Scientific Methodology and (Empirical|Experimental) Evaluation

Arnaud Legrand

Master 2 MOSIG
2024-2025

Overview

Teachers and contact




Resources

- Github (schedule, slides, homeworks)
- MOOC (Learn reproducible research by yourself)

Communication

- Pad (notes and homework reporting)
- Mattermost (questions, references, ...)

Quick Poll

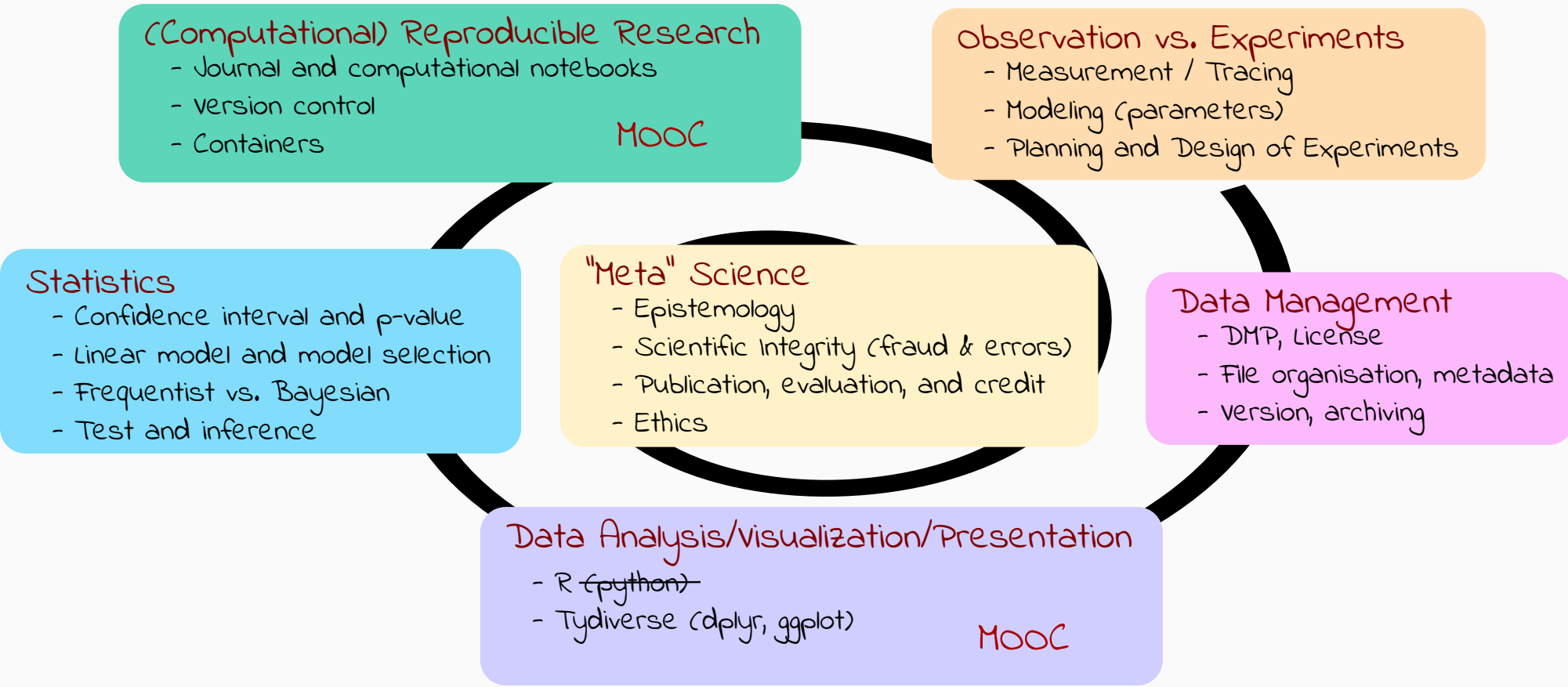
- Python
- R
- Notebooks (Rstudio, Jupyter, Org-Mode)
- Git
- Zenodo, Software Heritage
- Docker
- Confidence Interval
- P-value, P-hacking
- Linear regression
- Model Selection
- Design of Experiments
- Reinforcement Learning

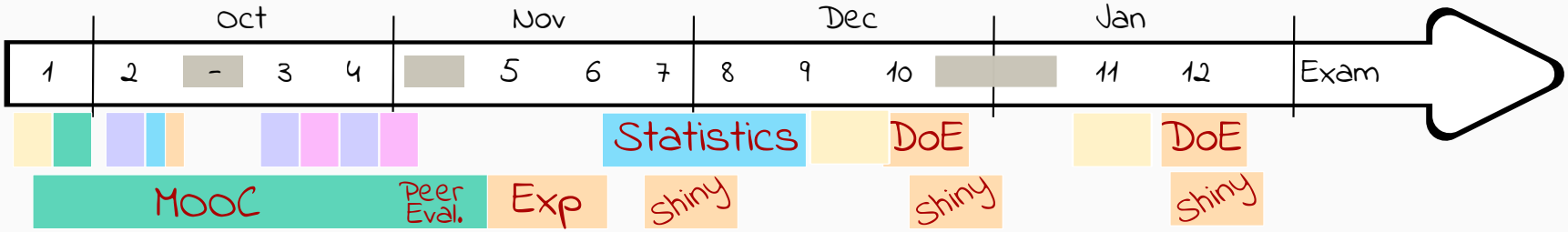
Tentative schedule

12 lectures Many practical homeworks (50%) A final Exam (50%)

1. Epistemology: scientific method and computer science — Presentation of the lecture
2. Visualization and Exploratory Data Analysis — Correlation/causality, spurious correlation
   - **No lecture**
3. Data curation with the tidyverse — Beautiful Viz with ggplot
4. Data management: theory — Data management: tools
   - **Fall Vacations**
5. Measurement and tracing
6. Clearly defining the experiment — Having the right tools to analyze (e.g., C.I)
7. The linear model
8. Multiple testing, ANOVA, and p-hacking
9. Scientific integrity
10. A Bayesian perspective on model selection — Sequential DoE (screening, LHS, D-opt, ...)
11. Ethics (AI and humain, climate change, societal challenges)
    - **Winter vacations**
12. Incremental DoE: Bayesian perspective, Reinforcement Learning, Surrogates (Bandit, GP)
    - **Exam**

