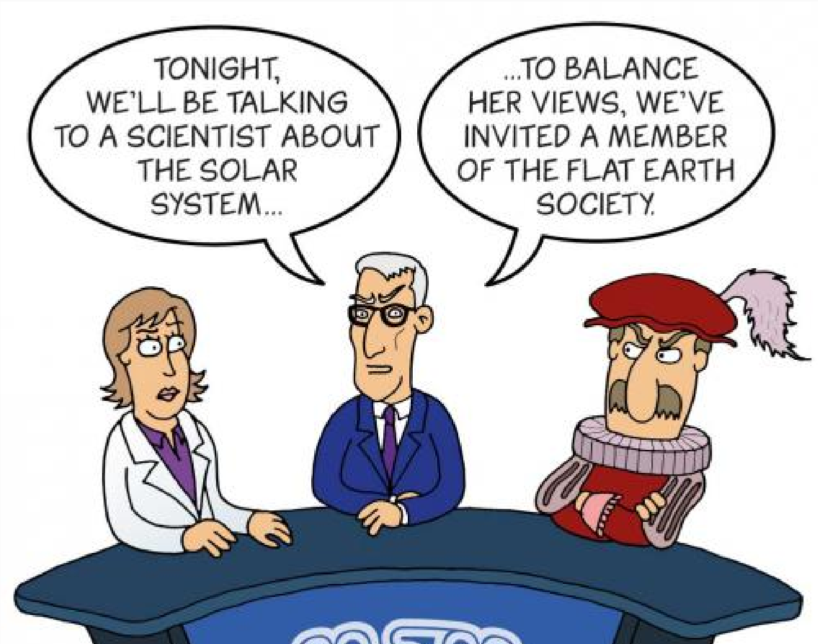
Reproducibility Research and Open Science

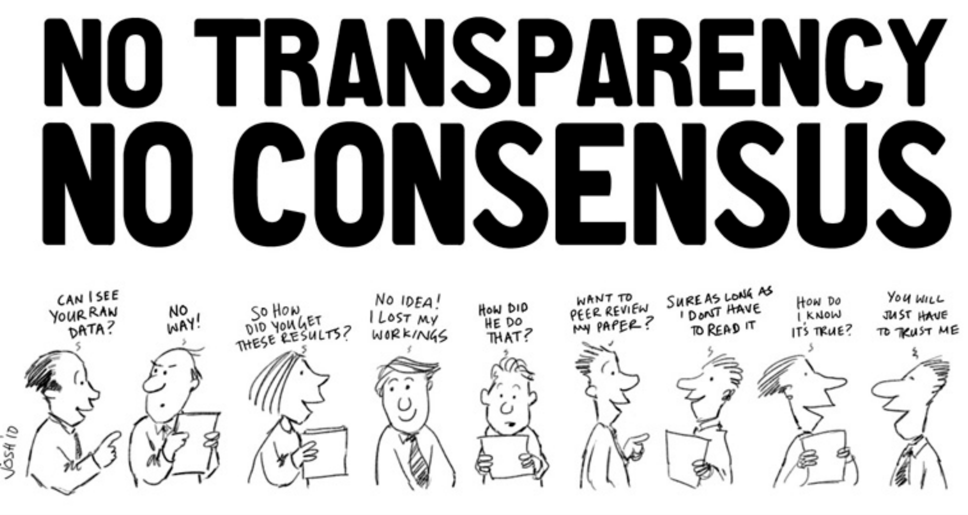
Arnaud Legrand

Doing a PhD, good practice and pitfalls to avoid
October 2023

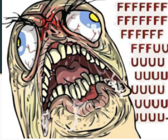
Scientific Consensus


TONIGHT, WE'LL BE TALKING TO A SCIENTIST ABOUT THE SOLAR SYSTEM...
...TO BALANCE HER VIEWS, WE'VE INVITED A MEMBER OF THE FLAT EARTH SOCIETY.

