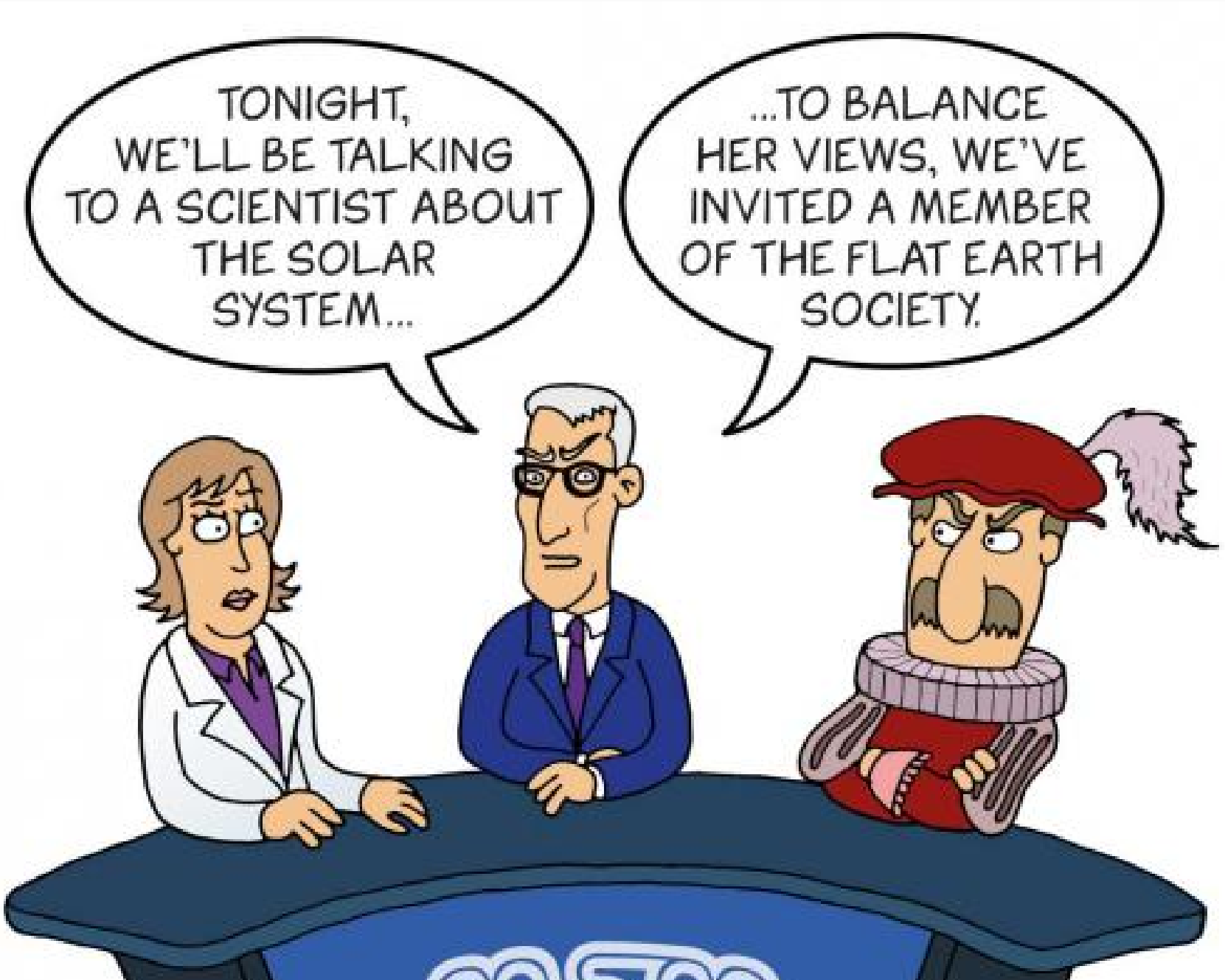
Reproducibility Research and Open Science

