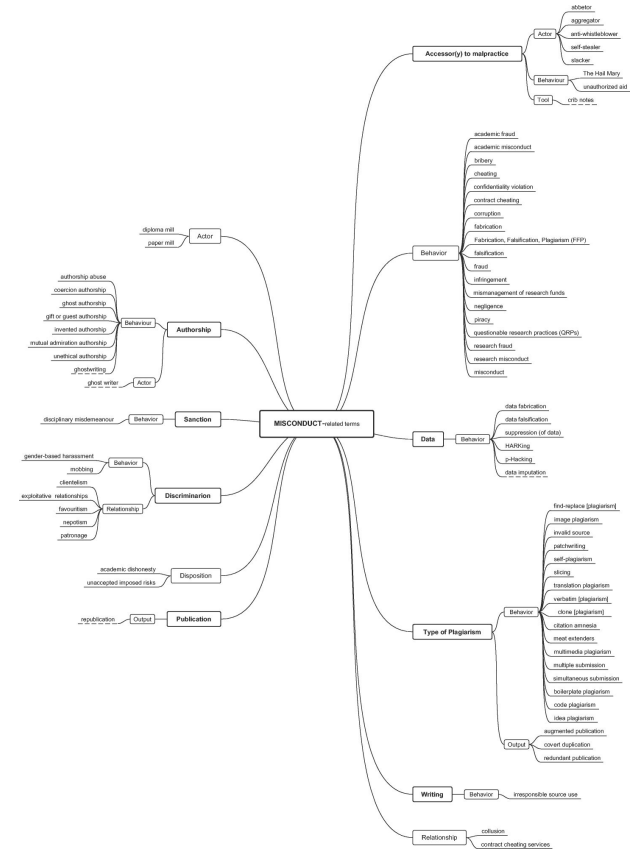
<https://medium.com/@jamesheathers/the-grim-test-a-method-for-evaluating-published-research-9a4e5f05e870>

More about Brown and Heathers:

<https://www.sciencemag.org/news/2018/02/meet-data-thugs-out-expose-shoddy-and-questionable-research>