Scientific Consensus


NO TRANSPARENCY
NO CONSENSUS
CAN I SEE YOUR RAW DATA?
NO WAY!
SO HOW DID YOU GET THESE RESULTS?
NO IDEA! I LOST MY WORKINGS
HOW DID HE DO THAT?
WANT TO PEER REVIEW MY PAPER?
SURE AS LONG AS I DON'T HAVE TO READ IT
HOW DO I KNOW IT'S TRUE?
YOU WILL JUST HAVE TO TRUST ME

Common Horror Stories 1/4: *What did I do?*



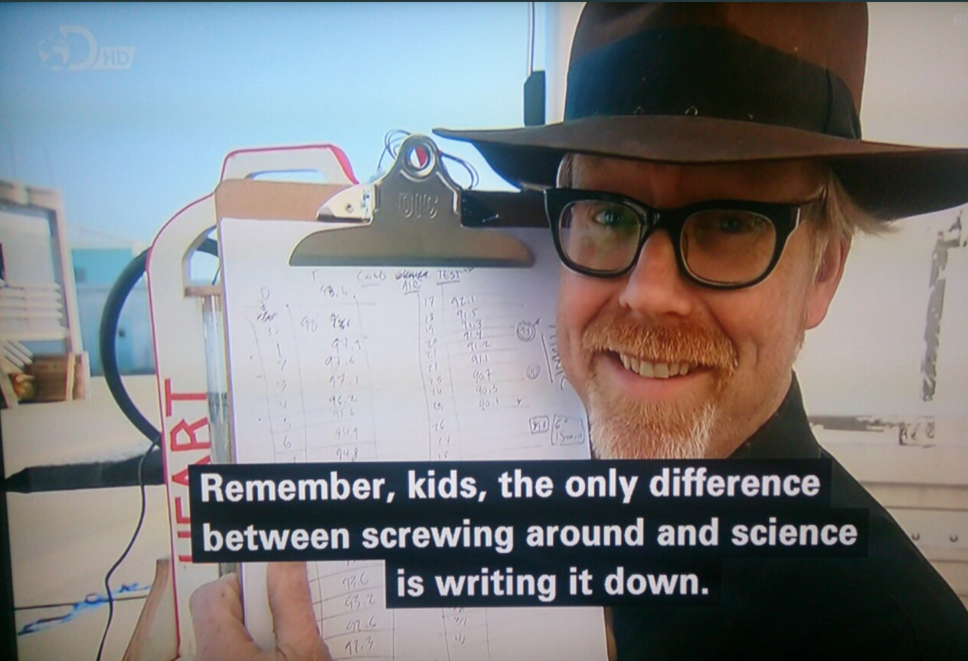
Author

- I thought I used the same parameters but I'm getting different results!
- The new student wants to compare with the method I proposed last year
- My advisor asked me whether I took care of setting this or this but I can't remember
- The damned fourth reviewer asked for a major revision and wants me to change Figure 3. Which code and which data set did I use?
- It worked yesterday! 6 months later: Why did I do that?

Reviewer

- As usual, there is no confidence interval, I wonder about the variability and whether the difference is significant or not
- That can't be true, I'm sure they removed some points
- Why is this graph in logscale? How would it look like otherwise? I'm not even sure of what this value means. If only I could access the generation script

Mythbusters: Science vs. Screwing Around



Remember, kids, the only difference between screwing around and science is writing it down.

Common Horror Stories 2/4: *Argh... damned computers*

- Hey! Here is my code. It’s on GitHub so feel free to play with it! I’m doing open science ☺
- **Alice**: I got 3.123123 **Bob**: I got segfault
- Damned! It used to work!!! Whenever I upgrade my computer, things break so I try to stay away from this ☹
- Whenever trying the code of my colleague, I had to install Foo but I broke everything and now neither his code nor mine works! ☹

Seriously ? It’s 21st century. 😉 How come all this is so painful ?

