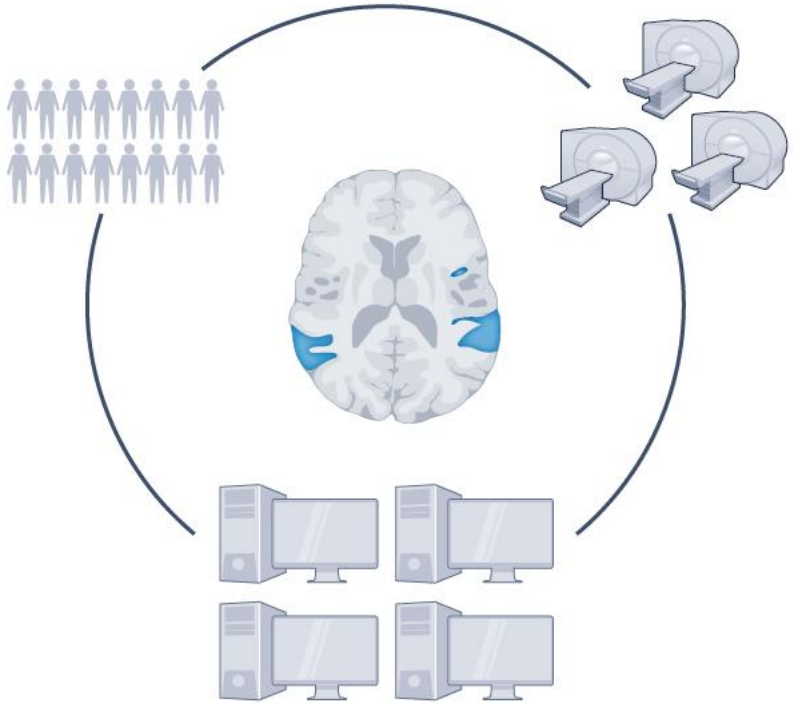
Reproducibility & replicability in MRI

Gaëlle Leroux

Reproducibility & replicability in MRI



Gaëlle Leroux, PhD

Research engineer

gaelle.leroux@cnrs.fr

https://orcid.org/0000-0003-2729-4945

Illustration from Vogt. Reproducibility in MRI. *Nat Methods* (2023)

Soutien Méthodologique aux projets d'imagerie

REMI Réseau d'Entraide Multicentrique en IRM

Ouvrir la science !

Research data College

Glossary

Generalisable
Replicable
Reproducible
Robust

		Data: Same	Data: Different
Code: Same		1	2
Code: Different		3	4

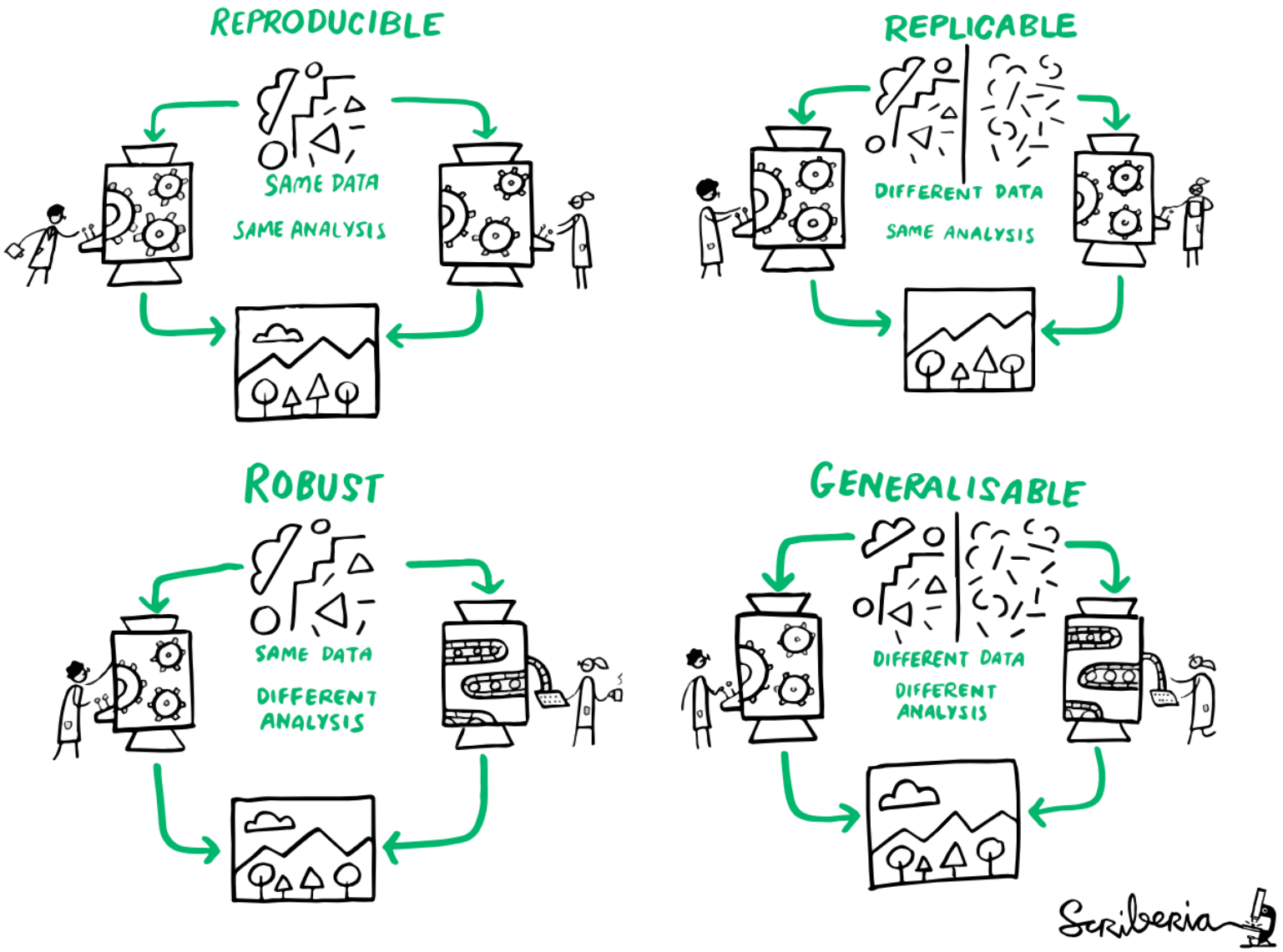
From Kirstie Whitaker 2017, doi: https://dx.doi.org/10.6084/m9.figshare.4244996

Glossary

		Data: Same	Data: Different
Code: Same		**Reproducible**	**Replicable**
Code: Different		**Robust**	**Generalisable**

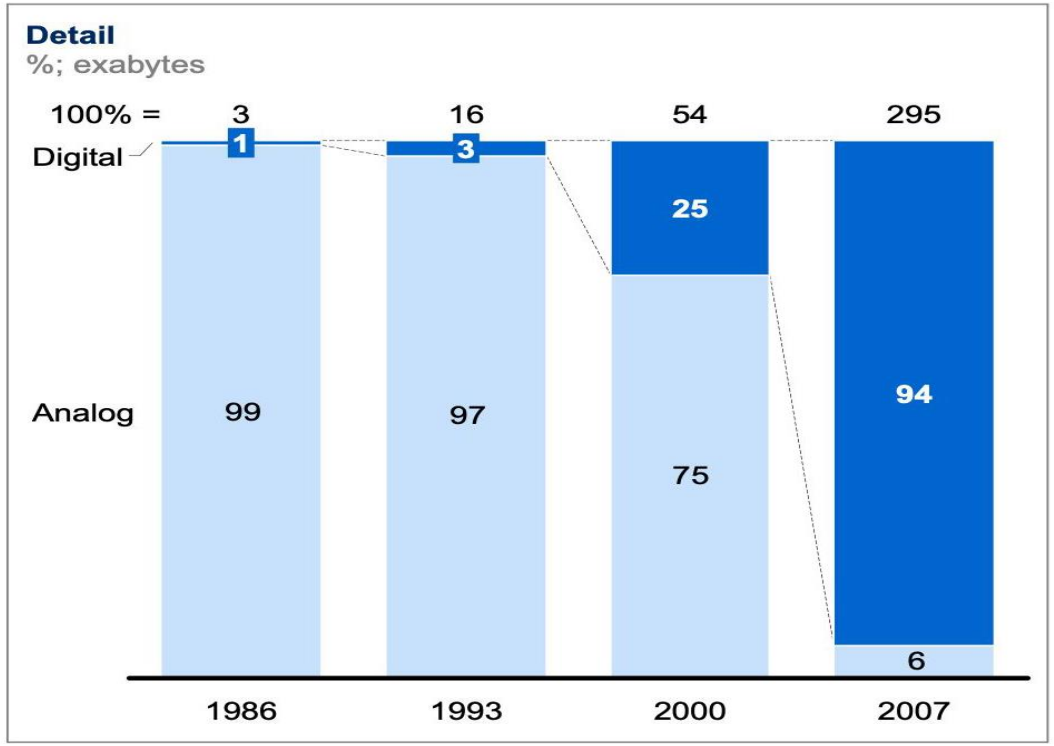
From Kirstie Whitaker 2017, doi: https://dx.doi.org/10.6084/m9.figshare.4244996