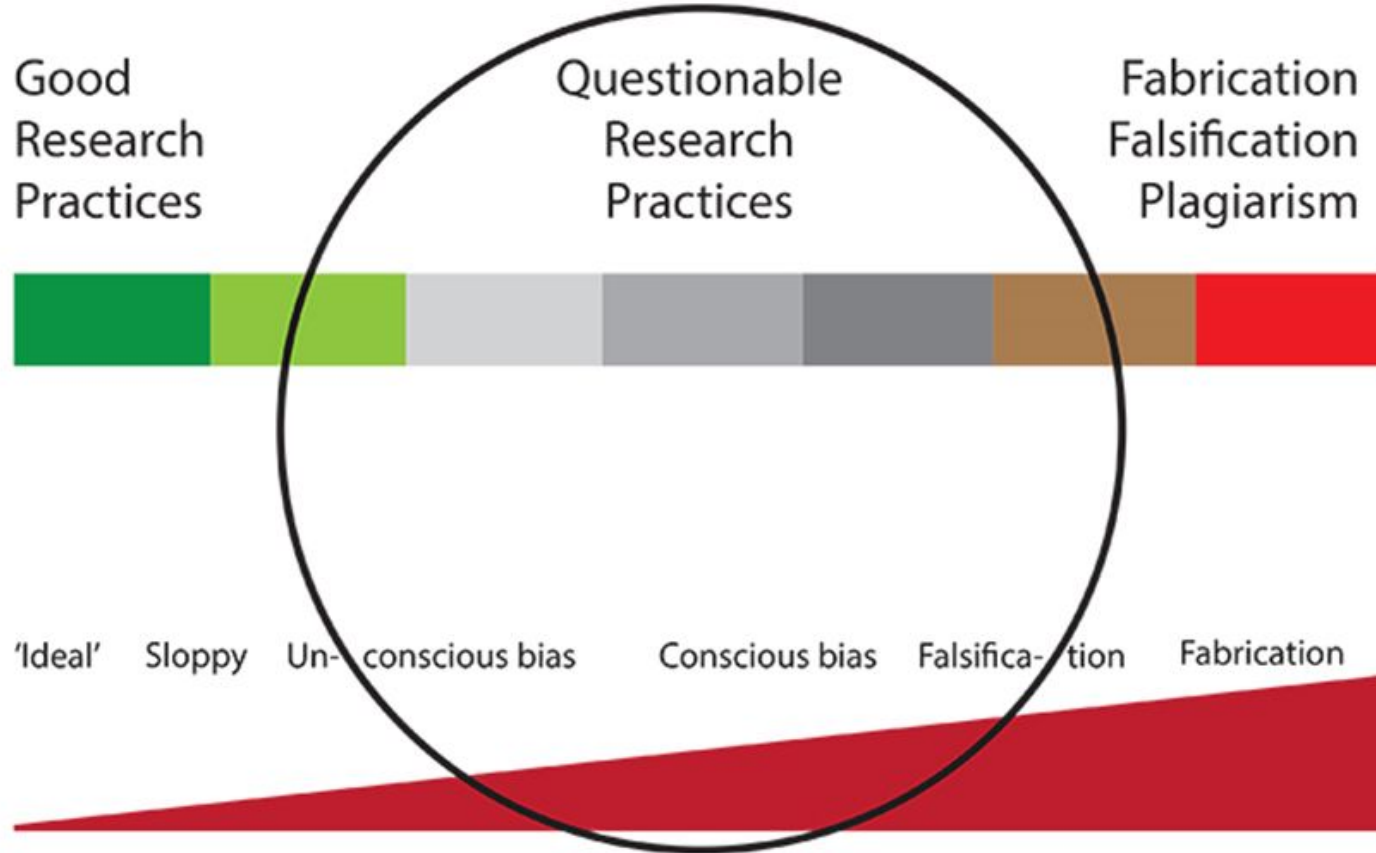
Taxonomy of the types of misconduct

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



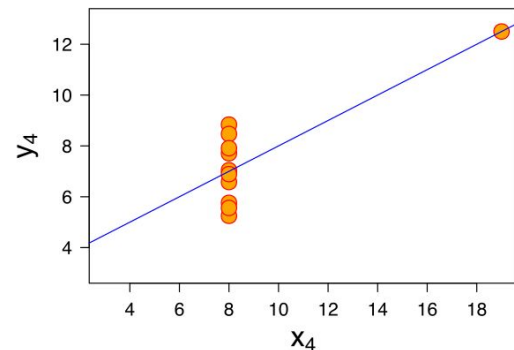
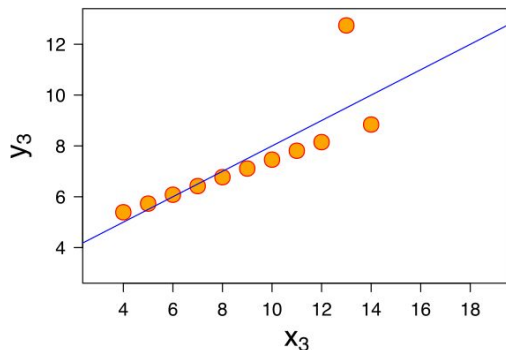
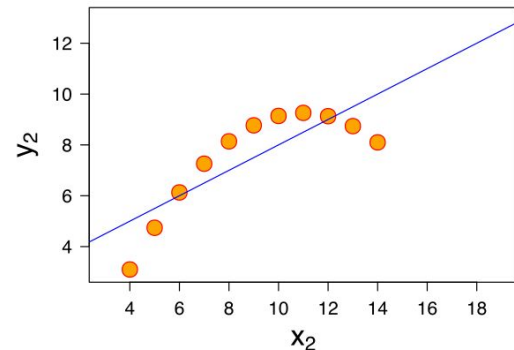
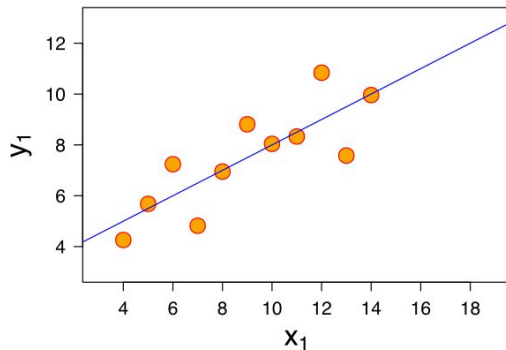
2.b What should be considered research misconduct in 2019?