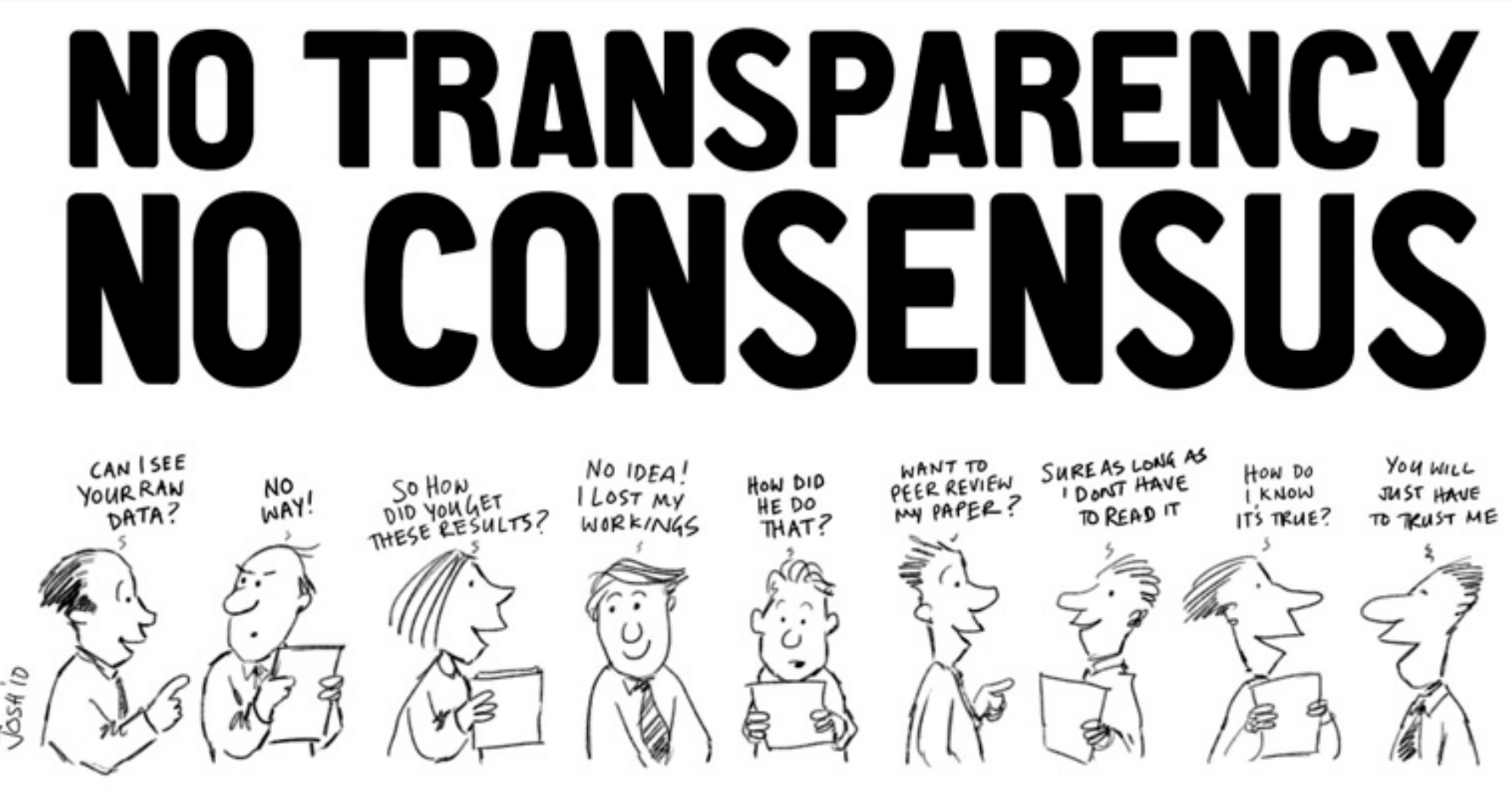
Arnaud Legrand et Konrad Hinsen

INSEE seminar on Open Science
January 206

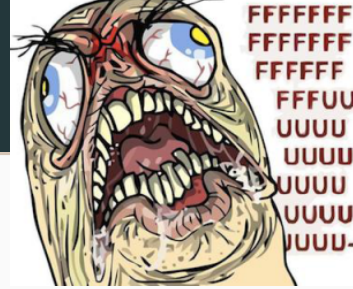
Scientific Consensus



Scientific Consensus