Containers and Package Managers

The good

Guix

The bad

The uggly

ReproZip

Automatic tracking

Containers and Package Managers

The good

Guix

The bad

The uggly

ReproZip

Automatic tracking

Containers

- Pros: Lightweight, Good isolation, Easy to use
  - Running as easy as `docker run <img> <cmd>`
  - Building images: `docker build -f <Dockerfile>`
  - Sharing through the Docker Hub: `docker pull/push <img>`

Containers and Package Managers

The good

Guix

The bad

The uggly

ReproZip

Automatic tracking

Containers

- Pros: Lightweight, Good isolation, Easy to use
- Cons: Opaque, Container build is generally not reproducible
  - Recipes rarely follow *reproducible good practices*

```
    FROM ubuntu:20.04
    RUN apt-get update
        && apt-get upgrade -y
        && apt-get install -y ...
```

- Choose a stable image (and the smallest possible)
- Include only the necessary libraries (e.g. no graphics libs)
- Avoid system updates (instead freeze sources)

Containers and Package Managers

The good The bad The uggly

Automatic tracking

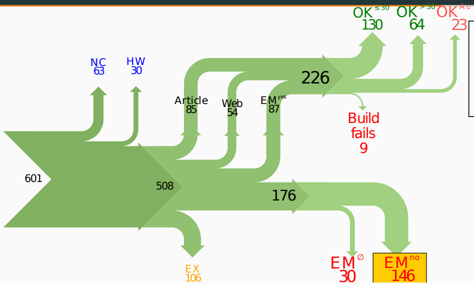
Containers

- Pros: Lightweight, Good isolation, Easy to use
- Cons: Opaque, Container build is generally not reproducible

Package managers (the uggly and the good)

- Language specific: `pip/pipenv/virtualenv`, `conda`, `CRAN/Bioconductor`
  - Limits: version management, durability, permeable, language centric
- GUIX/NiX = Full-fledged functional package manager
  - Native support for environment (*à la git*)
  - Isolation through `--pure`
  - Recompile from source (cache recommended)

Common Horror Stories 3/4: *Please hold on*



Collberg, Christian *et Al.*, *Measuring Reproducibility in Computer Systems Research*, http://reproducibility.cs.arizona.edu/

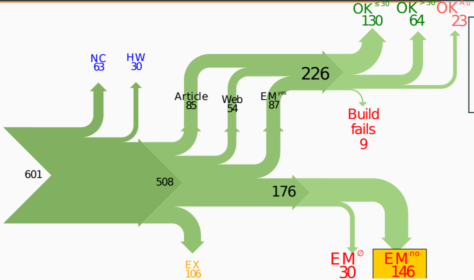
- 8 ACM conferences (ASPLOS'12, CCS'12, OOPSLA'12, OSDI'12, PLDI'12, SIGMOD'12, SOSP'11, VLDB'12) and 5 journals
- EM^{no}= the code cannot be provided

- Versionning Problems

Thanks for your interest in the implementation of our paper. The good news is that I was able to find some code. I am just hoping that it is a stable working version of the code, and matches the implementation we finally used for the paper. Unfortunately, I have lost some data when my laptop was stolen last year. The bad news is that the code is not commented and/or clean.

Attached is the (system) source code of our algorithm. I'm not very sure whether

Common Horror Stories 3/4: *Please hold on*



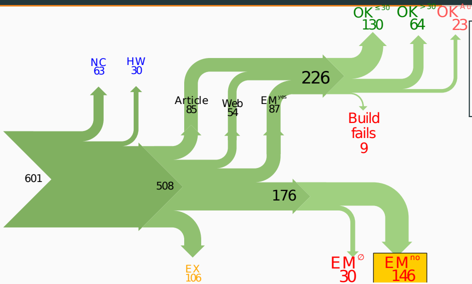
Collberg, Christian *et Al.*, *Measuring Reproducibility in Computer Systems Research*, http://reproducibility.cs.arizona.edu/

- 8 ACM conferences (ASPLOS'12, CCS'12, OOPSLA'12, OSDI'12, PLDI'12, SIGMOD'12, SOSP'11, VLDB'12) and 5 journals
- EM^{no}= the code cannot be provided

- Versionning Problems
- Bad Backup Practices

Unfortunately, the server in which my implementation was stored had a disk crash in April and three disks crashed simultaneously. While the help desk made significant effort to save the data, my entire implementation for this paper was not found.