From integrity to misconduct



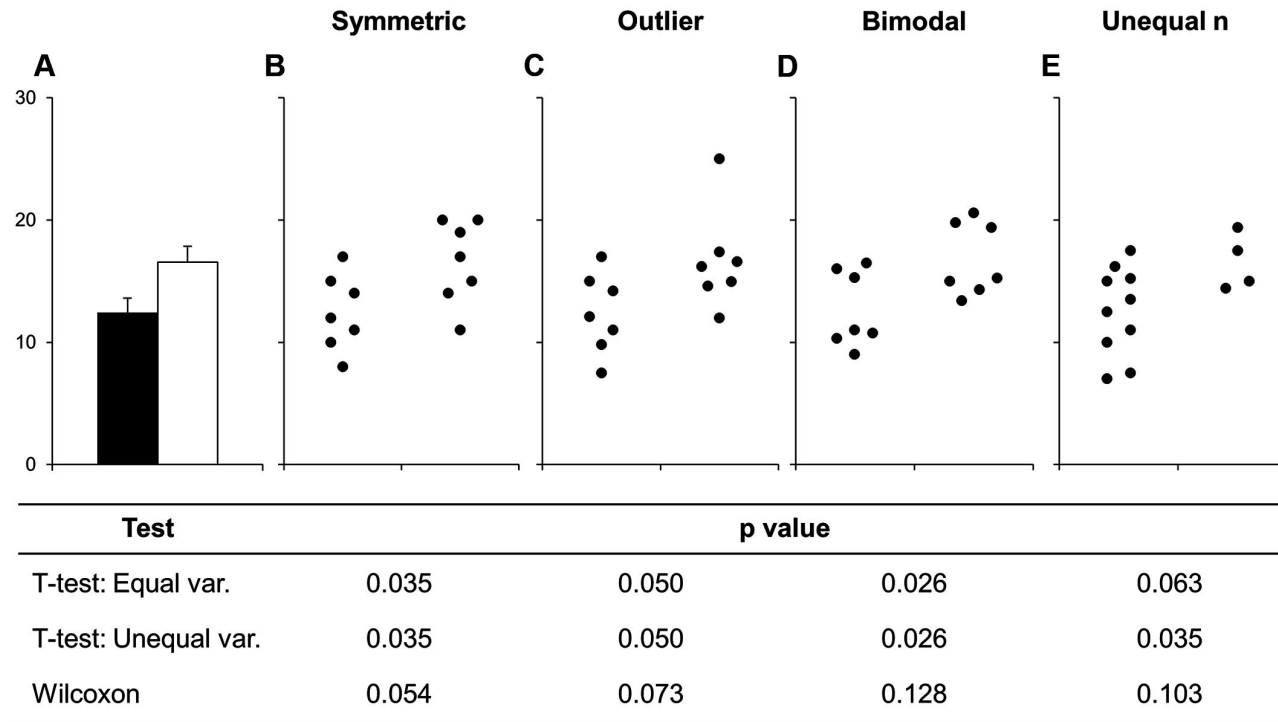
Honest mistakes

- Reporting $r = 0.816$ while...



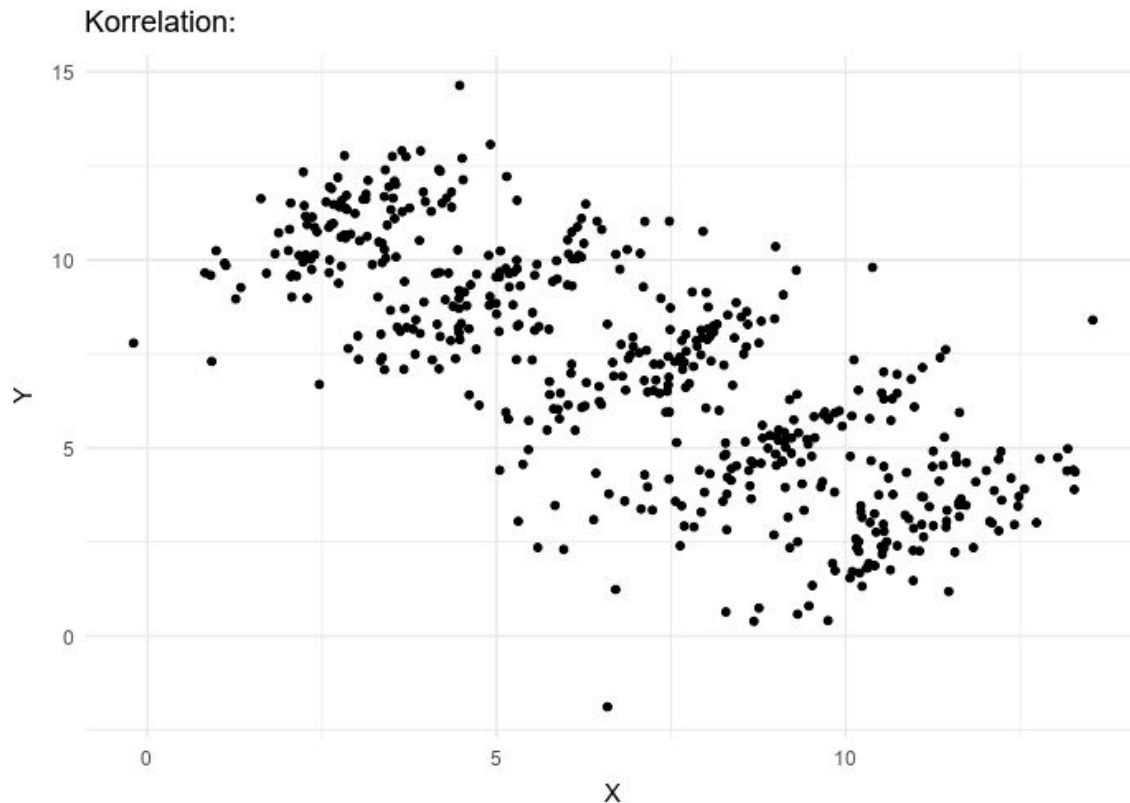
Honest mistakes

- Reporting a significant group difference while...



Honest mistakes

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



Problems caused by researchers 1

- Unconscious bias / Confirmatory bias / Seeing patterns that are not there / HARKing
- ... which leads to irreproducible findings, publication bias or to the file drawer effect
- Solution: eradicate the subjectivity from methods, automate data collection and analysis, make data and methods open, pre-register hypothesis, replication dataset