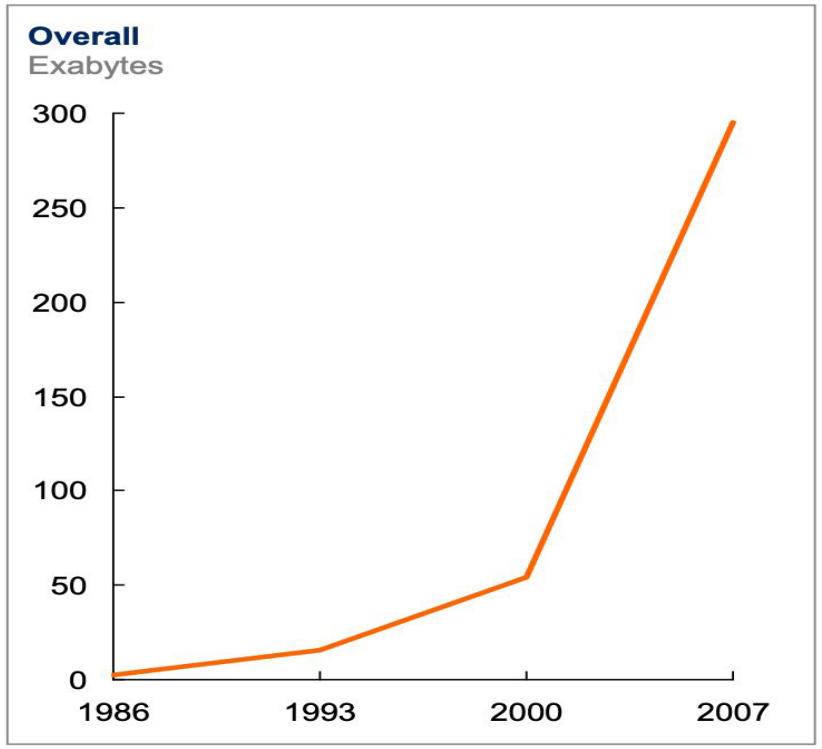
Glossary



The Turing Way project illustration by Scriberia. DOI: 10.5281/zenodo.3332807

1. An overview of the “reproducibility crisis”: origins

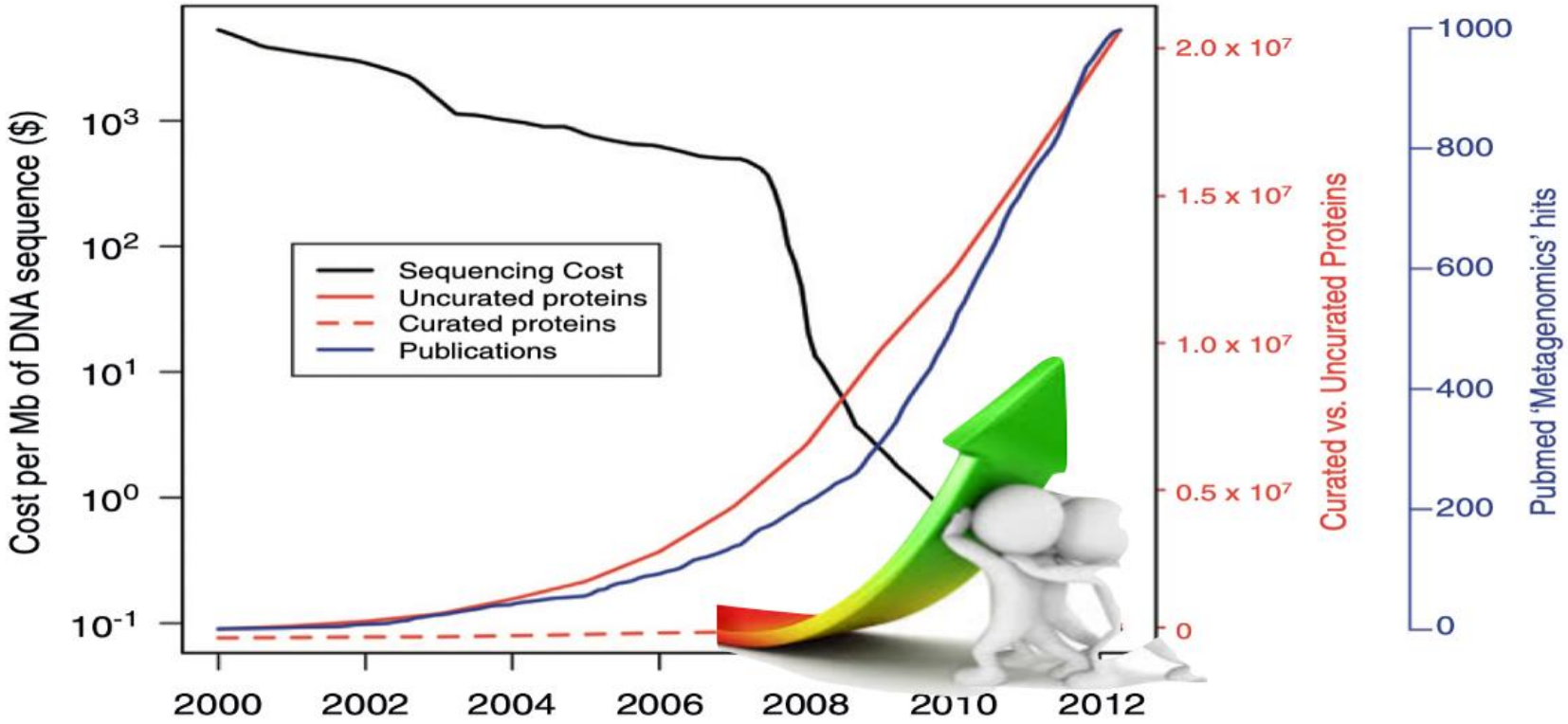




Source

1. An overview of the “reproducibility crisis”: origins

Human genome:
in 1990 = 13 years & $3 billion
in 2015 = a few hours & $1,000



➔ **The amount of data to be stored & to be analysed, explodes**
➔ **Analysis performance dramatically falls**

Source

1. An overview of the “reproducibility crisis”: research tasks have moved

Before

1. Design the experiment
2. Collect data
3. Analyze results

A change of paradigm

Now

1. Massive data production / reuse
2. Organize (store, document, annotate)
3. Analyze (extract information)
4. Share information

1. An overview of the “reproducibility crisis”: alarm in 2005

74,946 Save	7,388 Citation
3,001,940 View	9,576 Share

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

should be interpreted based only on *p*-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. “Negative” research is also very useful.

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R+1)$. The probability of a study finding a true relationship reflects the power $1-\beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2×2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2

Inadequate methodologies

A result is all the more likely to be false when:

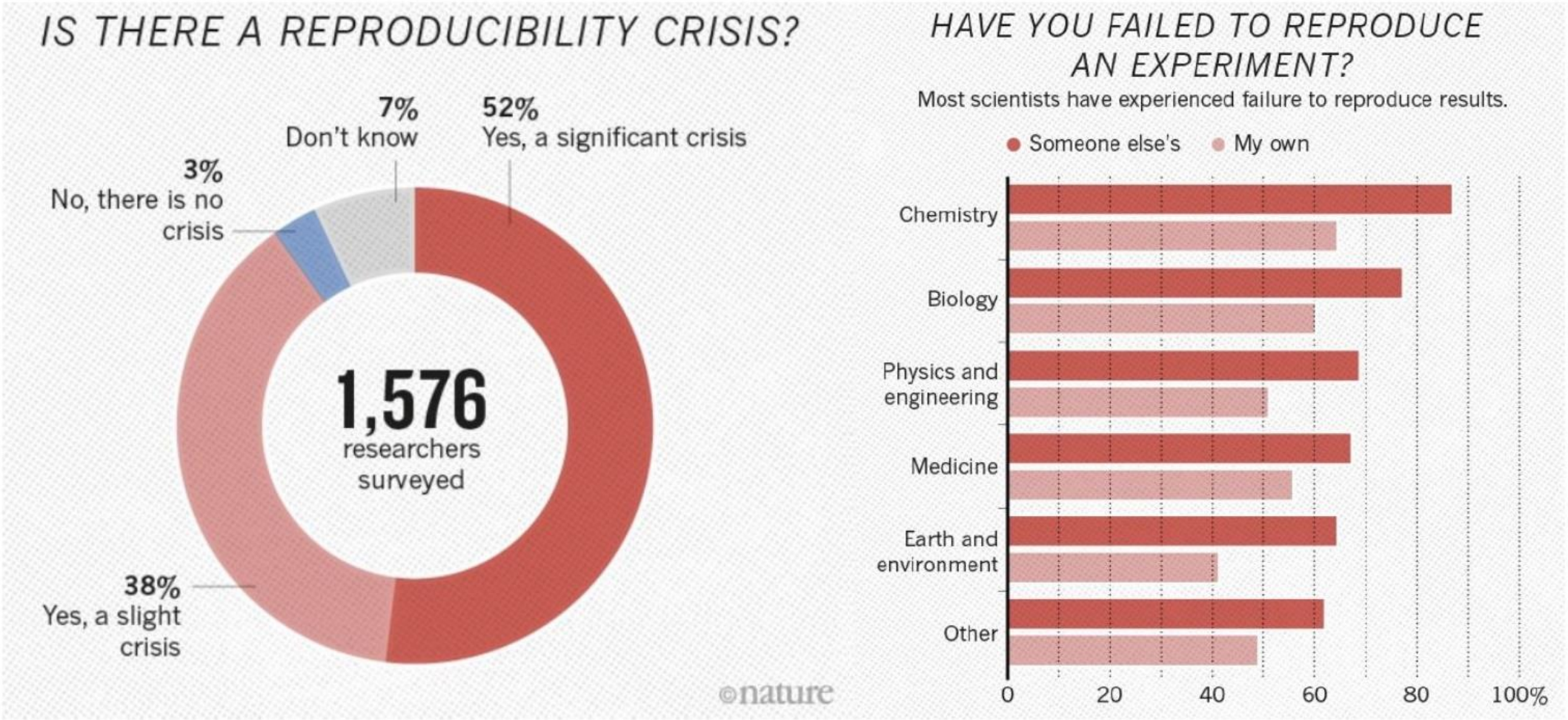
- **effect sizes** are small
- there is a **greater number and lesser preselection** of tested relationships
- there is **greater flexibility** in designs, definitions, outcomes, and analytical modes
- there is greater **financial and other interest** and prejudice

+ the propensity of **journals to publish only positive results**

+ the "**publish or perish**" culture for researchers

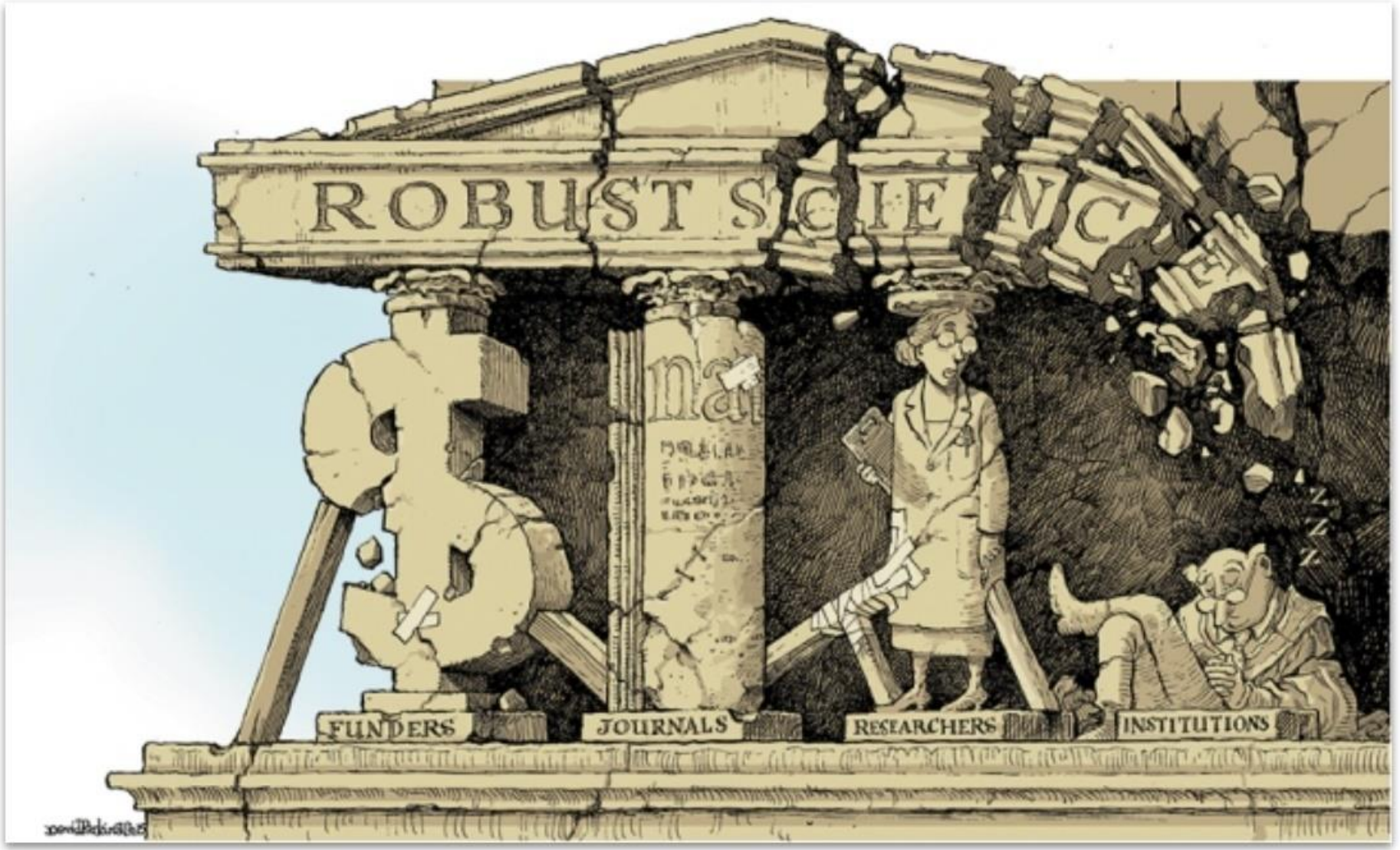
Ioannidis (2005) Why most published research findings are false. *PLoS Med* 2(8): e124 DOI: 10.1371/journal.pmed.0020124