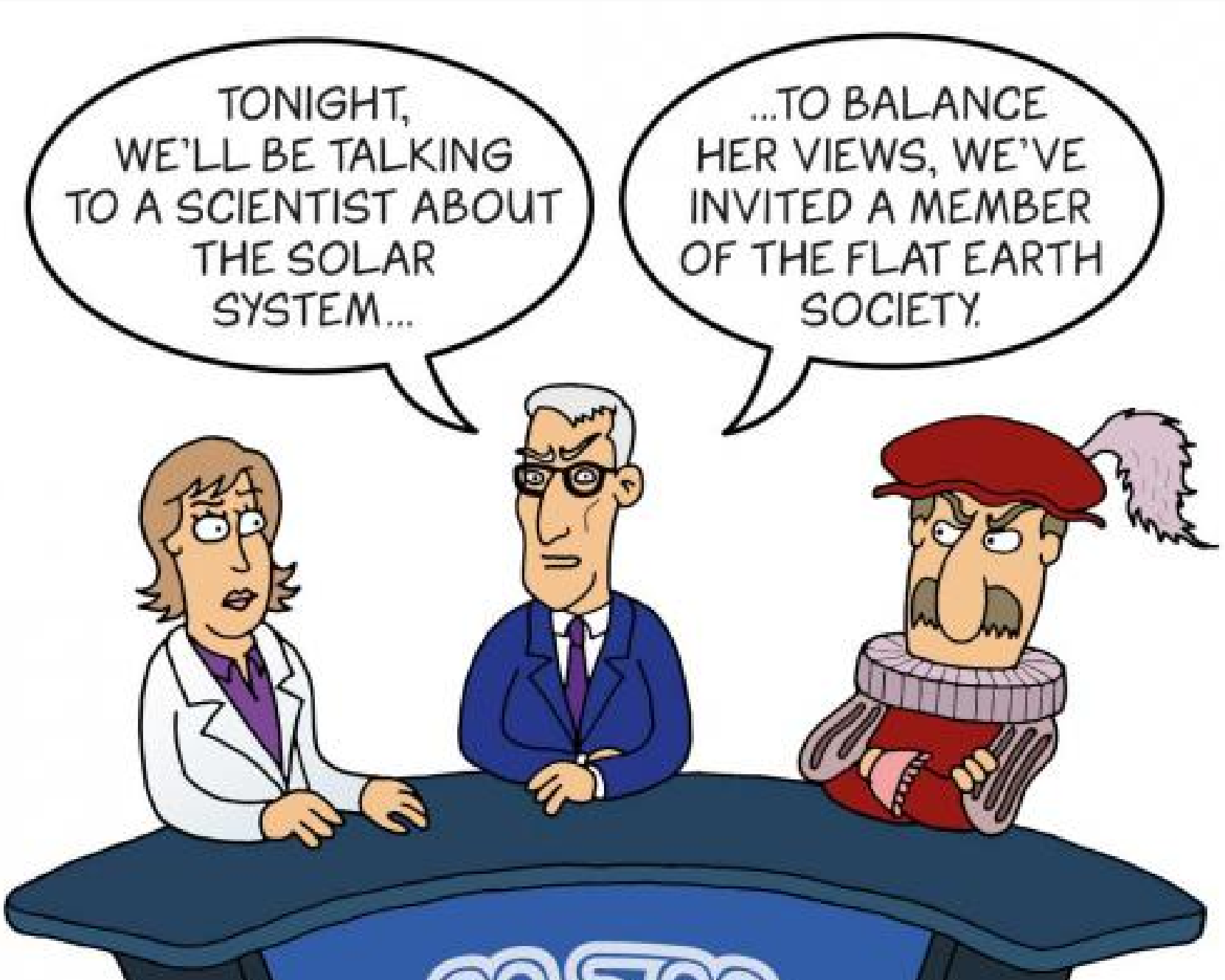
Towards a Rigorous and Ethical Computer Science