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



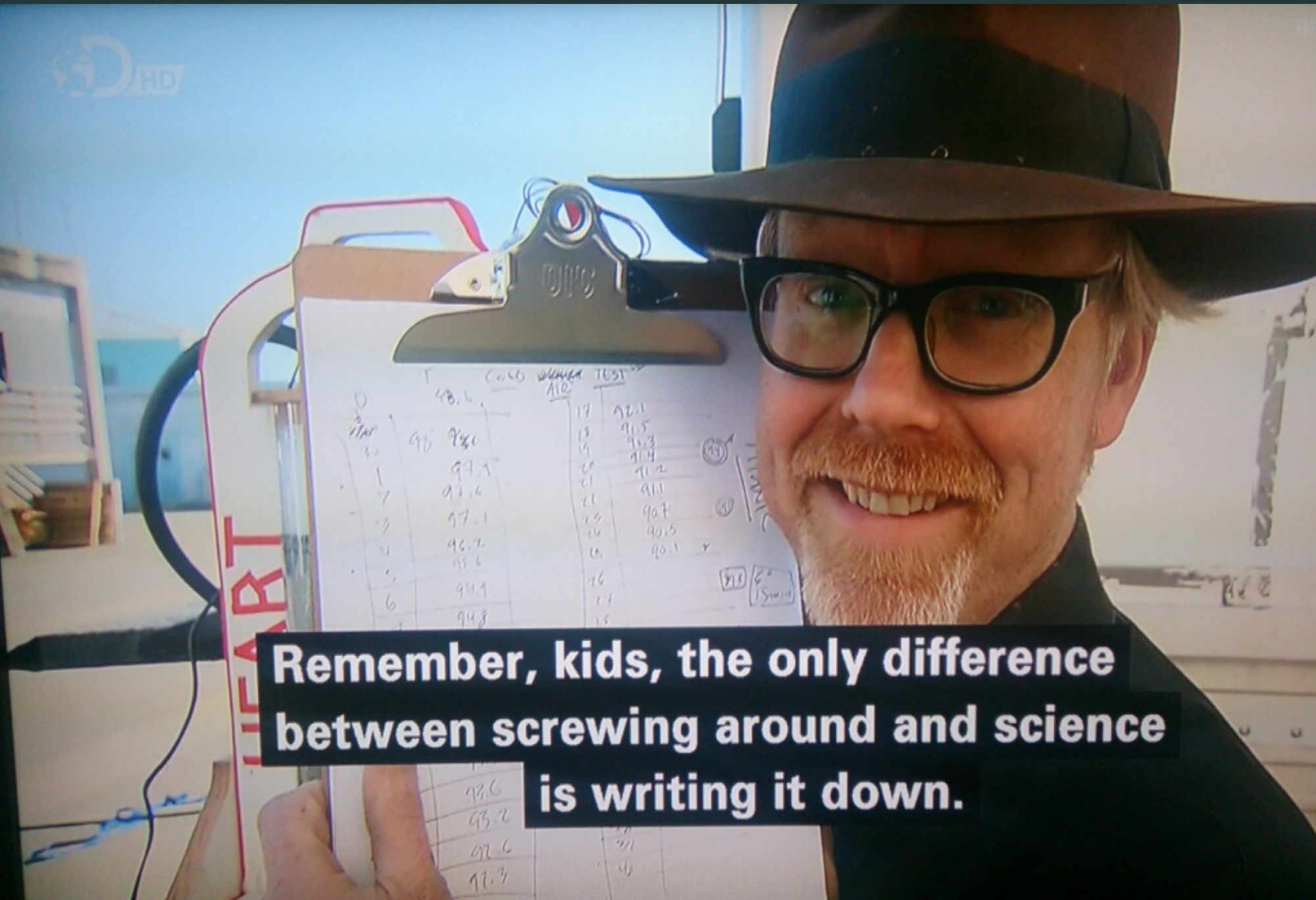
Author

- 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



Computational Documents...