1. An overview of the “reproducibility crisis”: awareness in 2016



"1,500 scientists lift the lid on reproducibility". Nature. 533: 452–454 - 2016

1. An overview of the “reproducibility crisis”: awareness in 2016



Source

1. An overview of the “reproducibility crisis”: not a binary phenomenon

Reproducibility is not black an white, but rather a continuum

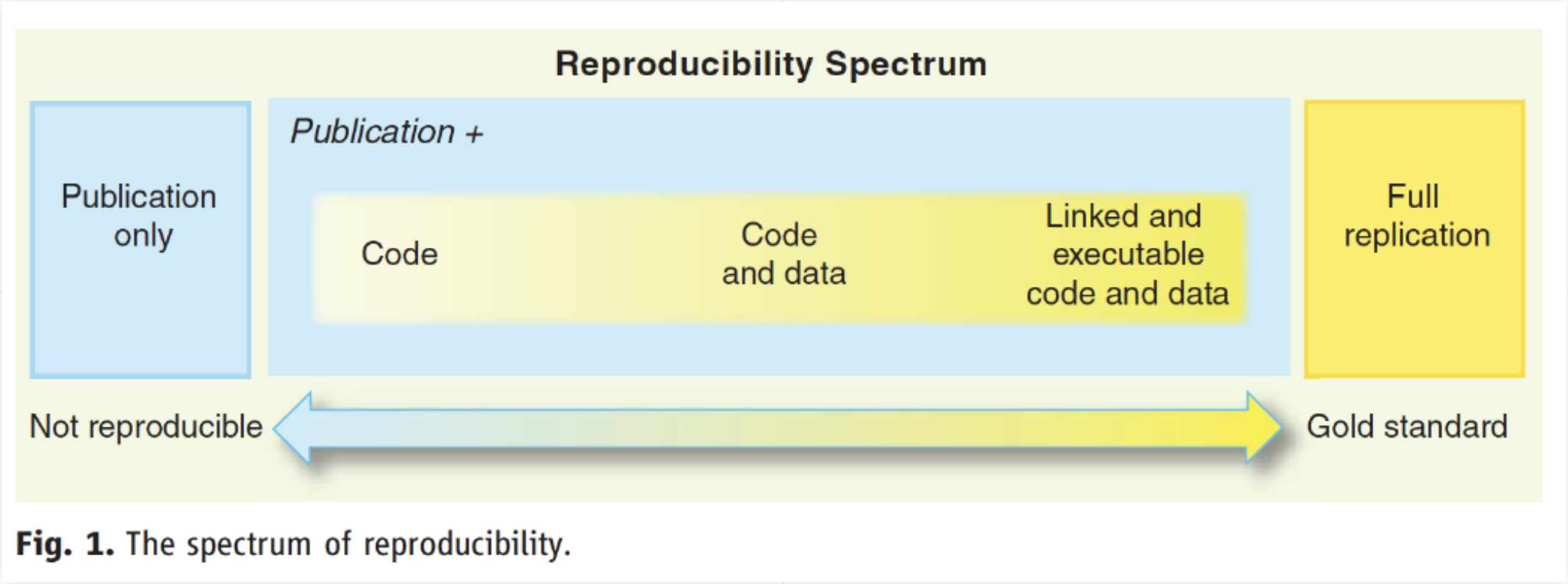


Fig. 1. The spectrum of reproducibility.

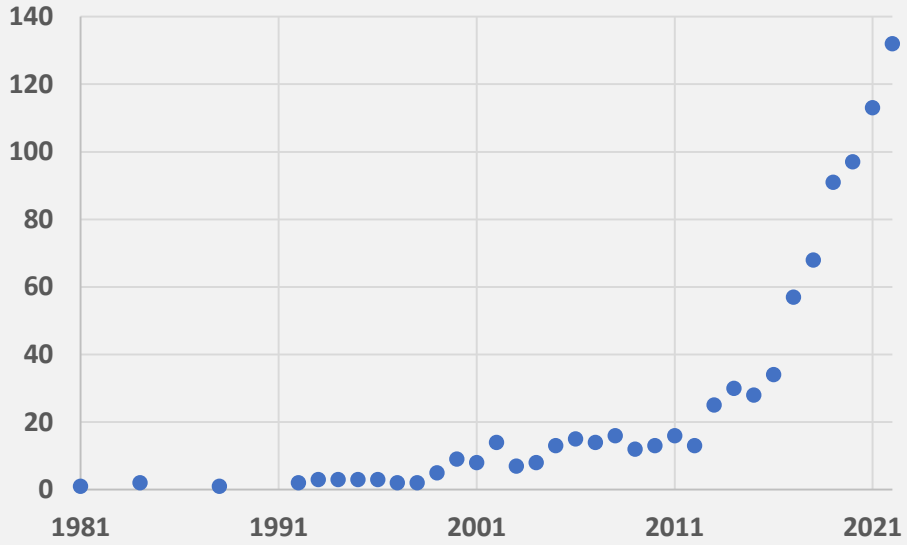
Peng, Science, 2011

What does research reproducibility mean?

The language and conceptual framework of "research reproducibility" are nonstandard and unsettled across the sciences.

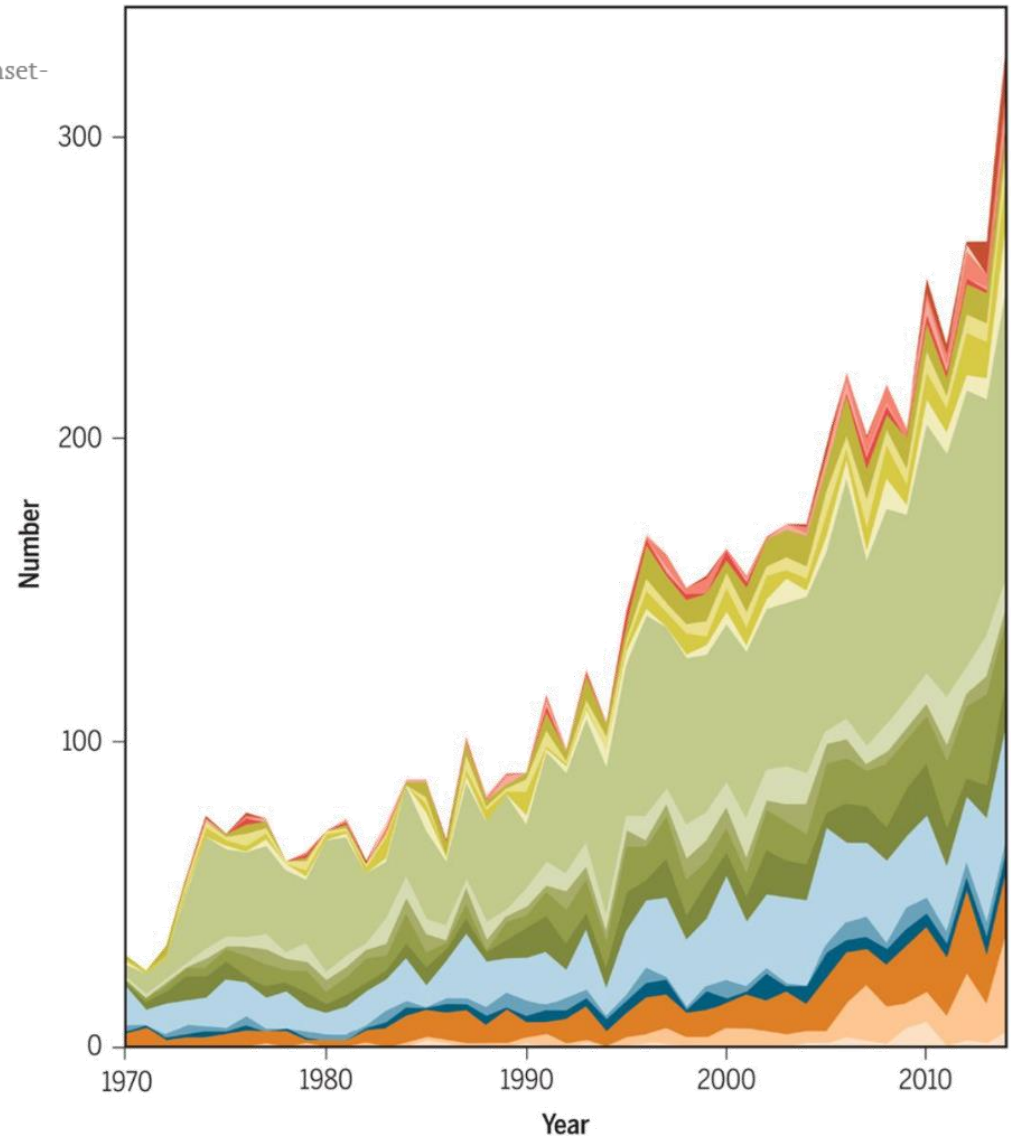
STEVEN N. GOODMAN, DANIELE FANELLI, AND JOHN P. A. IOANNIDIS Authors Info & Affiliations

SCIENCE TRANSLATIONAL MEDICINE · 1 Jun 2016 · Vol 8, Issue 341 · p. 341ps12 · DOI: 10.1126/scitranslmed.aaf5027