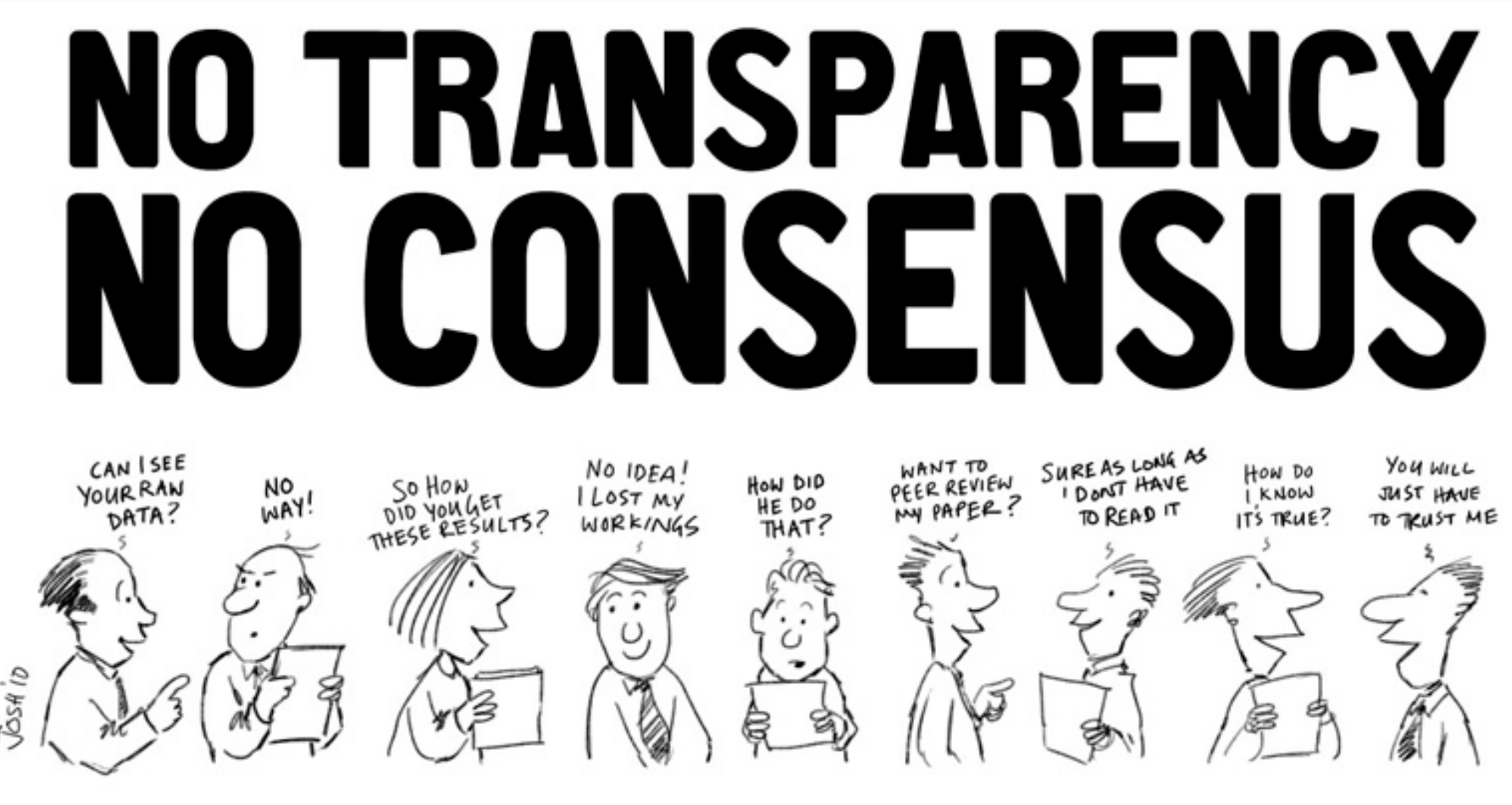


Reproducible Research

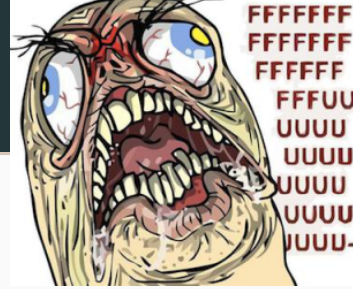
Scientific Consensus



Scientific Consensus



Common Horror Stories 1/3: *What did I do?*



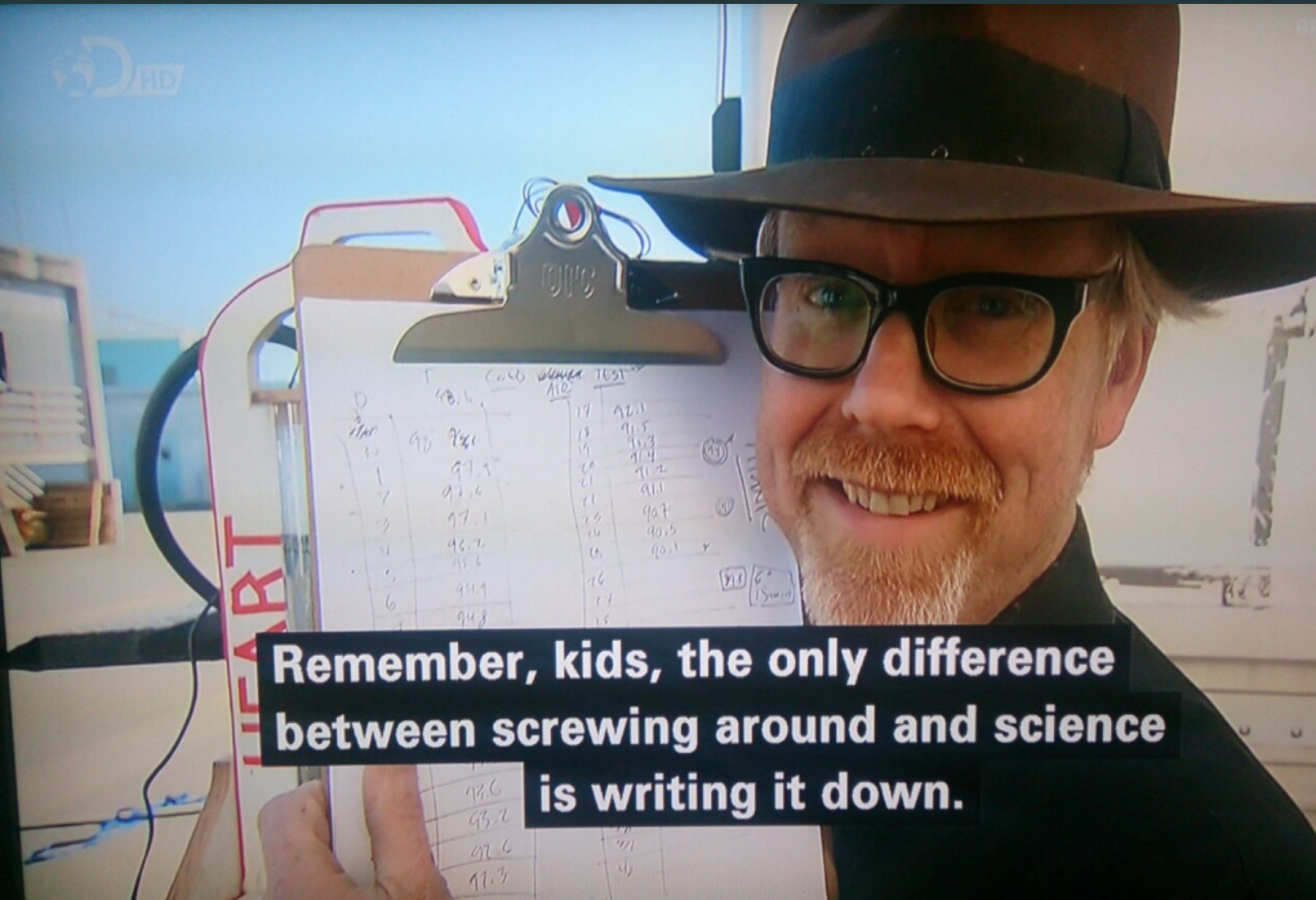
Author

- 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



Common Horror Stories 2/3: *Argh... damned computers*

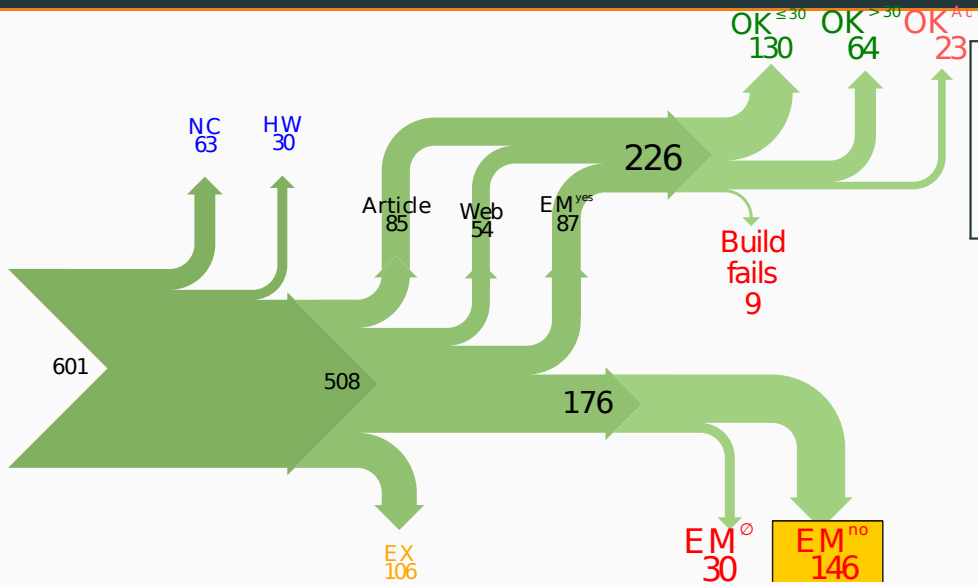
- **Alice**: I got 3.123123 **Bob**: I got segfault
- Damned! It used to work!!! Whenever I upgrade my computer, things break so I try to stay away from this ☹