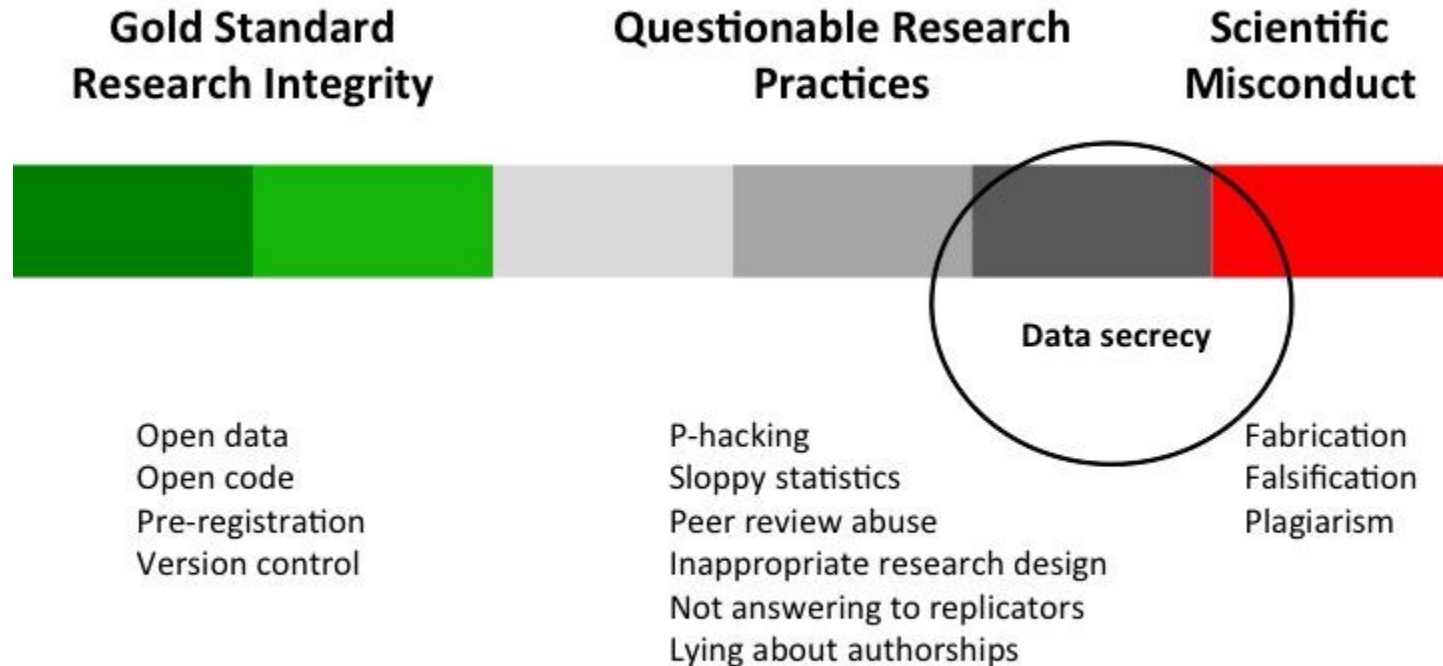


Problems caused by researchers 2

- Failure to understand statistics
- ... which often leads to p-hacking (trying multiple analysis and report only those that reached statistical significance): huge bias for false positives
- Solutions: simulated data, separate replication dataset, blind analysis with masked data, pre-registration of analysis, replication dataset

Problems caused by researchers 3

- **Data secrecy** (data privacy / confidentiality as an excuse)



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

THE LANCET

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

All Content

Search

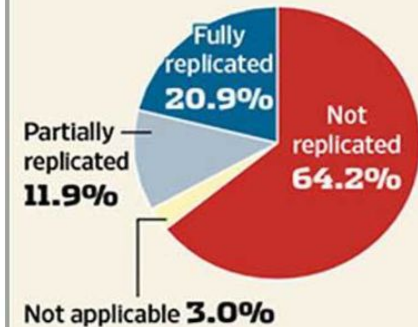
Advanced Search

Research: increasing value, reducing waste

Published: January 8, 2014

No Cure

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



Source: Nature Reviews Drug Discovery



NATURE | NEWS

First results from psychology's largest reproducibility test

Replicability crisis

Research misconduct and questionable research practices are at the basis of 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%

		Data	
		Same	Different
Code	Same	Reproducible	Replicable
	Different	Robust	Generalisable

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

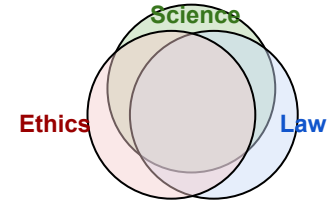
Researchers cut corners because of “the incentives”