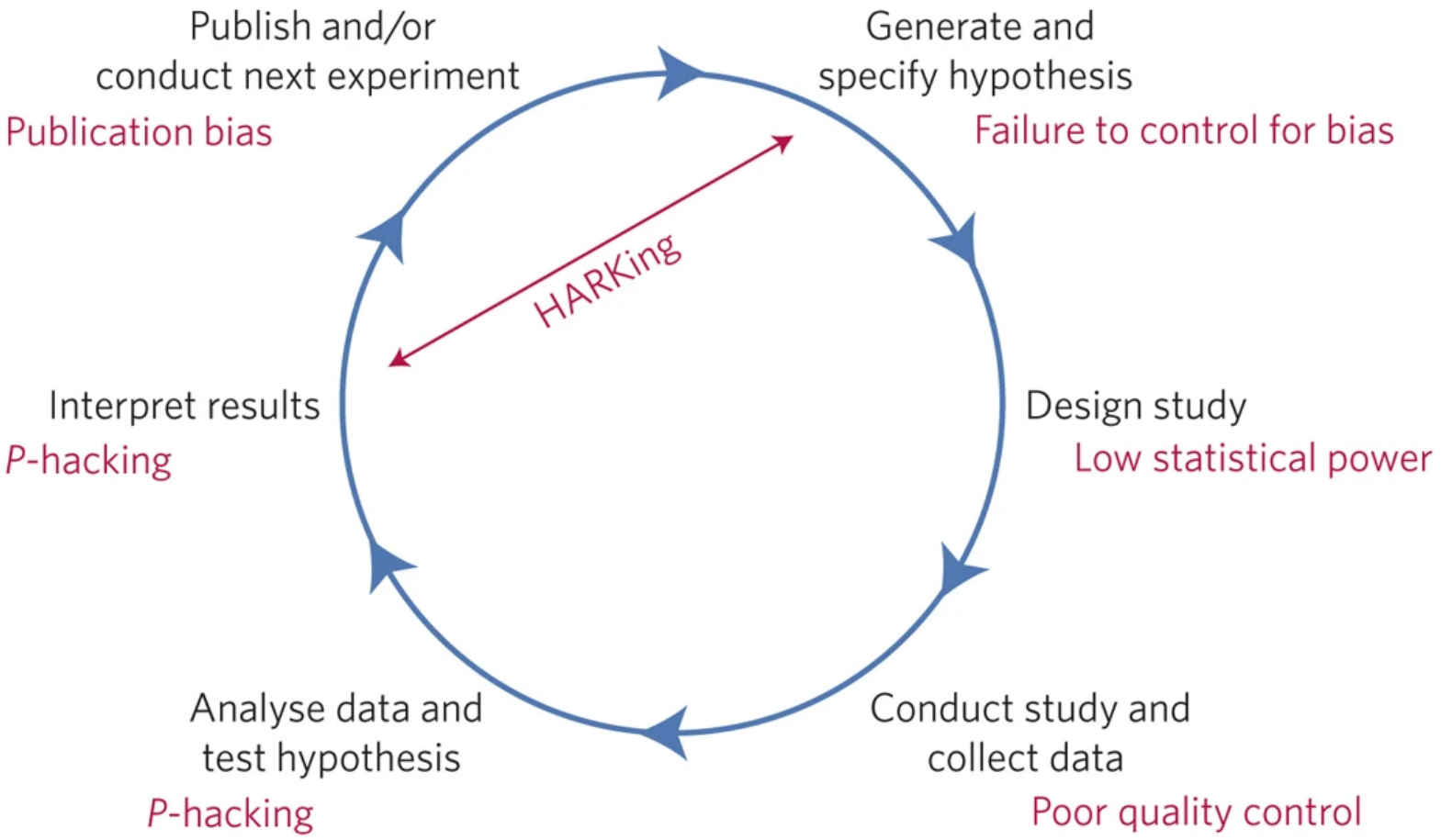
Query on Scopus:
same keywords as Goodman et al. but restricted to Neuroscience only

Conclusion: "any scientific field, require a clear specification of the kind of reproducibility being discussed (methods, results, or inferences)"



Keywords (at least 1 of the following expressions): research reproducibility, reproducibility of research, reproducibility of results, results reproducibility, reproducibility of study, study reproducibility, reproducible research, reproducible finding, or reproducible result

2. General tips to help with reproducibility: methods



Legend

Blue: An idealized version of the hypothetico-deductive model of the scientific method

Pink: threats to this model

Munafò *et al.* A manifesto for reproducible science. *Nat Hum Behav* **1**, 0021 (2017). https://doi-org.insb.bib.cnrs.fr/10.1038/s41562-016-0021

2. General tips to help with reproducibility: methods

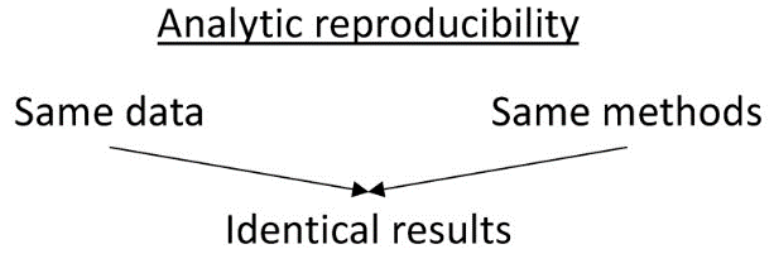
Table 1 | A manifesto for reproducible science.

Theme	Proposal	Examples of initiatives/potential solutions (extent of current adoption)	Stakeholder(s)
Methods	Protecting against cognitive biases	All of the initiatives listed below (* to ****) Blinding (**)	J, F
	Improving methodological training	Rigorous training in statistics and research methods for future researchers (*) Rigorous continuing education in statistics and methods for researchers (*)	I, F
	Independent methodological support	Involvement of methodologists in research (**) Independent oversight (*)	F
	Collaboration and team science	Multi-site studies/distributed data collection (*) Team-science consortia (*)	I, F
Reporting and dissemination	Promoting study pre-registration	Registered Reports (*) Open Science Framework (*)	J, F
	Improving the quality of reporting	Use of reporting checklists (**) Protocol checklists (*)	J
	Protecting against conflicts of interest	Disclosure of conflicts of interest (***) Exclusion/containment of financial and non-financial conflicts of interest (*)	J
Reproducibility	Encouraging transparency and open science	Open data, materials, software and so on (* to **) Pre-registration (**** for clinical trials, * for other studies)	J, F, R
Evaluation	Diversifying peer review	Preprints (* in biomedical/behavioural sciences, **** in physical sciences) Pre- and post-publication peer review, for example, Publons, PubMed Commons (*)	J
Incentives	Rewarding open and reproducible practices	Badges (*) Registered Reports (*) Transparency and Openness Promotion guidelines (*) Funding replication studies (*) Open science practices in hiring and promotion (*)	J, I, F

Estimated extent of current adoption: *, <5%; **, 5–30%; ***, 30–60%; ****, >60%. Abbreviations for key stakeholders: J, journals/publishers; F, funders; I, institutions; R, regulators.

Source

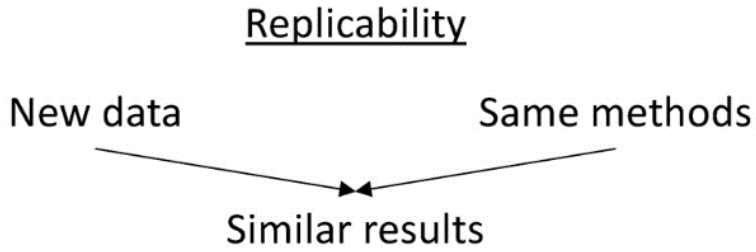
2. General tips to help with reproducibility: methods



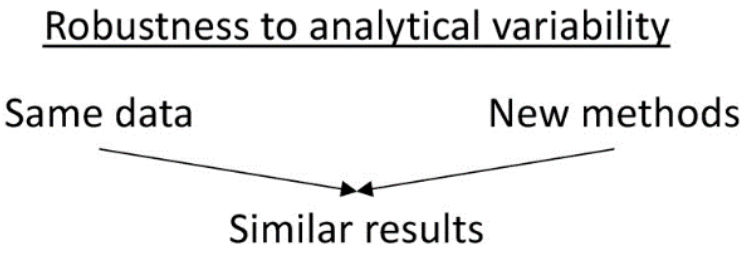
Level 1:
- Transparent, detailed reporting
- Novel publication platforms
- Tools for reporting error detection

Level 2:
- Data and code sharing
- Best practices for software engineering (version control, modularity, documentation, unit testing, DOI)
- Code reuse across labs, open-source tools
- Standardized data structures

Level 3:
- Reproduction of computing environment
- Containers



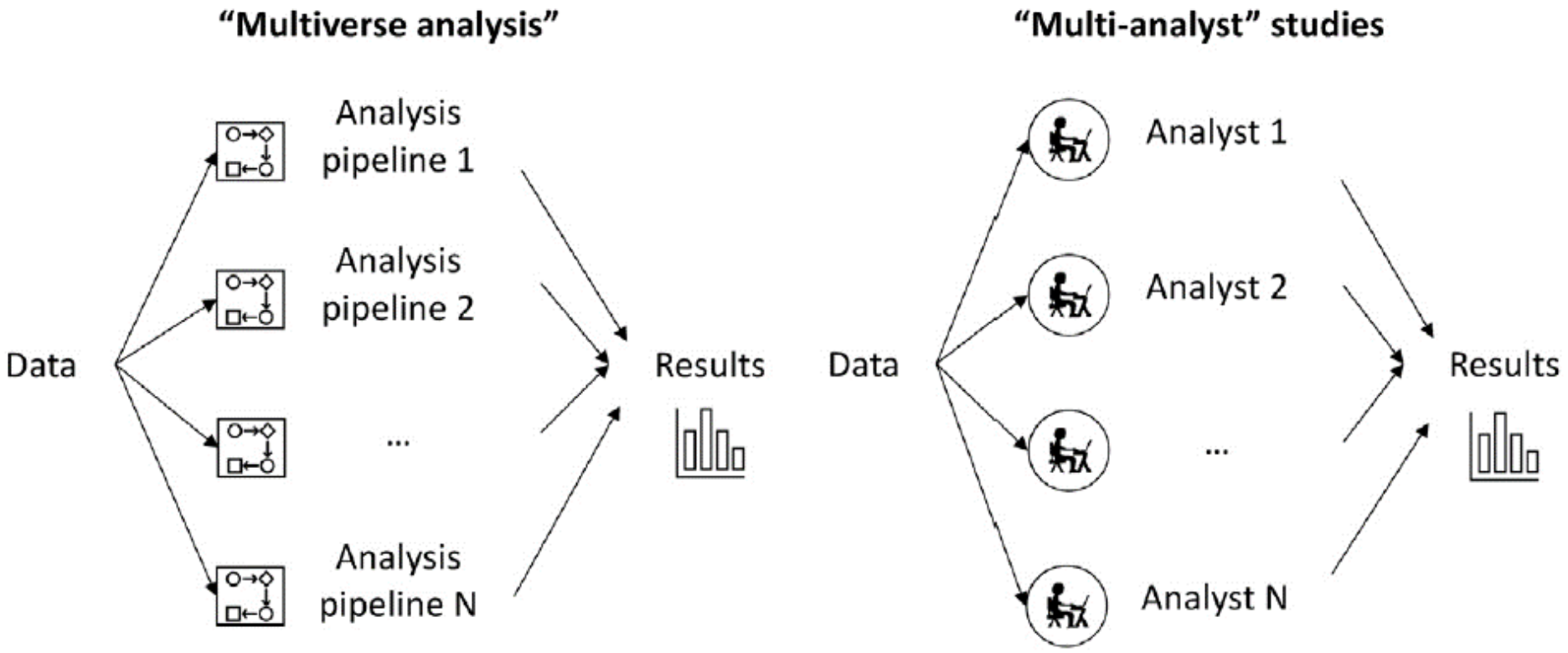
- Sample size based on power analyses
- Large-scale collaborations and consortia aggregating data across labs
- Pre-registration
- Positive controls
- External validation
- Optimizing number of participants and trials



- Transparent, detailed reporting
- Data and code sharing
- "Multiverse analysis", multi-analyst studies

Botvinik-Nezer & Wager. Reproducibility in Neuroimaging Analysis: Challenges and Solutions. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* (2022) https://doi.org/10.1016/j.bpsc.2022.12.006

2. General tips to help with reproducibility: methods



Botvinik-Nezer & Wager. Reproducibility in Neuroimaging Analysis: Challenges and Solutions. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* (2022) https://doi.org/10.1016/j.bpsc.2022.12.006