- Whenever trying the code of my colleague, I had to install Foo but I broke everything and now neither his code nor mine works! ☹

- But hey! Here is my code. It's on GitHub so feel free to play with it! I'm doing open science 😉

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

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



- Versionning Problems

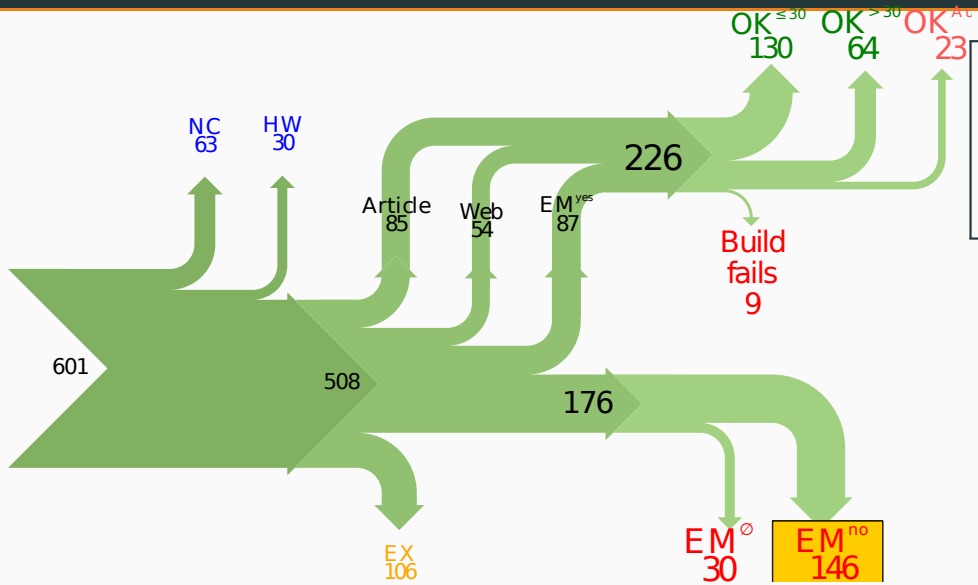
Collberg, Christian *et Al.*, *Measuring Reproducibility in Computer Systems Research*, http://reproducibility.cs.arizona.edu/

- 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

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 3/3: *Please hold on*



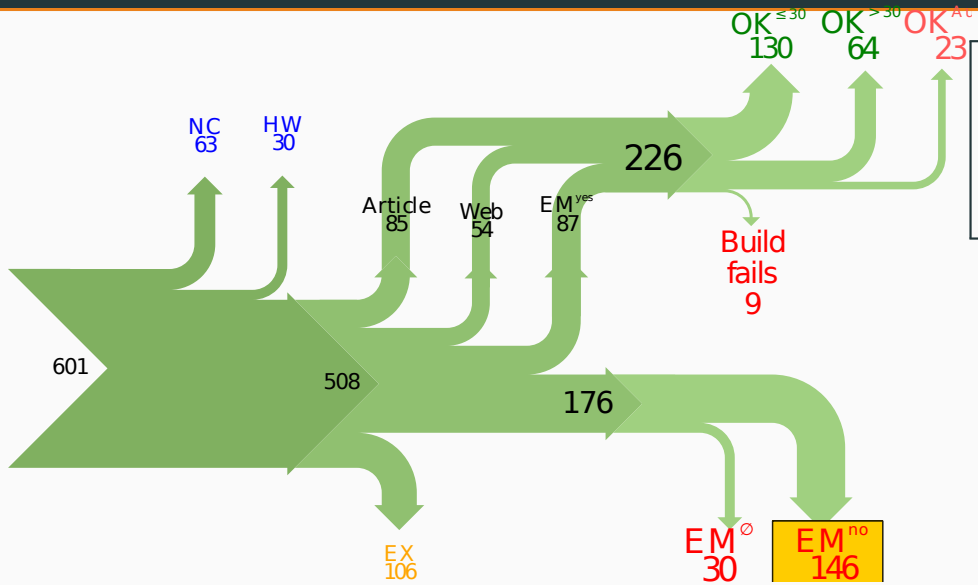
- Versionning Problems
- Bad Backup Practices

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

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 3/3: *Please hold on*



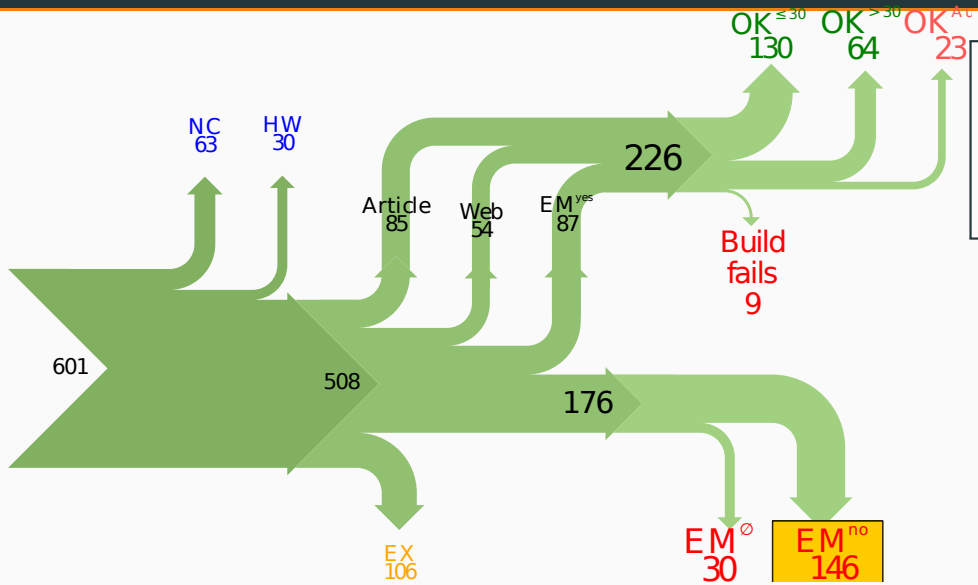
- Versionning Problems
- Bad Backup Practices
- Code Will be Available Soon