Common Horror Stories 3/4: *Please hold on*



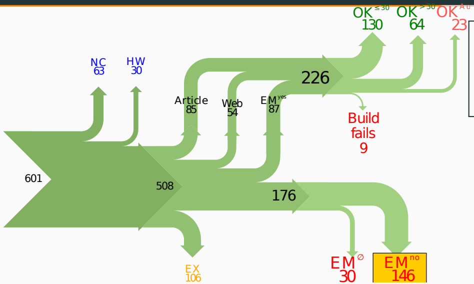
Collberg, Christian *et Al.*, *Measuring Reproducibility in Computer Systems Research*, http://reproducibility.cs.arizona.edu/

- 8 ACM conferences (ASPLOS'12, CCS'12, OOPSLA'12, OSDI'12, PLDI'12, SIGMOD'12, SOSP'11, VLDB'12) and 5 journals
- EM^{no}= the code cannot be provided

- Versionning Problems
- Bad Backup Practices
- Code Will be Available Soon

Unfortunately the current system is not mature enough at the moment, so it's not yet publicly available. We are actively working on a number of extensions and things are somewhat volatile. However, once things stabilize we plan to release it to outside users. At that point, we would be happy to send you a copy.

Common Horror Stories 3/4: *Please hold on*



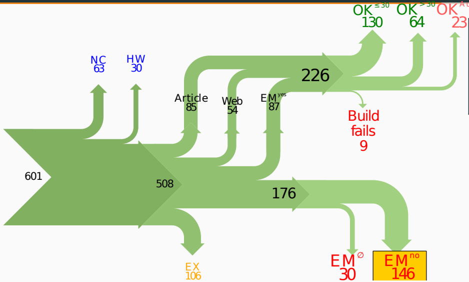
Collberg, Christian *et Al.*, *Measuring Reproducibility in Computer Systems Research*, http://reproducibility.cs.arizona.edu/

- 8 ACM conferences (ASPLOS'12, CCS'12, OOPSLA'12, OSDI'12, PLDI'12, SIGMOD'12, SOSP'11, VLDB'12) and 5 journals
- EM^{no}= the code cannot be provided

- Versionning Problems
- Bad Backup Practices
- Code Will be Available Soon
- No Intention to Release

I am afraid that the source code was never released. The code was never intended to be released so is not in any shape for general use.

Common Horror Stories 3/4: *Please hold on*



Collberg, Christian *et Al.*, *Measuring Reproducibility in Computer Systems Research*, http://reproducibility.cs.arizona.edu/

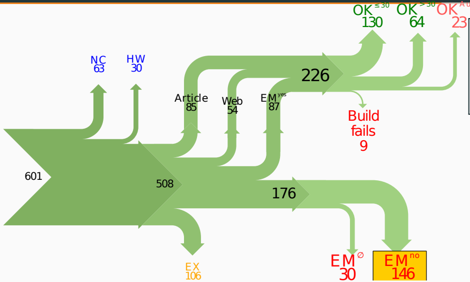
- 8 ACM conferences (ASPLOS'12, CCS'12, OOPSLA'12, OSDI'12, PLDI'12, SIGMOD'12, SOSP'11, VLDB'12) and 5 journals
- EM^{no}= the code cannot be provided
  - Programmer Left

- Versionning Problems
- Bad Backup Practices
- Code Will be Available Soon
- No Intention to Release

⟨STUDENT⟩ was a graduate student in our program but he left a while back so I am responding instead. For the paper we used a prototype that included many moving pieces that only ⟨STUDENT⟩ knew how to operate and we did not have the time to integrate them in a ready-to-share implementation before he left. Still, I hope you can build on the ideas/technique of the paper.