2. General tips to help with reproducibility: methods

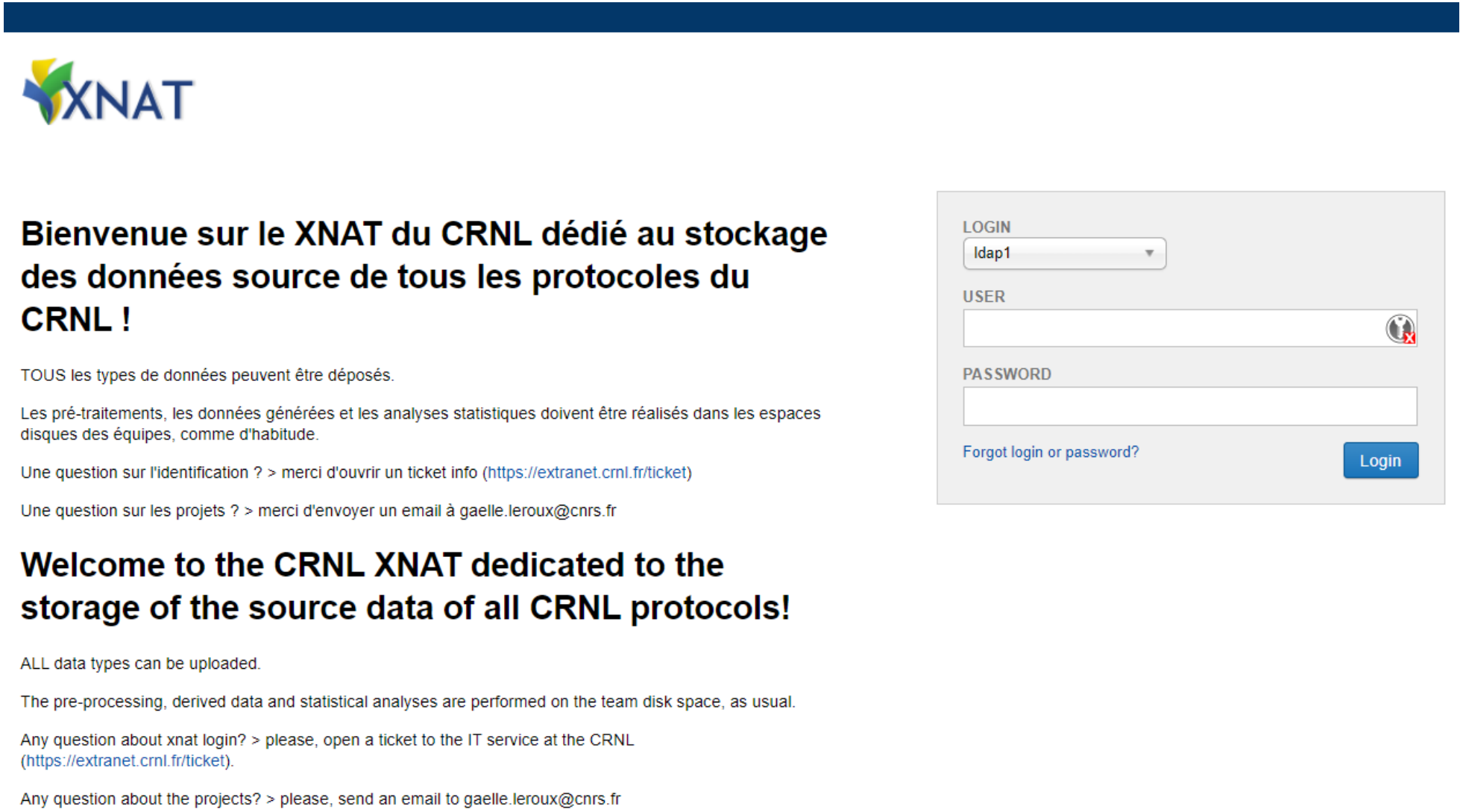
Organizational principles:

1. Put each project in its own project directory
2. Keep together what needs to be kept together
3. Keep the **source data**
4. Make file names **human-readable**
5. Make file names **machine-readable**
6. Make file names **easy to sort**
7. **Back up your data**

Source

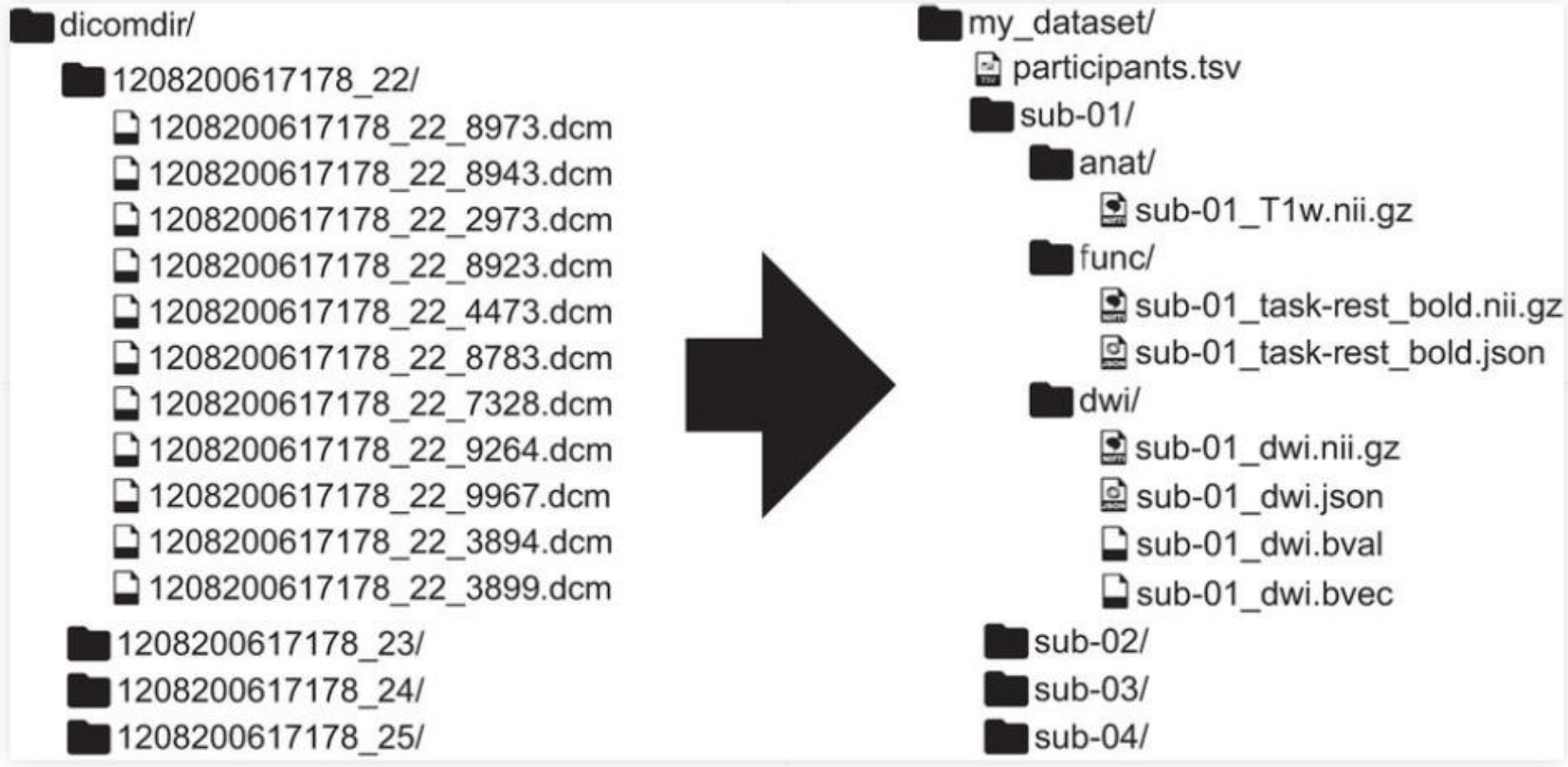
2. General tips to help with reproducibility: methods

XNAT @ CRNL (see wiki for details)



2. General tips to help with reproducibility: methods

BIDS
BRAIN IMAGING DATA STRUCTURE

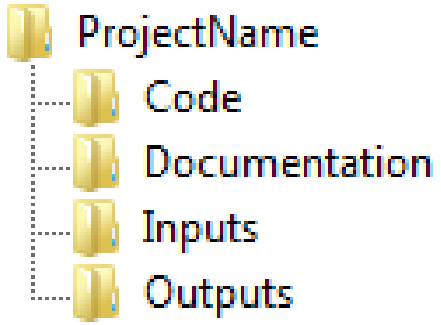


Source & further reading: Gorgolewski et al. *Sci Data*, 2016, http://bids.neuroimaging.io

2. General tips to help with reproducibility: methods

Organizational principles:

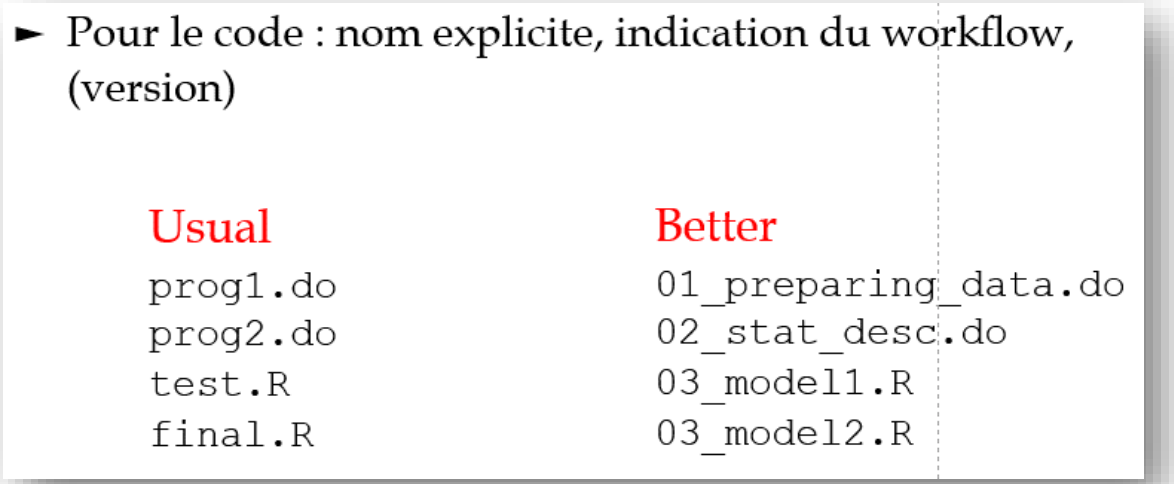
1. Put each project in its own project directory
2. Keep together what needs to be kept together
3. Keep the source data
4. Make file names human-readable
5. Make file names machine-readable
6. Make file names easy to sort
7. Back up your data



Source

Documentation principles:

8. Create an **overview** of your project
9. Make software **dependencies** explicit
10. Describe variables in a **codebook**
11. **Document** your code and its use
12. Give **meaningful names** to functions and variables
13. **Keep track** of how each result was produced



Source

2. General tips to help with reproducibility: methods

git

GitLab

GitLab @ CRNL (see wiki for details)

or https://gitlab.in2p3.fr/

GitHub

Bitbucket

2. General tips to help with reproducibility: methods

Variable Name	Description	Type	Values or Characteristics
patientid	Patient ID	Numeric	Integers
sex	Patient Gender	Numeric	1=Female, 2=Male
proceduredate	Date of Procedure	Date, M/D/YYYY	
dob	Date of Birth	Date, M/D/YYYY	
patientrace	Patient Race	Numeric	0=Not Latino, 1=Latino,
tx	Treatment Group	Numeric	1-Tx 1, 2=Tx 2, 3=Tx 3
dischargeppt	PPt at Discharge	Numeric	2 decimal places
notes	Notes	Character	
location	Location	Numeric	1=local, 2=regional, 3=distant
patientdied	Patient is Dead	Numeric	0=No, 1=Yes
deathdate	Date of Death	Date, M/D/YYYY	
clinicaloutcome	Clinical Outcome	Character	
outcomecomments	Outcome Comments	Character	