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

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 3/3: *Please hold on*



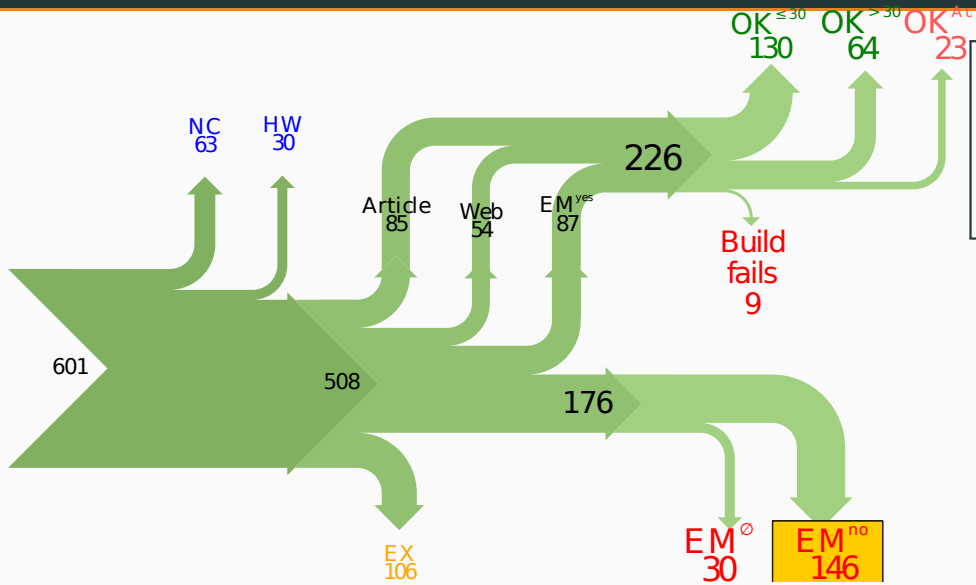
Collberg, Christian *et Al.*, *Measuring Reproducibility in Computer Systems Research*, http://reproducibility.cs.arizona.edu/

- 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 3/3: *Please hold on*



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

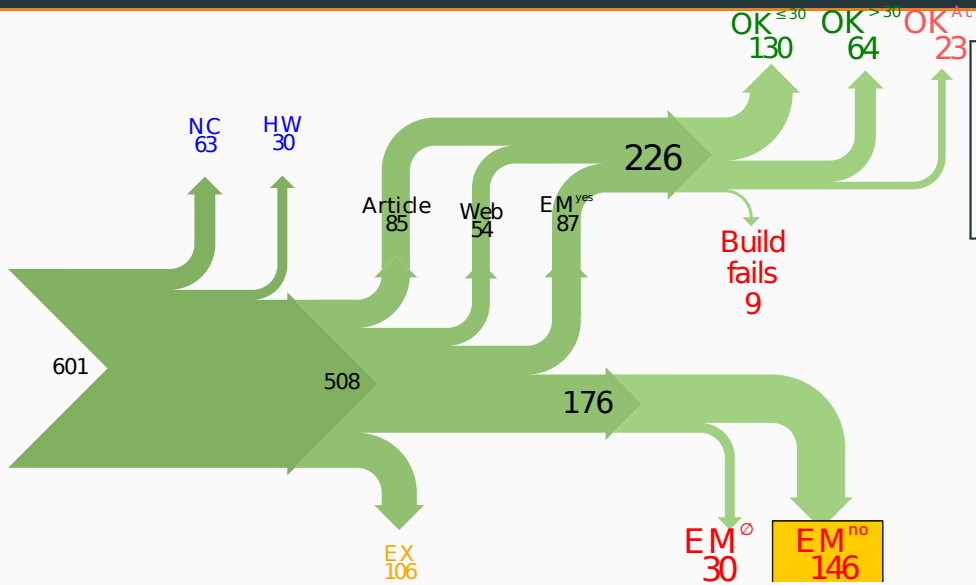
- 8 ACM conferences (ASPLOS'12, CCS'12, OOPSLA'12, OSDI'12, PLDI'12, SIGMOD'12, SOSP'11, VLDB'12) and 5 journals
- EM^{no} = the code cannot be provided
  - Programmer Left

- 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 3/3: *Please hold on*



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

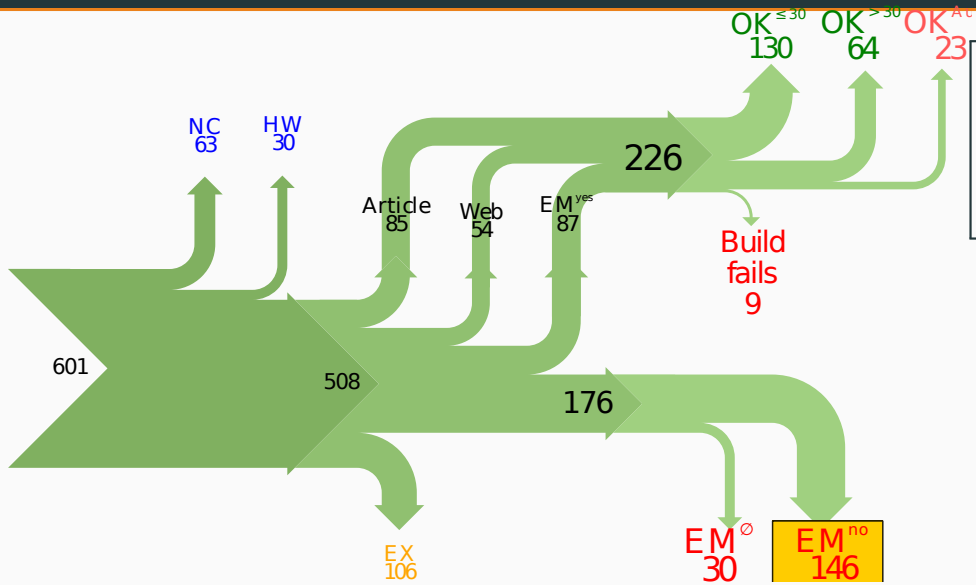
- 8 ACM conferences (ASPLOS'12, CCS'12, OOPSLA'12, OSDI'12, PLDI'12, SIGMOD'12, SOSP'11, VLDB'12) and 5 journals
- EM^{no} = the code cannot be provided
  - Programmer Left
  - Commercial Code

- 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 3/3: *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
  - Proprietary Academic Code