Unfortunately, the author who has done most of the coding for this paper has

Common Horror Stories 3/4: *Please hold on*



Collberg, Christian *et Al.*, *Measuring Reproducibility in Computer Systems Research*, http://reproducibility.cs.arizona.edu/

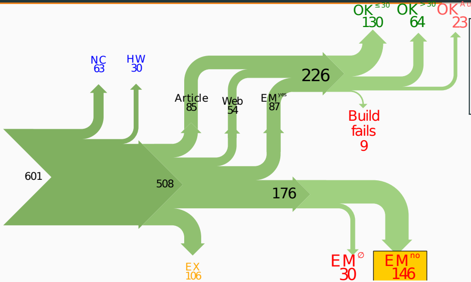
- 8 ACM conferences (ASPLOS'12, CCS'12, OOPSLA'12, OSDI'12, PLDI'12, SIGMOD'12, SOSP'11, VLDB'12) and 5 journals
- EM^{no}= the code cannot be provided
  - Programmer Left
  - Commercial Code

- Versionning Problems
- Bad Backup Practices
- Code Will be Available Soon
- No Intention to Release

Since this work has been done at ⟨COMPANY⟩ we don't open-source code unless there is a compelling business reason to do so. So unfortunately I don't think we'll be able to share it with you.

The code owned by ⟨COMPANY⟩, and AFAIK the code is not open-source. Your best bet is to reimplement :(Sorry.

Common Horror Stories 3/4: *Please hold on*



- Versionning Problems
- Bad Backup Practices
- Code Will be Available Soon
- No Intention to Release

Collberg, Christian *et Al.*, *Measuring Reproducibility in Computer Systems Research*, http://reproducibility.cs.arizona.edu/

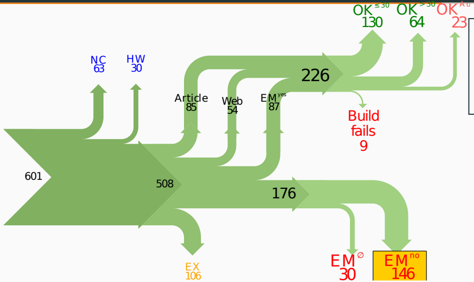
- 8 ACM conferences (ASPLOS'12, CCS'12, OOPSLA'12, OSDI'12, PLDI'12, SIGMOD'12, SOSP'11, VLDB'12) and 5 journals
- EM^{no}= the code cannot be provided
  - Programmer Left
  - Commercial Code
  - Proprietary Academic Code

Unfortunately, the ⟨SYSTEM⟩ sources are not meant to be opensource (the code is partially property of ⟨UNIVERSITY 1⟩, ⟨UNIVERSITY 2⟩ and ⟨UNIVERSITY 3⟩.)

If this will change I will let you know, albeit I do not think there is an intention to make the ⟨SYSTEM⟩ sources opensource in the near future.

If you're interested in obtaining the code, we only ask for a description of the re-

Common Horror Stories 3/4: *Please hold on*



Collberg, Christian *et Al.*, *Measuring Reproducibility in Computer Systems Research*, http://reproducibility.cs.arizona.edu/

- 8 ACM conferences (ASPLOS'12, CCS'12, OOPSLA'12, OSDI'12, PLDI'12, SIGMOD'12, SOSP'11, VLDB'12) and 5 journals
- EM^{no}= the code cannot be provided

- Versionning Problems
- Bad Backup Practices
- Code Will be Available Soon
- No Intention to Release

- Programmer Left
- Commercial Code
- Proprietary Academic Code
- Research vs. Sharing

In the past when we attempted to share it, we found ourselves spending more time getting outsiders up to speed than on our own research. So I finally had to establish the policy that we will not provide the source code outside the group.

Horror Stories 4/4: Fighting Information Loss with Archives

or $=$ awesome collaborations ($\neq$ archive)

- D. Spinellis. *The Decay and Failures of URL References*. CACM, 46(1), 2003
  The half-life of a referenced URL is approximately 4 years
- P. Habibzadeh. *Decay of References to Web sites in Articles Published in General Medical Journals: Mainstream vs Small Journals*. Applied Clinical Informatics. 4 (4), 2013

  half life ranged from 2.2 years in EMHJ to 5.3 years in BMJ
- Discontinuated forges: Code Space, Gitorious, Google code, Inria Gforge

Horror Stories 4/4: Fighting Information Loss with Archives

or = awesome collaborations (≠ archive)