Source: http://domstat.med.ucla.edu/pages/codebook2

Further reading: Codebook cookbook - A guide to writing a good codebook for data analysis projects in medicine, Creating a codebook - Document, Discover, and Interoperate alliance

Tools: codebook generator for R: memisc

Source

2. General tips to help with reproducibility: methods

Organizational principles:

1. Put each project in its own project directory
2. Keep together what needs to be kept together
3. Keep the source data
4. Make file names human-readable
5. Make file names machine-readable
6. Make file names easy to sort
7. Back up your data

Documentation principles:

8. Create an overview of your project
9. Make software dependencies explicit
10. Describe variables in a codebook
11. Document your code and its use
12. Give meaningful names to functions and variables
13. Keep track of how each result was produced

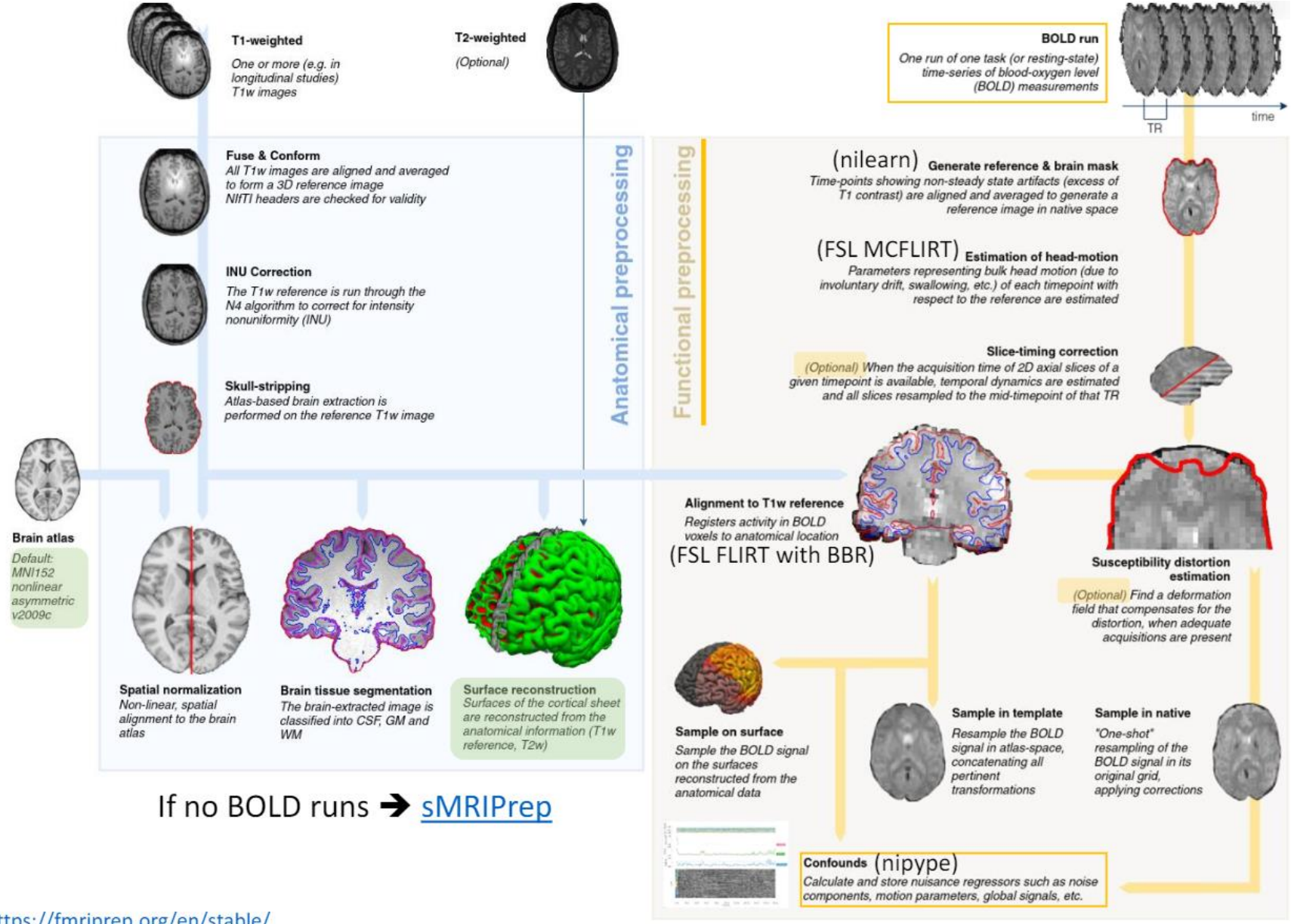
Principles of automation :

14. **Avoid manual** data manipulation steps
15. **Modularize code** rather than copy and paste
16. **Reuse code** rather than rewrite it
17. Don't comment or uncomment code sections to control program behavior
18. **Add assertions** to programs to check their operation
19. Save the random number generator seed
20. Record all intermediate results

Source

2. General tips to help with reproducibility: methods

fMRIPrep workflow



If no BOLD runs ➔ sMRIPrep

Source: https://fmriprep.org/en/stable/

2. General tips to help with reproducibility: methods

Organizational principles:

1. Put each project in its own project directory
2. Keep together what needs to be kept together
3. Keep the source data
4. Make file names human-readable
5. Make file names machine-readable
6. Make file names easy to sort
7. Back up your data

Documentation principles:

8. Create an overview of your project
9. Make software dependencies explicit
10. Describe variables in a codebook
11. Document your code and its use
12. Give meaningful names to functions and variables
13. Keep track of how each result was produced

Principles of automation :

14. Avoid manual data manipulation steps
15. Modularize code rather than copy and paste
16. Reuse code rather than rewrite it
17. Don't comment or uncomment code sections to control program behavior
18. Add assertions to programs to check their operation
19. Save the random number generator seed
20. Record all intermediate results

Distribution principles:

21. Store data in **open formats**
22. Process data with **open software**
23. Eliminate hard-coded paths in your code
24. Link text statements to underlying results
25. Submit code and data to a **DOI-issuing repository** and reference it in your article
26. Make the **license** of code and data explicit

Source

File format	Recommendation	Avoid
Document texte	PDF, TXT, ODT	MS Word, RTF
Feuille de calcul	ODS, CSV	MS Excel, PDF, OOXML
Base de données	SQL, SIARD, DB tables (.CSV)	MS Access, dBase (.dbf), HDF5
Données statistiques	SPSS Portable, STATA, XML, CSV, TXT	SAS et R
Images	JPEG, TIFF, PNG	DICOM
Audio	BWF, MXF, Matroska (.mka), FLAC, OPUS	WAVE, MP3, AAC, AIFF, OGG
Video	MXF, MKV	MPEG-4, MPEG-2, AVI, QuickTime (.mov, .qt)
Information géographique	GML, MIF/MID	ESRI Shapefiles, MapInfo, KML
Images géoréférencées	GeoTIFF (.tif, .tiff)	TIFF World File
Raster	ASCII GRID (.asc, .txt)	ESRI GRID

https://facile.cines.fr/ service de validation des formats

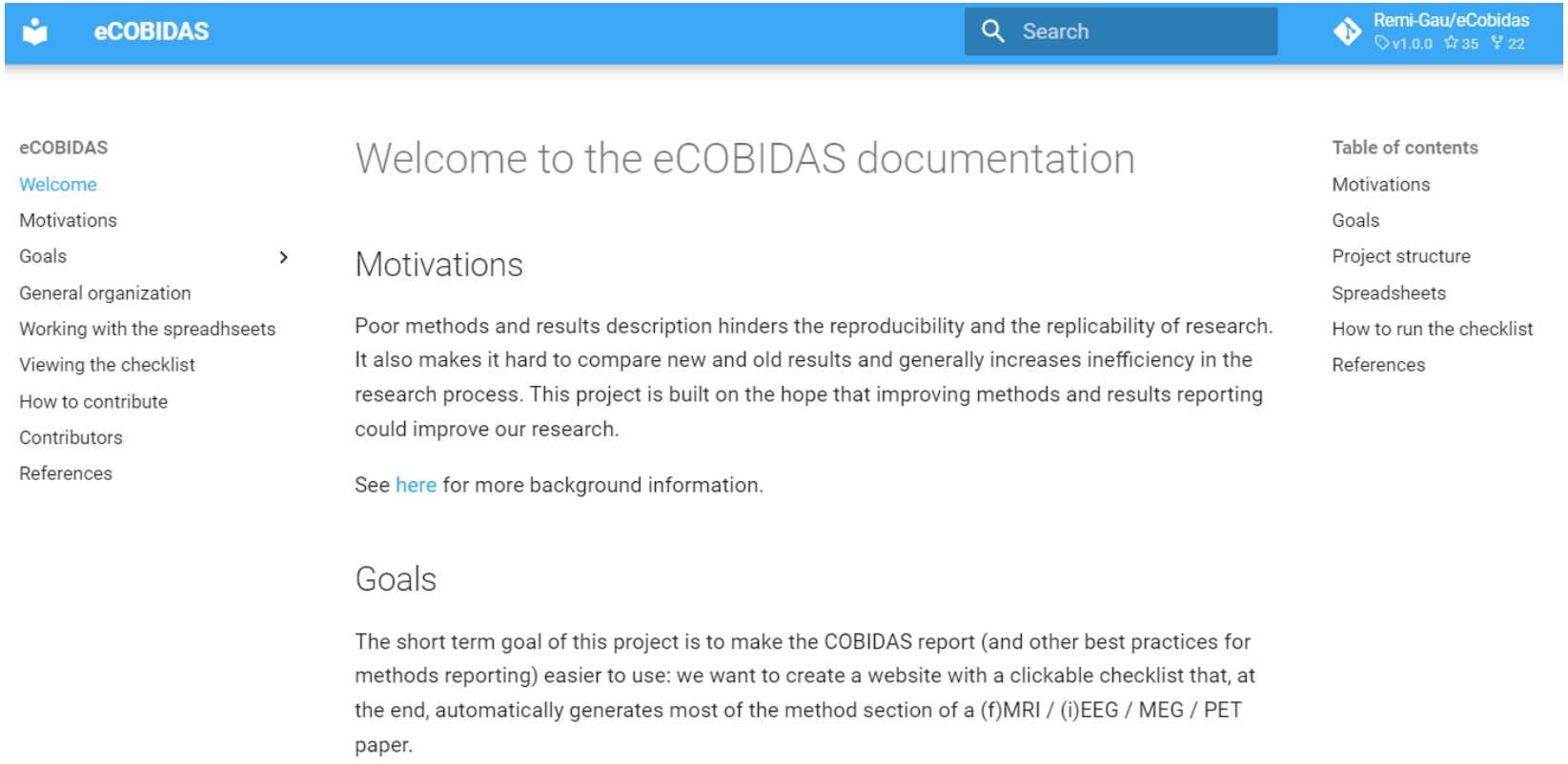
In practice, you can often work with a popular closed format and convert it to an open format. But you need to check whether the conversion alters the information, and take compensatory measures if necessary. Example: XLSX -> CSV conversion loses formatting.

Source

eCOBIDAS checklist

1. experimental design reporting, **2.** image acquisition reporting, **3.** preprocessing reporting, **4.** statistical modeling, **5.** results reporting, **6.** data sharing, **7.** reproducibility

For each of 7 areas of a study, a tabular listing of over 100 items to help plan, execute, report, and share research in the most transparent fashion is provided.