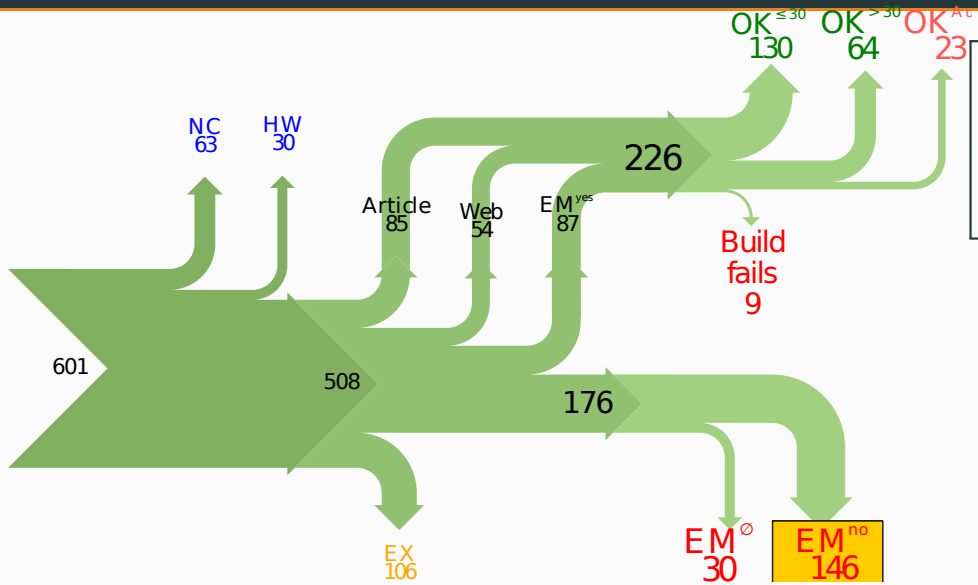
- Versionning Problems
- Bad Backup Practices
- Code Will be Available Soon
- No Intention to Release

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 3/3: *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.

Different Reproducibility Concerns in Modern Science

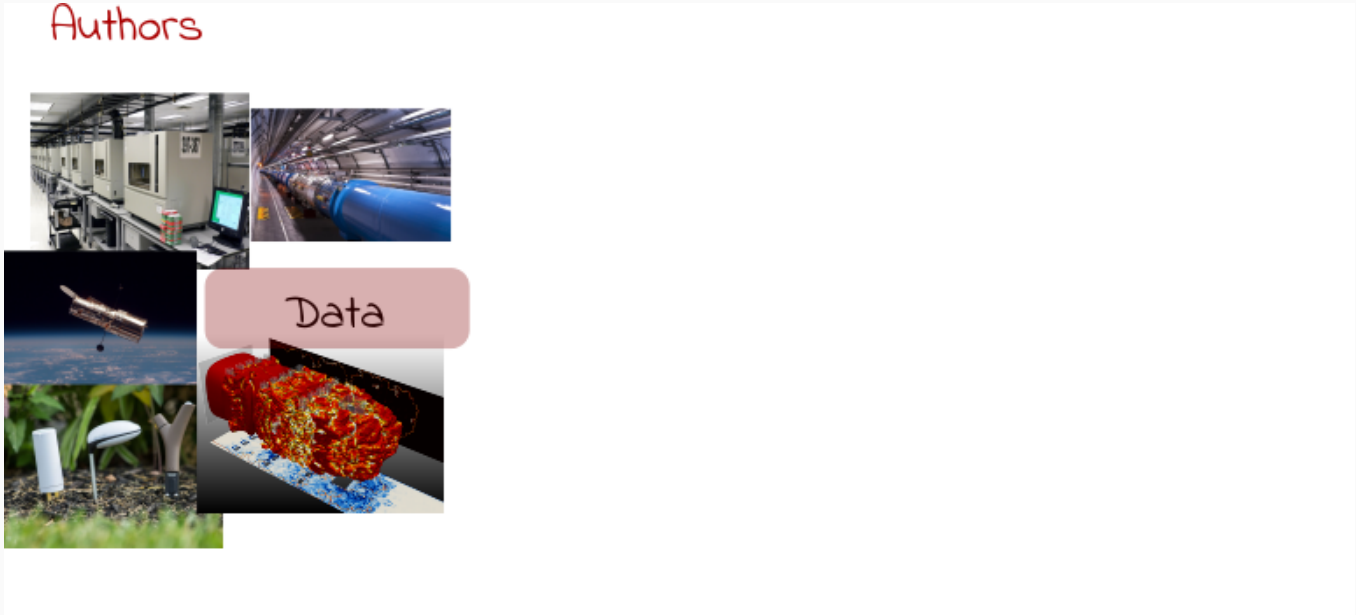
Social Sciences, Oncology, ... methodology, statistics, pre-registration

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

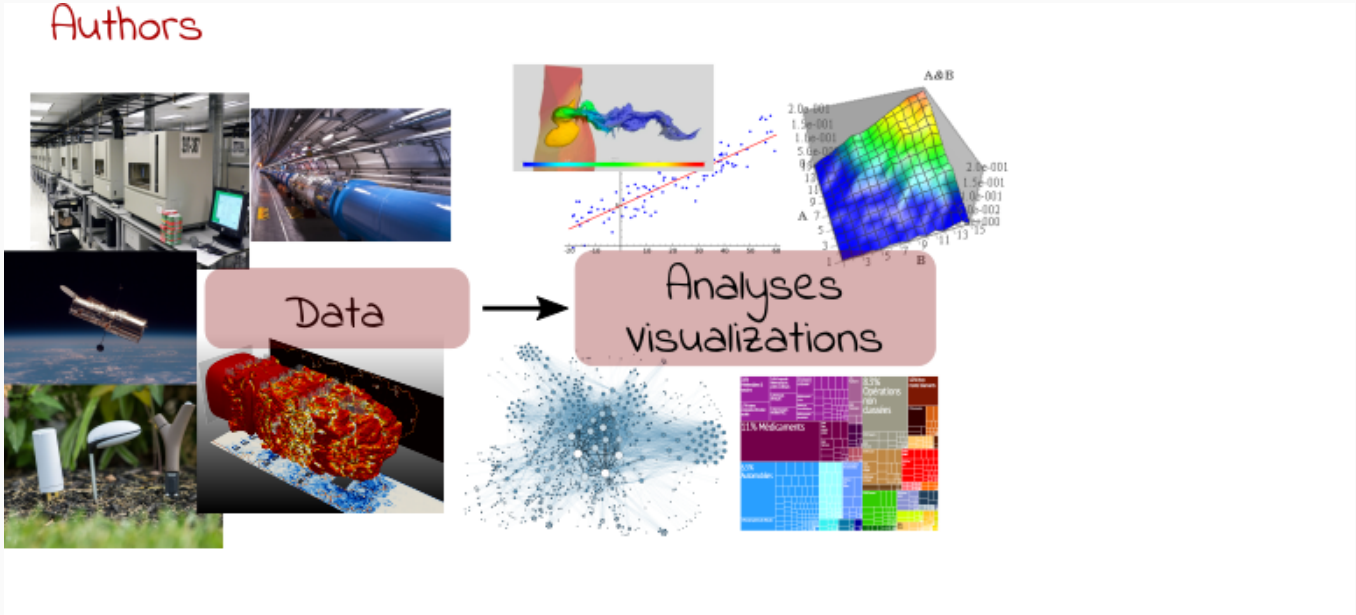
Social Sciences, Oncology, ... methodology, statistics, pre-registration