- D. Spinellis. *The Decay and Failures of URL References*. CACM, 46(1), 2003
  The half-life of a referenced URL is approximately 4 years
- P. Habibzadeh. *Decay of References to Web sites in Articles Published in General Medical Journals: Mainstream vs Small Journals*. Applied Clinical Informatics. 4 (4), 2013
  half life ranged from 2.2 years in EMHJ to 5.3 years in BMJ
- Discontinuated forges: Code Space, Gitorious, Google code, Inria Gforge

Article archives arXiv.org HAL archives-ouvertes.fr

Data archives figshare zenodo

Software Archive Software Heritage Collect/Preserve/Share

INTERNET ARCHIVE

Separation between articles, code, and data is not so simple though

Different Reproducibility Concerns in Modern Science

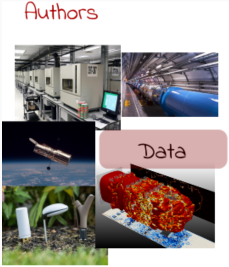
Social Sciences, Oncology, … methodology, statistics, pre-registration

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical issues

Artificial Intelligence most of the above

The processing steps between raw observations and findings have gotten increasingly numerous and complex



Different Reproducibility Concerns in Modern Science

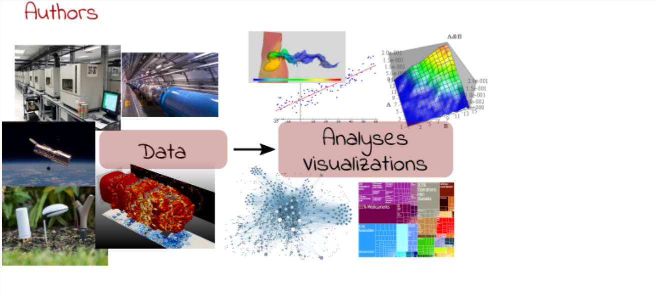
Social Sciences, Oncology, ... methodology, statistics, pre-registration

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical issues

Artificial Intelligence most of the above

The processing steps between raw observations and findings have gotten increasingly numerous and complex



Different Reproducibility Concerns in Modern Science

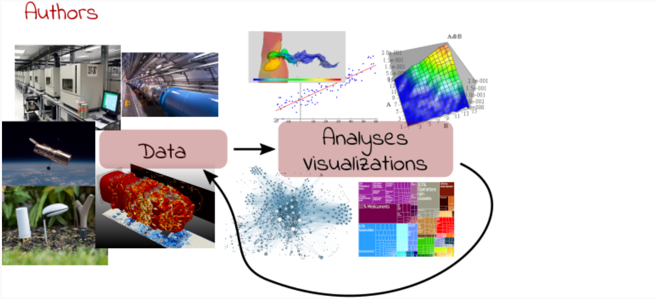
Social Sciences, Oncology, ... methodology, statistics, pre-registration

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical issues

Artificial Intelligence most of the above

The processing steps between raw observations and findings have gotten increasingly numerous and complex



Different Reproducibility Concerns in Modern Science

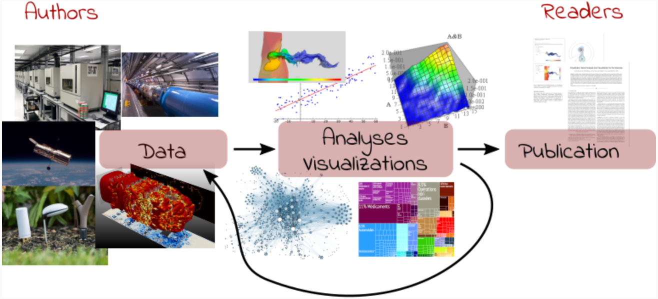
Social Sciences, Oncology, … methodology, statistics, pre-registration

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical issues

Artificial Intelligence most of the above

The processing steps between raw observations and findings have gotten increasingly numerous and complex