Document initial dans son environnement

jupyter example_pi (autosaved)

File Edit View Insert Cell Kernel Widgets Help Hide Code Python 3

Code CellToolbar

Hide Prompts Hide Code Hide Outputs

```
# Un document computationnel
```

Hide Prompts Hide Code Hide Outputs

```
Mon ordinateur m'indique que $\pi$ vaut *approximativement*
```

In [1]:

Hide Prompts Hide Code Hide Outputs

```
from math import *
print(pi)
```

```
3.141592653589793
```

Hide Prompts Hide Code Hide Outputs

```
Mais calculé avec la __méthode__ des [aiguilles de Buffon]
(https://fr.wikipedia.org/wiki/Aiguille_de_Buffon), on obtiendrait comme
__approximation__ :
```

In [2]:

Hide Prompts Hide Code Hide Outputs

```
import numpy as np
N = 1000000
x = np.random.uniform(size=N, low=0, high=1)
theta = np.random.uniform(size=N, low=0, high=pi/2)
2/(sum((x+np.sin(theta))>1)/N)
```

Out[2]: 3.1437198694098765

Hide Prompts Hide Code Hide Outputs

```
On peut inclure des formules mathématiques comme $\frac{1}{\sigma
\sqrt{2\pi}} \exp \left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des
_dessins qui n'ont rien à voir_ avec $\pi$ (si ce n'est une constante de
normalisation... ☺).
```

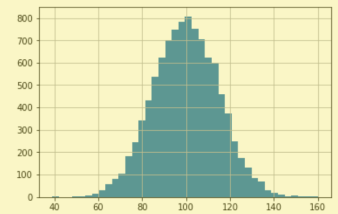
In [3]:

Hide Prompts Hide Code Hide Outputs

```
%matplotlib inline
import matplotlib.pyplot as plt

mu, sigma = 100, 15
x = mu + sigma*np.random.randn(10000)

plt.hist(x,40)
plt.grid(True)
plt.show()
```



Export

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut *approximativement*

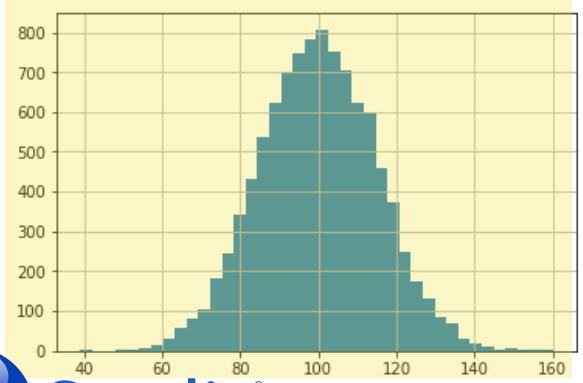
```
3.141592653589793
```

Mais calculé avec la **méthode** des aiguilles de Buffon, on obtiendrait comme **approximation** :

```
import numpy as np
N = 1000000
x = np.random.uniform(size=N, low=0, high=1)
theta = np.random.uniform(size=N, low=0, high=pi/2)
2/(sum((x+np.sin(theta))>1)/N)
```

```
3.1437198694098765
```

On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des *dessins qui n'ont rien à voir* avec π (si ce n'est une constante de normalisation... ☺).



... and Laboratory Notebooks

Subject: Fin des pièces jointes dans Mattermost (Inria) (Nov. 2025)

> Mattermost est le principal outil de communication et d'échanges dans certaines équipes. Nous avons un canal par projet, un canal avec chaque doctorant, etc. Ce ne sont pas de simples chats "informels"... ce sont de vraiment outils de travail dans la durée !

> Comme XXX, nous aussi l'utilisons beaucoup pour échanger des fichiers, des PDF de papiers, faire la biblio

- *Nous avons positionné Mattermost comme un service de discussion instantanée (ou asynchrone, mais pour des messages à court terme).*
- *Sa finalité ([..]) n'est ni la ~~gestion documentaire~~, ni ~~l'archivage de documents~~, ni le ~~suivi des expérimentations~~ (au sens carnet de laboratoire), même si, finalement, on peut l'utiliser de cette manière.*
- *Si des documents importants pour votre activité au sein d'Inria sont stockés dans Mattermost, et conservés seulement ici, c'est un risque pour Inria (et pour vous): que se passe-t-il en cas de départ des agents? remplacement du service par un autre ? évolution d'une équipe de recherche ? Qui est le propriétaire d'un document partagé ici ?*