- 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/>

4. How can we fix things?

Transparency in science

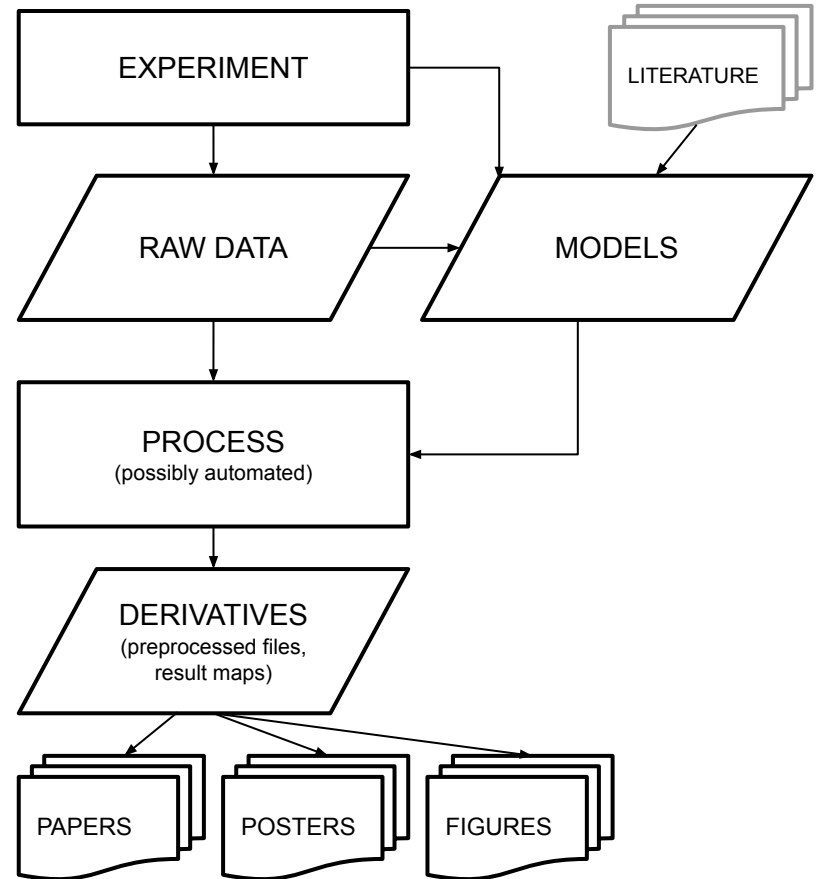


Transparency is the principle that should lead us towards the sweet spot of ethical, lawful, reproducible science

- **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!**

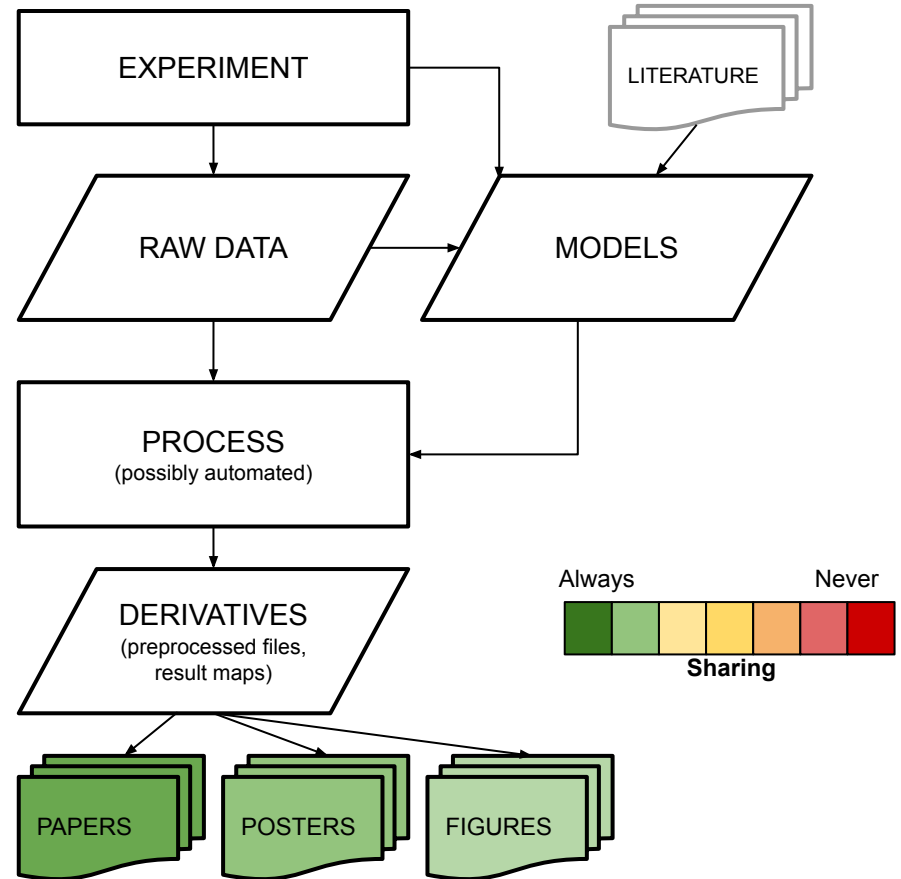
The (brain) experiment pipeline

- 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)