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

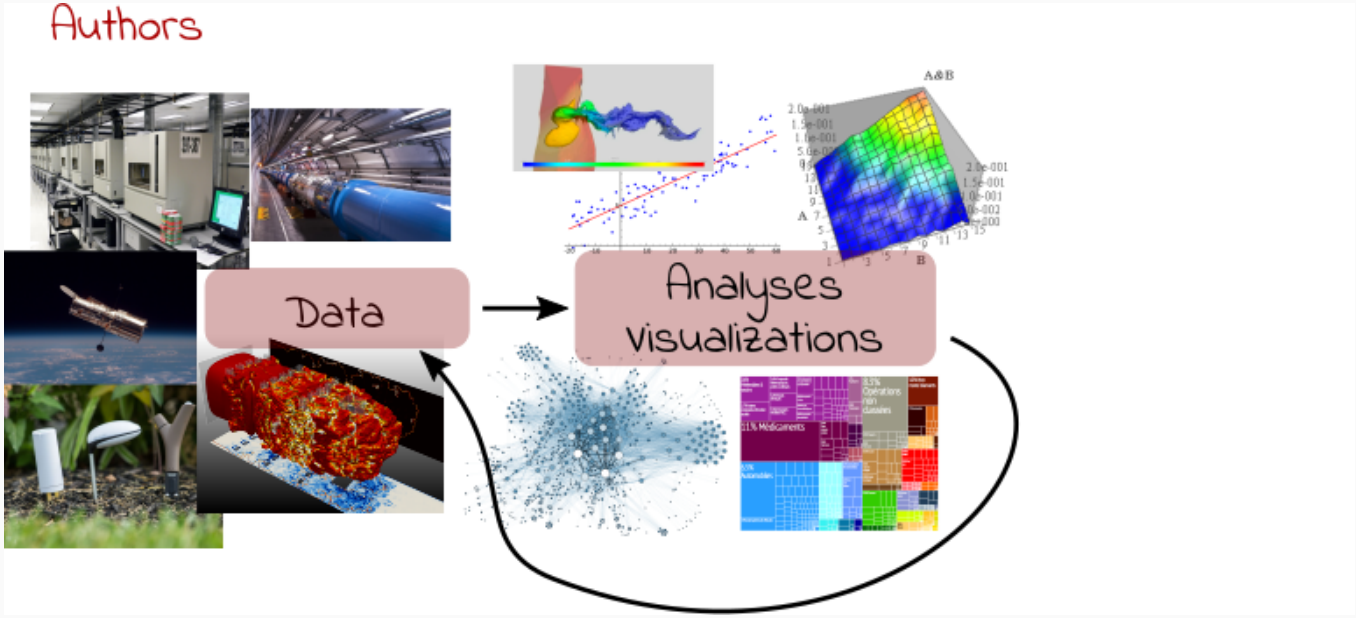
Social Sciences, Oncology, ... methodology, statistics, pre-registration

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

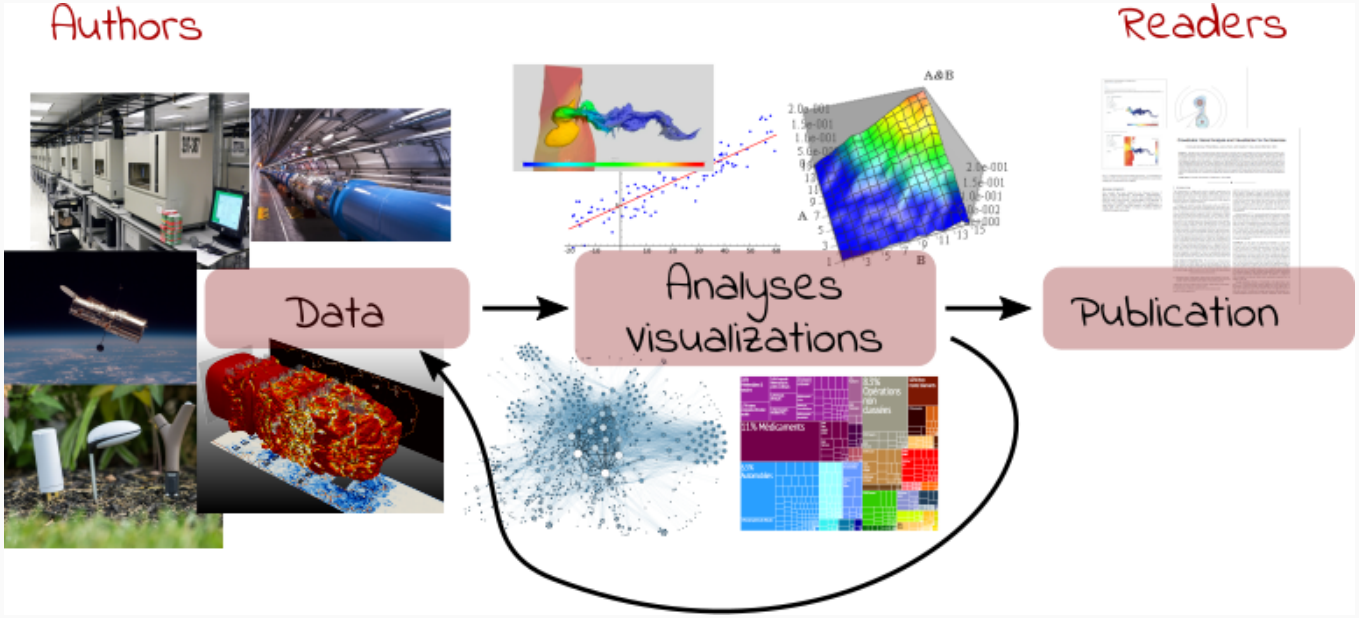
Social Sciences, Oncology, ... methodology, statistics, pre-registration