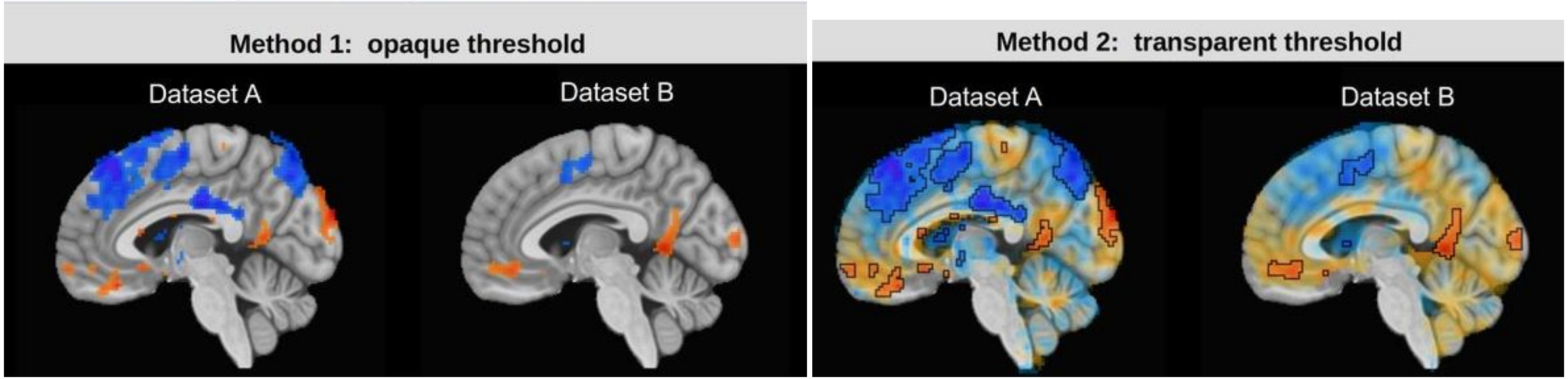


https://remi-gau.github.io/eCobidas/

2. General tips to help with reproducibility: results of fMRI data

- Most studies do not present all results of their analysis, **hiding subthreshold ones**.
- **Hiding results negatively affects the interpretation and understanding of the study**.
- **Neuroimagers should present all results of their study, highlighting key**

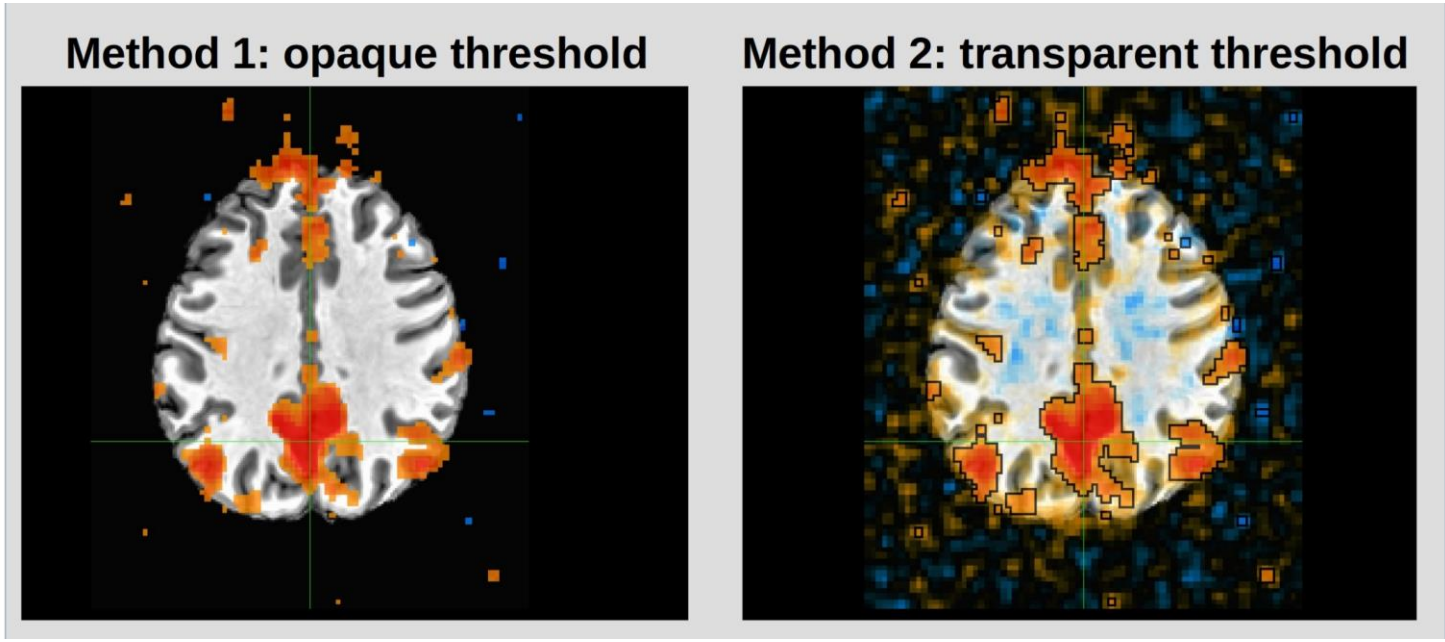
Q: Which method for viewing FMRI data makes it easier to understand the results, and to interpret how they might be similar and/or different? How would your description of the "reproducibility" of the results change when moving from the 1st to the 2nd?



Thresholds hide interesting effects!

Taylor & al. Highlight results, don't hide them: Enhance interpretation, reduce biases and improve reproducibility. *NeuroImage* (2023)

2. General tips to help with reproducibility: results of fMRI data



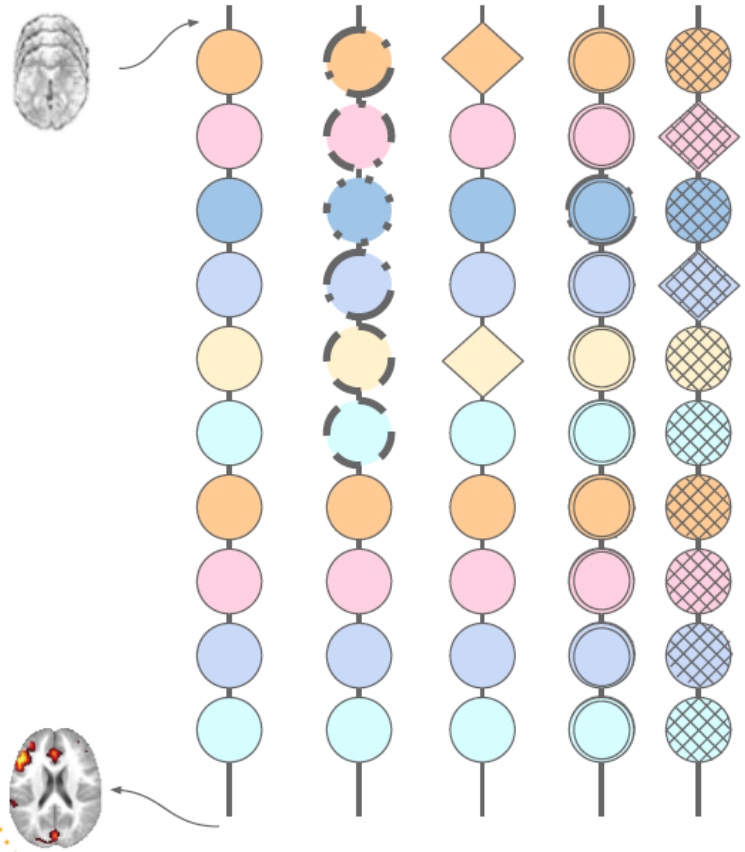
& other times, thresholds hide artifacts and badness.

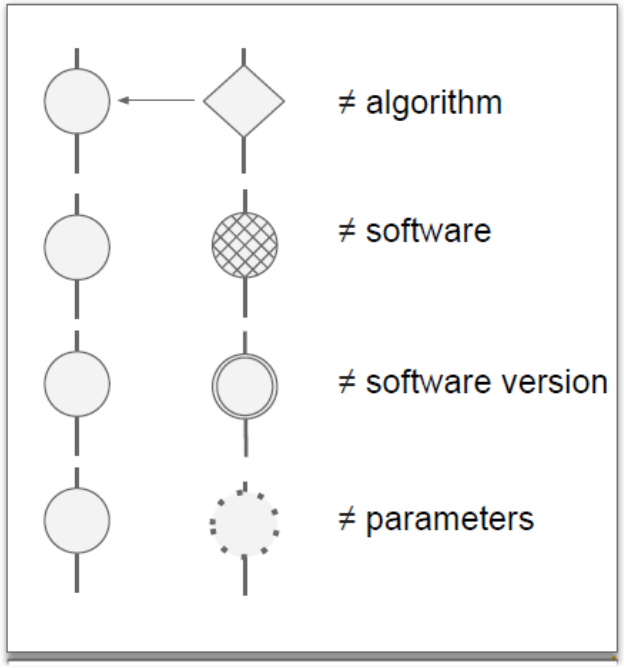
"We need to see each. Showing more results helps with quality control (QC) evaluation.
Become more confident in (or appropriately worried about!) your data." Paul Taylor

Taylor & al. Highlight results, don't hide them: Enhance interpretation, reduce biases and improve reproducibility. *NeuroImage* (2023)