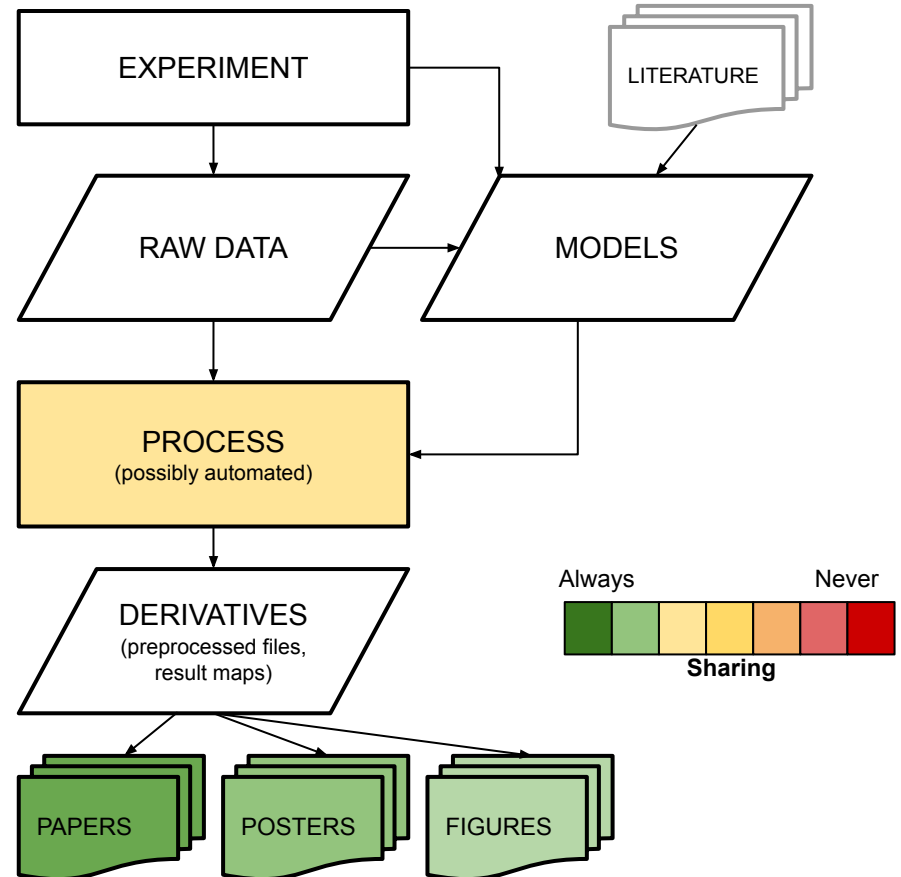


**Should we just trust
other scientists for all
the other bits?**

Younger me: yes!
Current me: no!