Different Reproducibility Concerns in Modern Science

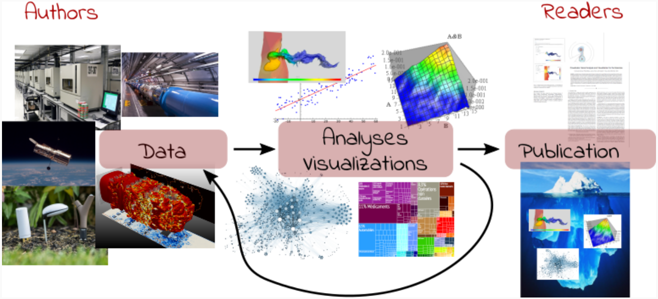
Social Sciences, Oncology, ... methodology, statistics, pre-registration

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical issues

Artificial Intelligence most of the above

The processing steps between raw observations and findings have gotten increasingly numerous and complex



Different Reproducibility Concerns in Modern Science

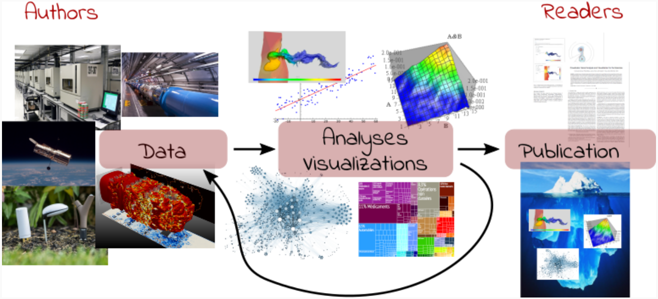
Social Sciences, Oncology, … methodology, statistics, pre-registration

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical issues

Artificial Intelligence most of the above

The processing steps between raw observations and findings have gotten increasingly numerous and complex



Reproducible Research = Bridging the Gap by working Transparently

Changing Research Practices

Soft. Engineering, Statistics, and Reproducible Research in the curricula

Manifesto: *"I solemnly pledge"* (WSSSPE, Lorena Barba, FAIR)

1. I will teach my graduate students about reproducibility
2. All our research code (and writing) is under version control
3. We will always carry out verification and validation
4. We will share data, plotting script & figure under CC-BY
5. We will upload the preprint to arXiv at the time of submission of a paper
6. We will release code at the time of submission of a paper
7. We will add a "Reproducibility" declaration at the end of each paper
8. I will keep an up-to-date web presence

Learn and Teach using online resources like

- Software Carpentry, The Turing Way, ...

Changing Publishing Practices

Artifact evaluation and ACM badges

Artifact CGO AEC Evaluated
Artifact PPoPP AEC Evaluated
Artifacts Available acm
Artifacts Evaluated Functional acm
Artifacts Evaluated Reusable acm
Results Replicated acm
Results Reproduced acm

Major conferences

- Supercomputing: Artifact Description (AD) mandatory, Artifact Evaluation (AE) still optional, Double blind vs. RR
- NeurIPS, ICLR: open reviews, reproducibility challenge

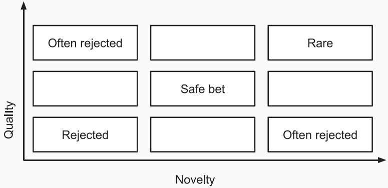
  Joelle Pineau @ NeurIPS'18

- ACM SIGMOD 2015-2019, Most Reproducible Paper Award…

Mentalitie are evolving people care, make stuff available, errors are found and fixed

Changing Academic Practices (~~Publish or Perish~~)

- Goodhart's Law: Are Academic Metrics Being Gamed?, M. Fire 2019
  - AI: over 1,000 ranked journals (×10 in 15 years)
  - Shorter papers with increasing self references
  - More and more papers without any citation
  - Sharp increase in the number of new authors publishing at a much faster rate given their career age
- The Truth, The Whole Truth, and Nothing But the Truth: A Pragmatic, Guide to Assessing Empirical Evaluations, *TOPLAS* 2016