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

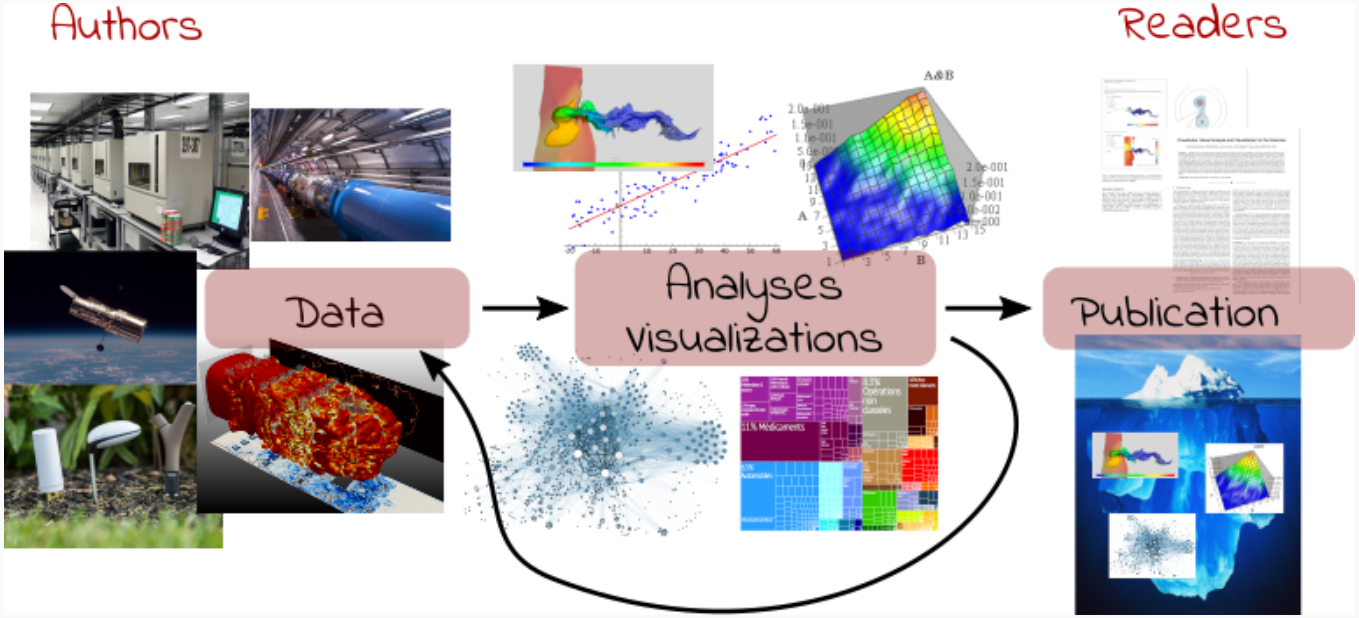
Social Sciences, Oncology, ... methodology, statistics, pre-registration

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

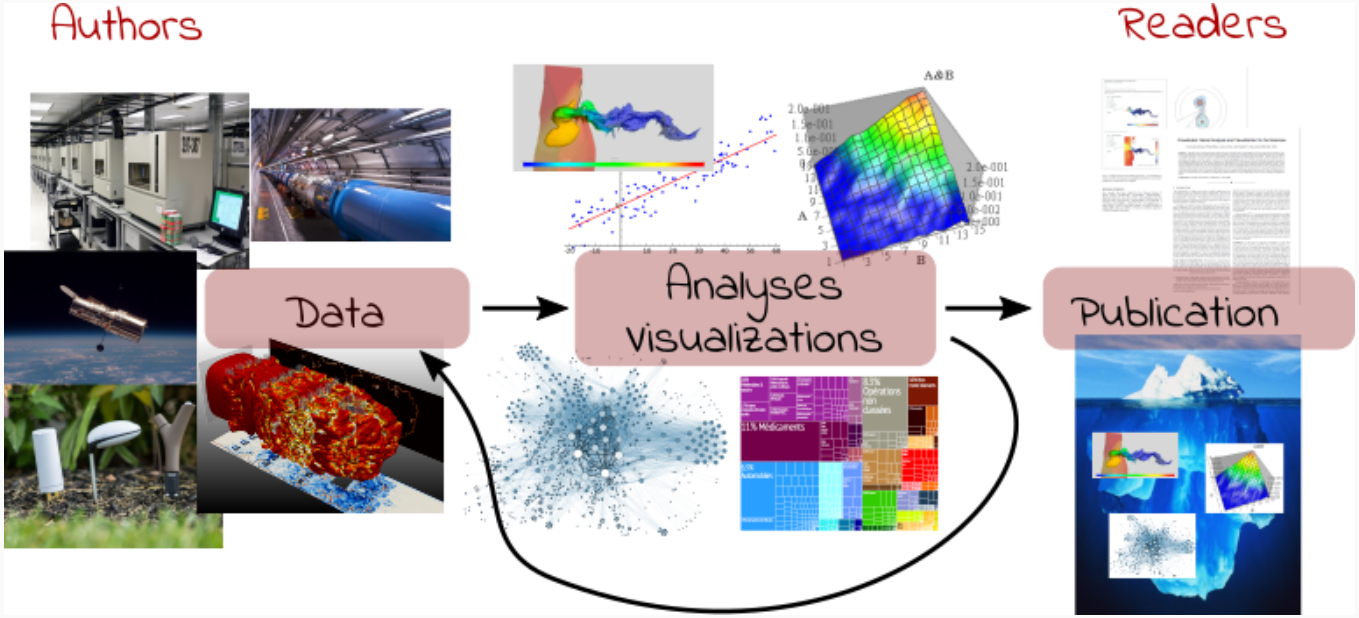
Social Sciences, Oncology, ... methodology, statistics, pre-registration

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

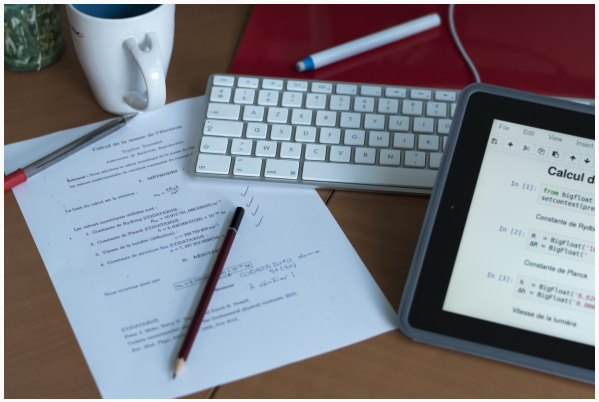
Manifesto: *"I solemnly pledge"* (WSSSPE, Lorena Barba, FAIR)

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, ...

The Reproducible Research MOOC



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

- Konrad Hinsen, Christophe Pouzat
- 3rd Edition: March 2020 – March 2023 (15,000+)
- In French but fully subtitled in English

Module 1 Taking notes (markdown, git)

Module 2 Computational document (Jupyter, Rstudio, Org-Mode)

Data Analysis of the Space Shuttle: Pre-Challenger Prediction of Failure

Module 3 Reproducible analysis

Data analysis : 7 possible subjects and a Peer evaluation

Module 4 Reproducibility Pitfalls (Hell)

The 2nd MOOC "Advanced RR" has been released in ~~2021~~ 2024

- Managing data
- Software environment control
- Scientific workflow

Homeworks

- ☐ Indicate your name on the Pad.
- ☐ Register on the Mattermost.
- ☐ Set up a public github or gitlab project for this lecture.
- ☐ Register to the MOOC
- ☐ Follow modules 1 + 2 of the MOOC with as much exercises as possible
- ☐ Set up a computational document system (e.g., Rstudio or Jupyter)
- ☐ Report the URL of your git project, your mattermost ID on the Pad.
- ☐ Start learning R by reading my short *R crash course for computer scientists*