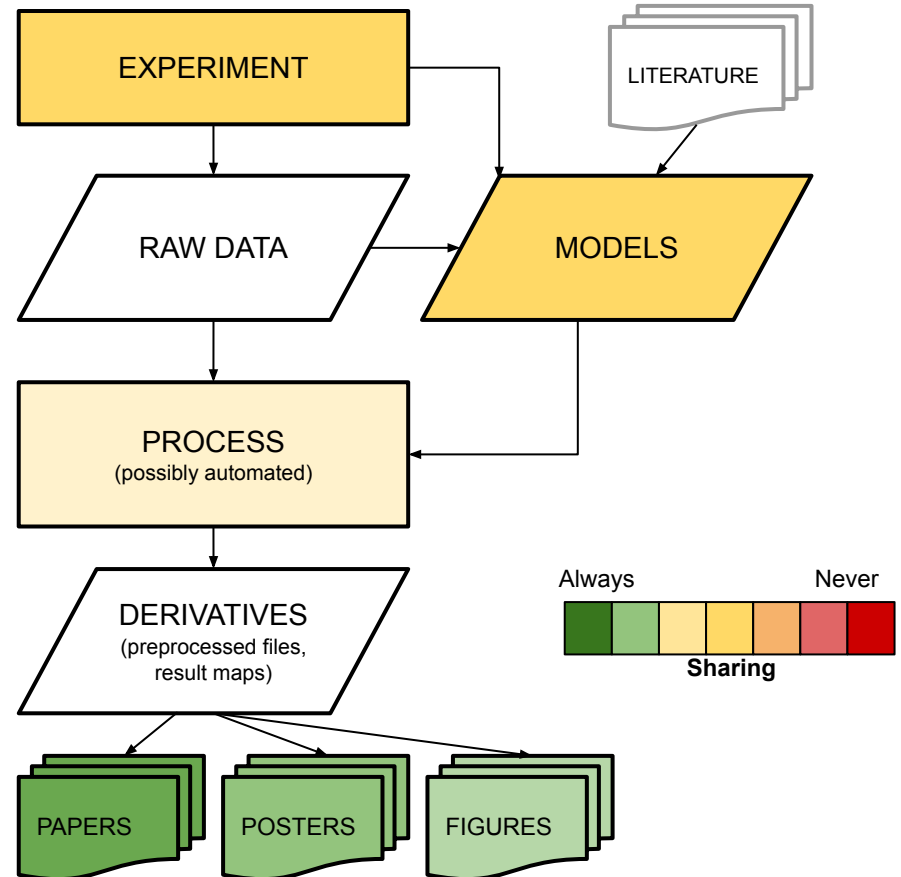
Sharing is fundamental in science

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



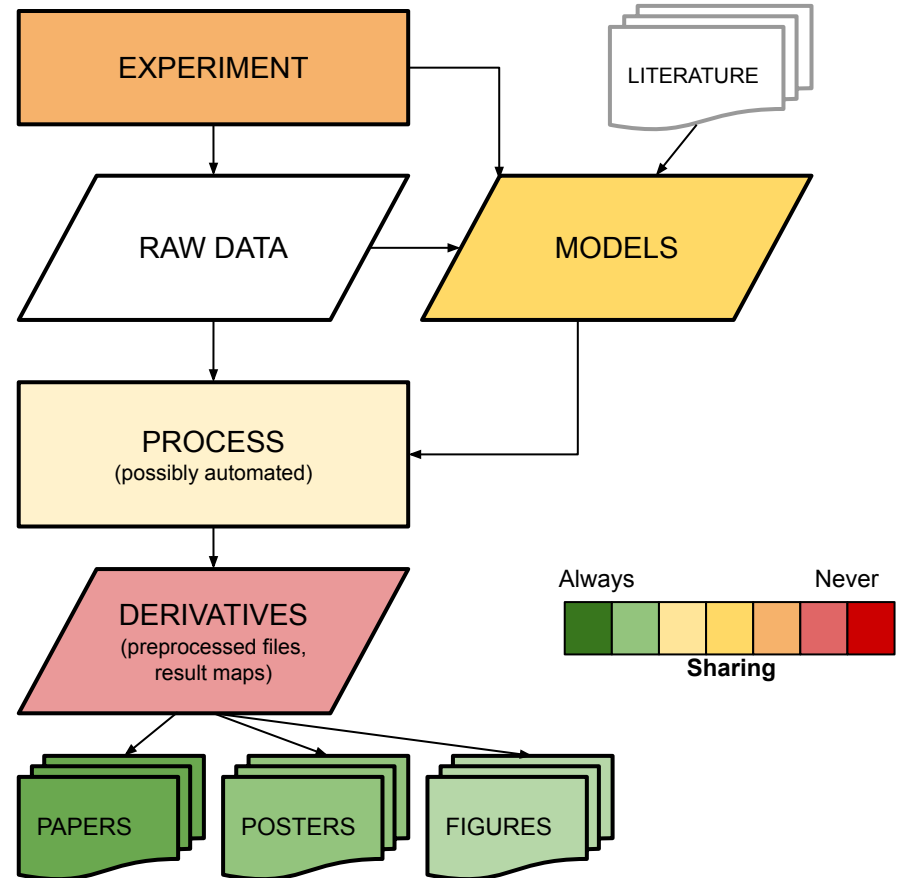
Sharing is fundamental in science

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



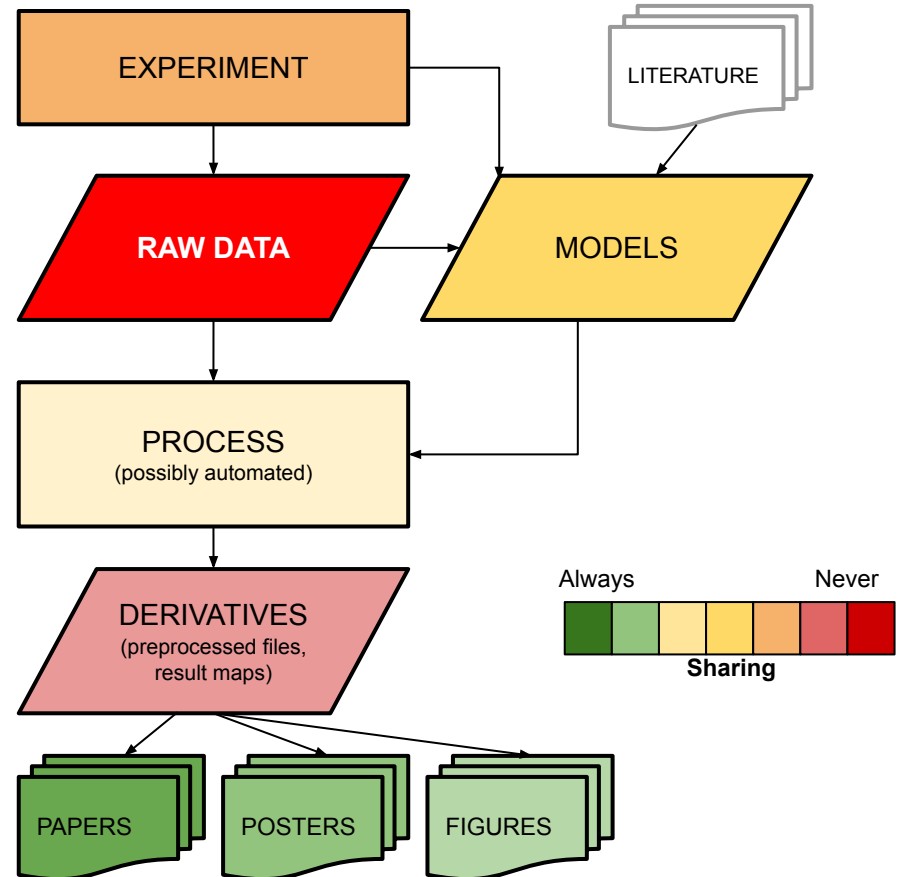
Sharing is fundamental in science

- **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)