2. General tips to help with reproducibility: results of fMRI data

Analytical pipelines

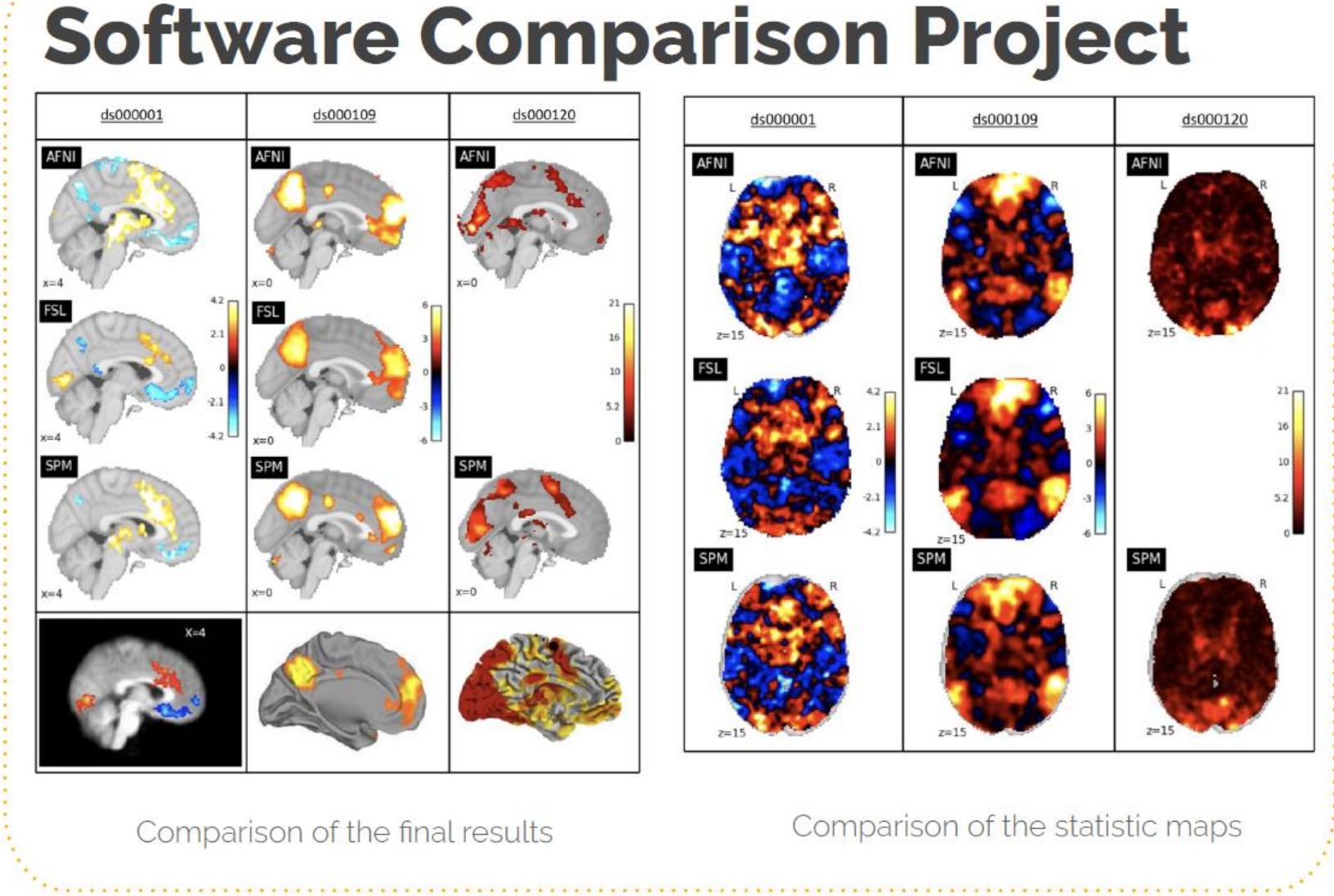




A family of "acceptable" pipelines
over 10^30 combinations…

Maumet. Different analysis pipelines, different results... How to deal with analytical variability? 20th IEEE International Symposium on Biomedical Imaging (ISBI 2023), Cartagena (Colombia). ffhal-04073419f

2. General tips to help with reproducibility: results of fMRI data



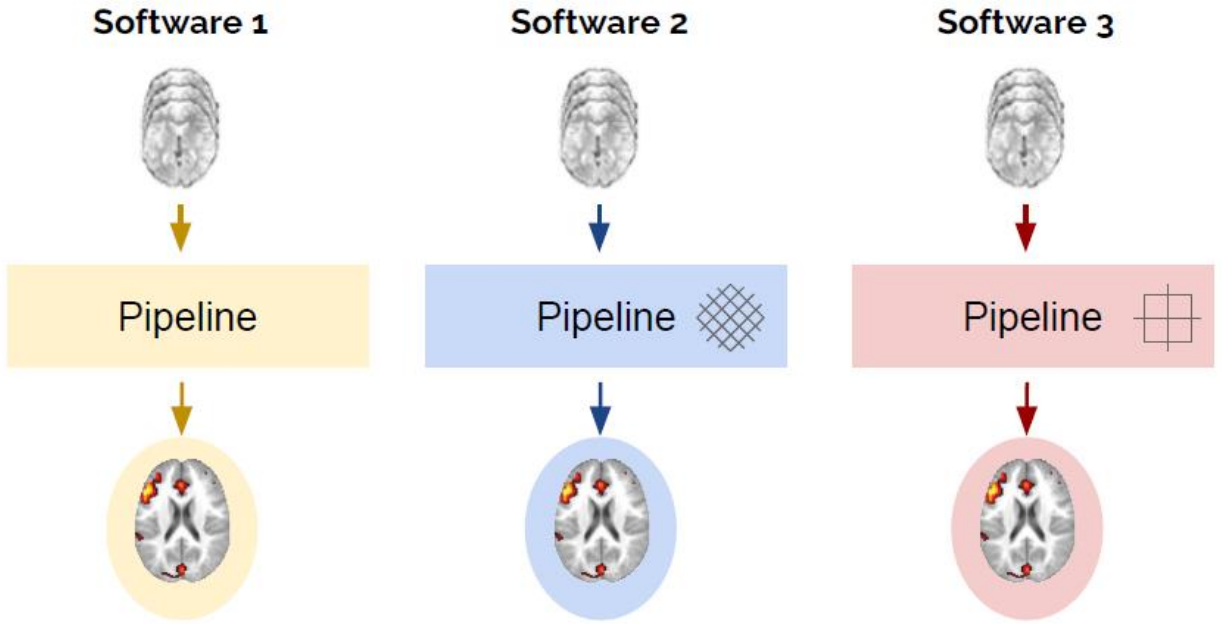
Maumet. Different analysis pipelines, different results... How to deal with analytical variability? 20th IEEE International Symposium on Biomedical Imaging (ISBI 2023), Cartagena (Colombia). ffhal-04073419f

2. General tips to help with reproducibility: results of fMRI data

Variability across software

- Reproduced 3 published functional MRI studies
  - ds001, ds109, ds120 (from Openneuro)
- Using 3 different software packages
  - AFNI, FSL, SPM

Alex Bowring Tom Nichols



Maumet. Different analysis pipelines, different results... How to deal with analytical variability? 20th IEEE International Symposium on Biomedical Imaging (ISBI 2023), Cartagena (Colombia). ffhal-04073419f

The Neuroimaging Analysis Replication and Prediction Study

What happens when
70 independent groups of researchers
analyze the same fMRI dataset?

www.narps.info

2. General tips to help with reproducibility: results of fMRI data

Conclusions of the study:

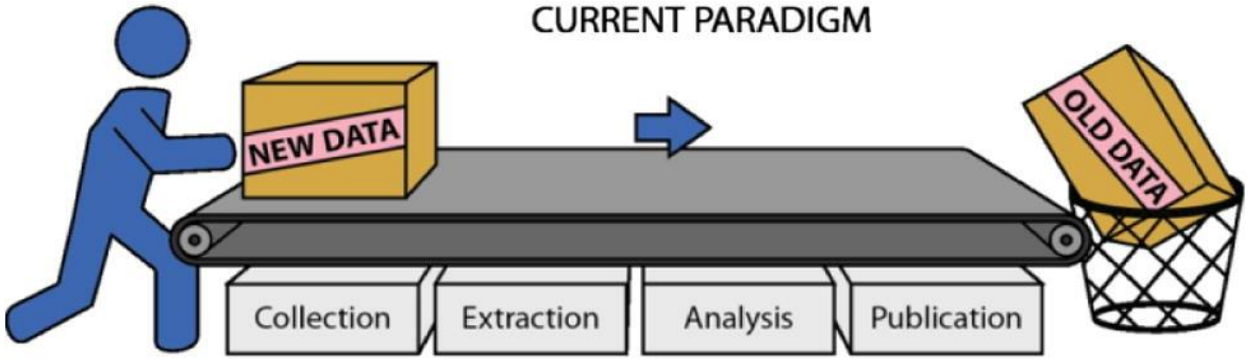
The strongest factors in explaining the differences between results:

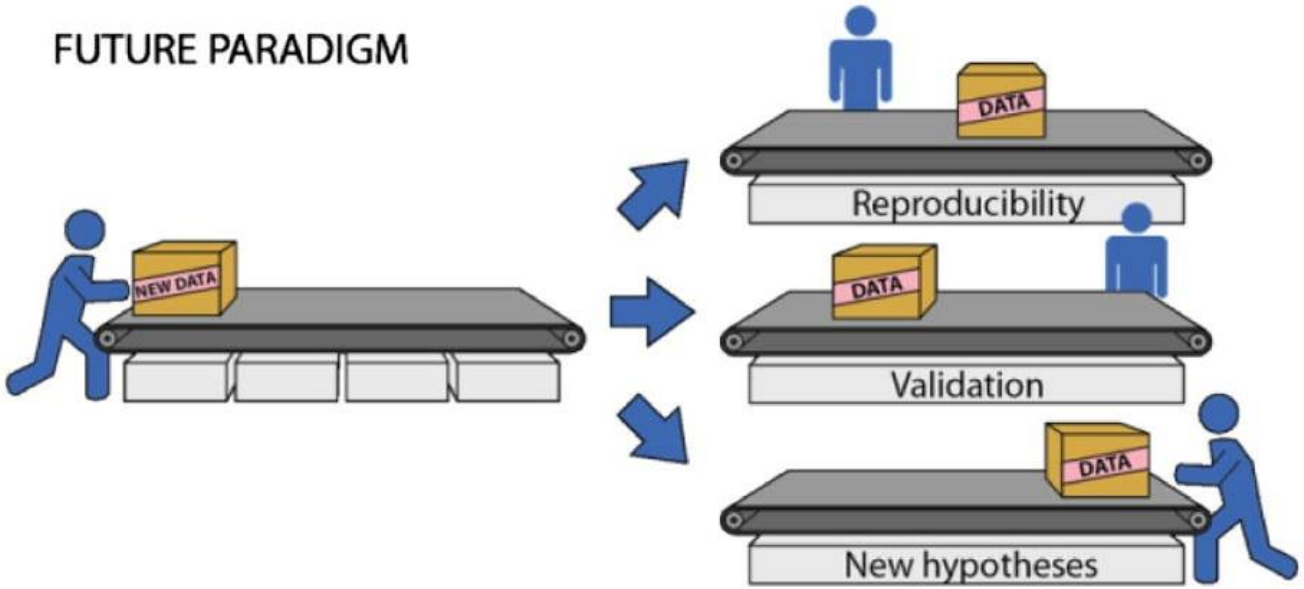
1. **Spatial smoothness** : higher FWHM lead to a higher likelihood of significant outcomes
2. **FSL vs other softs**: FSL has higher likelihood of significant results across all hypotheses
3. **Multiple correction method**
4. **No significant effect of standardized** vs customized preprocessing pipelines

Recommendations to:

1. **Pre-register** the study, hypotheses, methods
2. **Share the pipelinee, data and code** because the Mat & Methods section is never enough
3. **Share un- and thresholded maps** to allow image-based meta-analyses (e.g. on NeuroVault)
4. **Use multiple pipelines** to analyse the data ("multiverse analysis"), several teams
5. Report them and **find the consensus using a meta-analysis**, not the "best" one.
6. Use **pipeline-optimization tools** to reduce analysis flexibility (e.g. FitLins & LaConte article)
7. Use **sensitivity analyses** to evaluate the effects of pipeline decisions

Botvinik-Nezer *et al.* Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* **582**, 84–88 (2020). https://doi-org.insb.bib.cnrs.fr/10.1038/s41586-020-2314-9





"My hope is that others will be able to reproduce some of our core findings and extend them in interesting new directions."
–Pr. James R. Booth

Lambin et al. *Radiother. Oncol*. 2013 doi: 10.1016/j.radonc.2013.07.007

https://lab.vanderbilt.edu/boothlab/datasets/

Resources

Open science in the doctoral school https://nsco.universite-lyon.fr/science-ouverte/

French network on reproductible science: https://www.recherche-reproductible.fr/
UK Reproductibility Network https://www.ukrn.org/
OSIRIS: OPEN SCIENCE TO INCREASE REPRODUCIBILITY IN SCIENCE

"Scientific integrity" Club(CRNL): https://project.crnl.fr/IntegriteScientifique/
NeuroImaging club meetings: https://osf.io/sxkgq/

Additional tools to the ones cited in the slides:
Electronic labbooks https://datacc.elab.one/login.php & https://cle.inserm.fr/users/sign_in
Research Data Plan https://dmp.opidor.fr/
Support to data management & sharing https://doranum.fr/