Often rejected

Rare

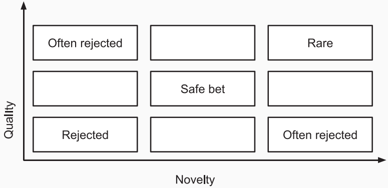
Safe bet

Rejected

Often rejected

Changing Academic Practices (~~Publish or Perish~~)

- Goodhart's Law: Are Academic Metrics Being Gamed?, M. Fire 2019
  - AI: over 1,000 ranked journals (×10 in 15 years)
  - Shorter papers with increasing self references
  - More and more papers without any citation
  - Sharp increase in the number of new authors publishing at a much faster rate given their career age
- The Truth, The Whole Truth, and Nothing But the Truth: A Pragmatic, Guide to Assessing Empirical Evaluations, *TOPLAS* 2016



Often rejected		Rare
	Safe bet	
Rejected		Often rejected

We redo Science !
GITHUB.COM/RESCIENCE

- Impact factor abandoned by Dutch university in hiring and promotion, decisions. Nature, June 2021. *Faculty and staff members at Utrecht University will be evaluated by their commitment to open science*

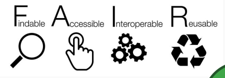
Reproducible Research ∼ Open Science ?

Plan National pour la Science Ouverte (BSN ⇝ CoSO)

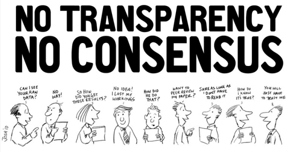
- CNRS, Inria, INRAE, ...
- Many flavors: *Citizen Science*

Main pillars:

1. Open access
2. Open data

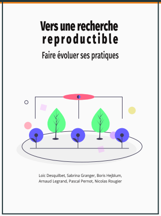


3. Open source
    - *Open hardware*
4. **Open methodology** (Reproducible Research)
    - *Open-notebook science*
    - *Open science infrastructures*
5. Open peer review (avoid collusion)
6. Open educational resources


NO TRANSPARENCY
NO CONSENSUS
CAN I SEE YOUR RAW DATA?
NO WAY!
SO HOW DID YOU GET THESE RESULTS?
NO IDEA! I LOST MY WORKINGS
HOW DID HE DO THAT?
WANT TO PEER REVIEW MY PAPER?
SURE AS LONG AS I DON'T HAVE TO READ IT
HOW DO I KNOW IT'S TRUE?
YOU WILL JUST HAVE TO TRUST ME
JOSH ID

Obviously making code/data available for the reproduction of results from published papers has become the new norm

Resources and Acknowledgments



Vers une recherche reproductible

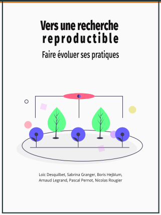
Faire évoluer ses pratiques

Loïc Desquilbet, Sabrina Granger, Boris Hejblum, Arnaud Legrand, Pascal Pernot, Nicolas Rougier

A non-technical introduction to reproducibility issues (in French)

- Loïc Desquilbet, Sabrina Granger, Boris Hejblum, Pascal Pernot, Nicolas Rougier

Resources and Acknowledgments



Vers une recherche reproductible

Faire évoluer ses pratiques

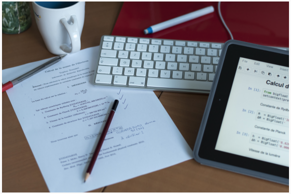
Loïc Desquilbet, Sabrina Granger, Boris Hejblum, Arnaud Legrand, Pascal Pernot, Nicolas Rougier

A non-technical introduction to reproducibility issues (in French)

- Loïc Desquilbet, Sabrina Granger, Boris Hejblum, Pascal Pernot, Nicolas Rougier

MOOC Reproducible Research: Methodological principles for a transparent science, Learning Lab Inria

- Konrad Hinsen, Christophe Pouzat
- 3rd Edition: March 2020 – March 2024 (17,000+)



Stay tuned for the MOOC "Advanced RR" planned for ~~2021~~ ~~2022~~ ~~2023~~ 2024

- Managing data (`git annex`, Zenodo, SWH)
- Software environment control (`docker`, `singularity`, `guix`)
- Scientific workflow (`make`, `snakemake`)

That's all Folks!