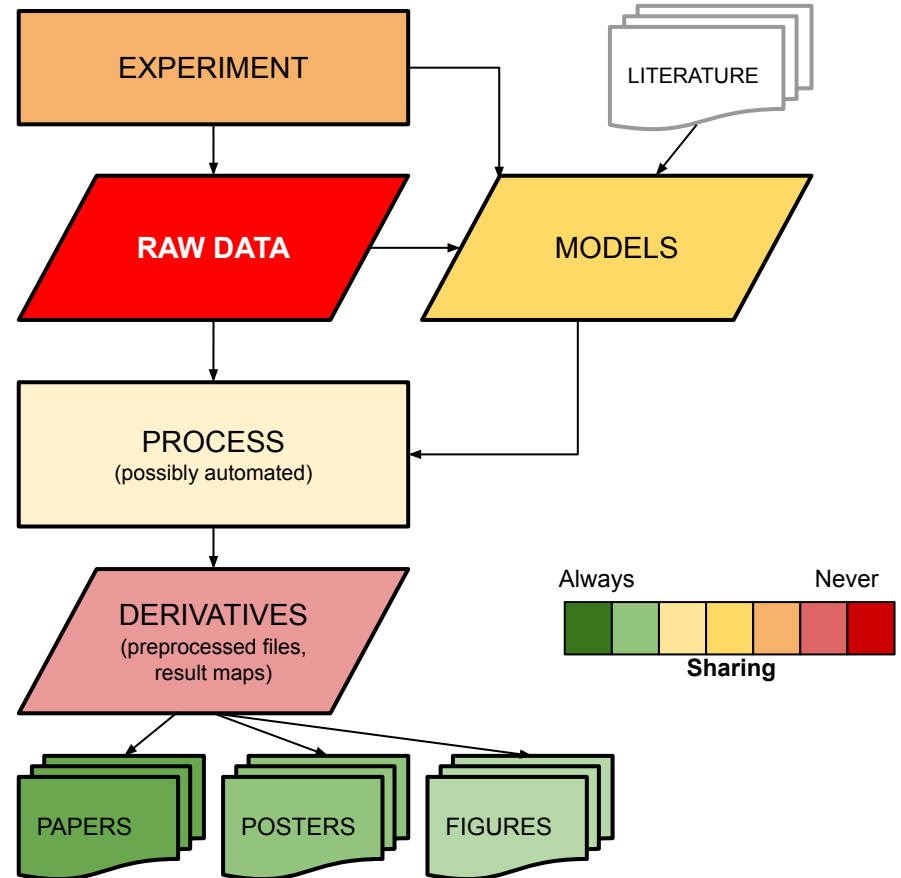
**Sharing everything
should be at the basis of
the scientific process**

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

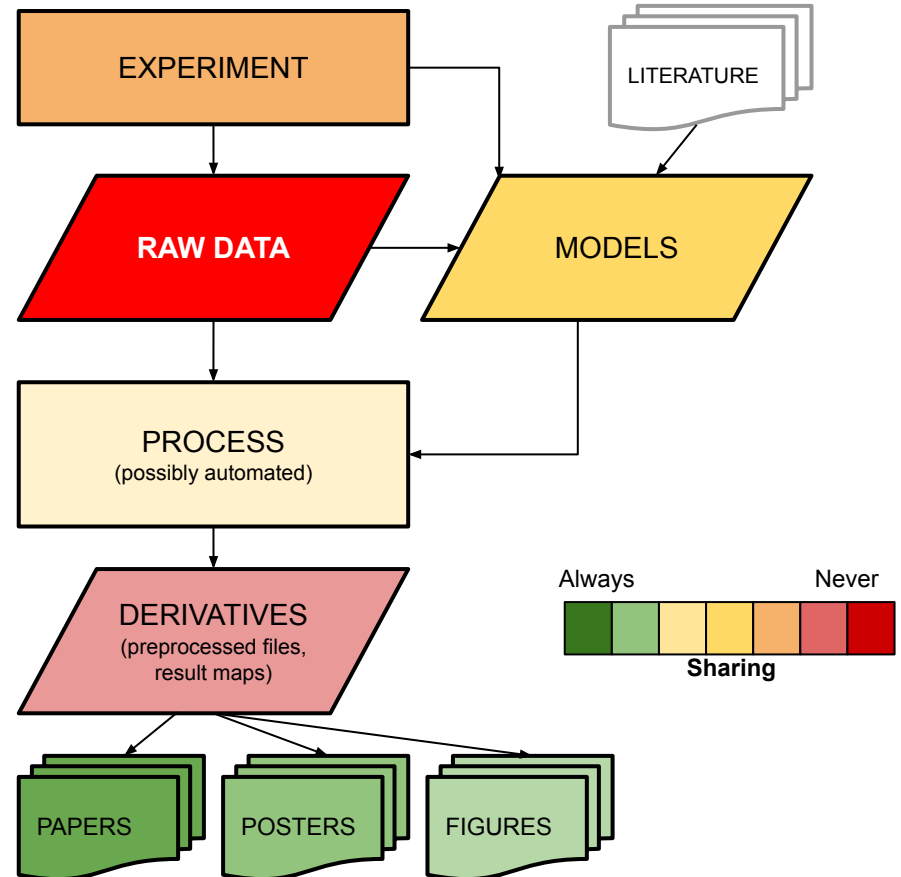
- **Experiment/models**

Zenodo, figshare, eudat

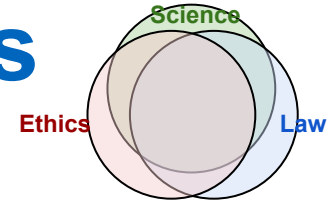
- **Derivatives**

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.

(NOTE, MY OPINION): 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

Make literature research reproducible

Systematic literature reviews (e.g. PRISMA guidelines)