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

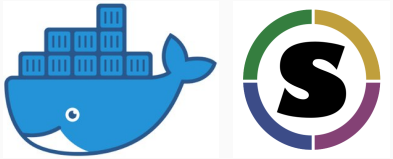
- Hey! Here is my code. It's on GitHub so feel free to play with it! I'm doing open science 🙂
  - **Alice**: I got 3.123123 **Bob**: I got segfault **Cal**: I got 3.123125
- Damned! It used to work!!! Whenever I upgrade my computer, things break so I try to stay away from this ☹
- Whenever trying the 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

The bad



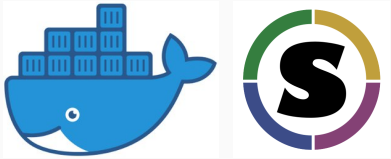
The uggly

Automatic tracking

Containers and Package Managers

The good

The bad



The uggly

Automatic tracking

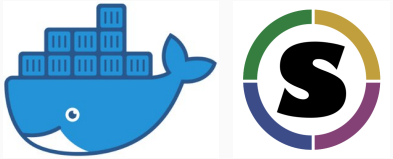
Containers

- 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

The bad



The uggly

Automatic tracking

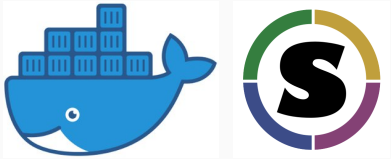
Containers

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

Containers and Package Managers

The good

The bad



The uggly

Automatic tracking

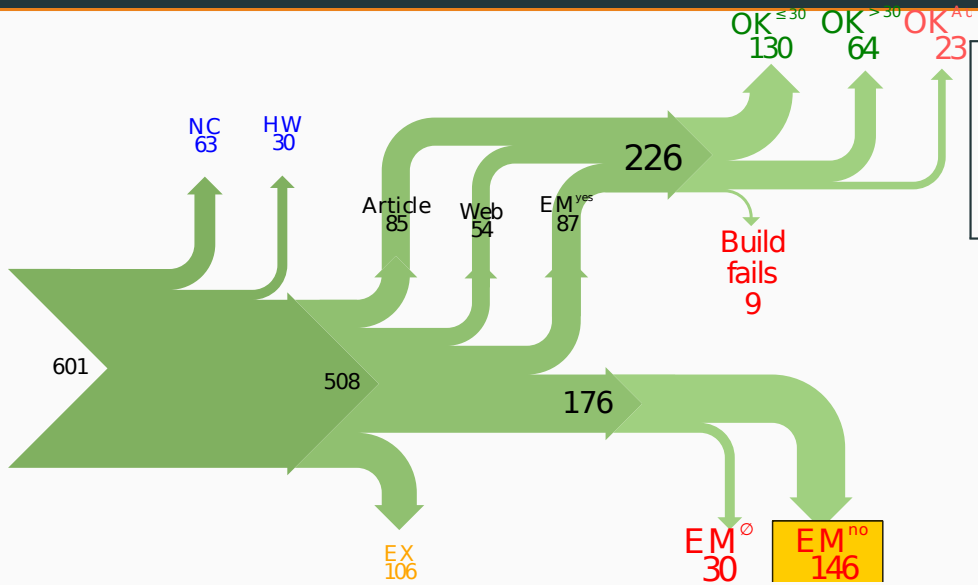
Containers

- 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` or through containers
  - 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/ 2013

- 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
  - Programmer Left
  - No Intention to Release
  - Commercial Code
  - Proprietary Academic Code
  - Research vs. Sharing

The good news is that I was able to find some code. [..] 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.

⟨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 [...]

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.

Changing Research/Publishing Practices

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

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

Learn and Teach using online resources like Software Carpentry, The Turing Way, ...

Reforming reviewing/publishing practices through incentives

Artifact evaluation and ACM badges

Major conferences

- Supercomputing: Artifact Description (AD) mandatory, Artifact Evaluation (AE) still optional, Double blind vs. RR
- NeurIPS, ICLR: open reviews, reproducibility challenge
- ACM SIGMOD 2015-2019, Most Reproducible Paper Award...

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

INTERNET ARCHIVE

Software Archive Software Heritage Collect/Preserve/Share

Plan for disaster with `git` and `git-annex` (not `git LFS`!)

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

Different Reproducibility Concerns in Modern Science

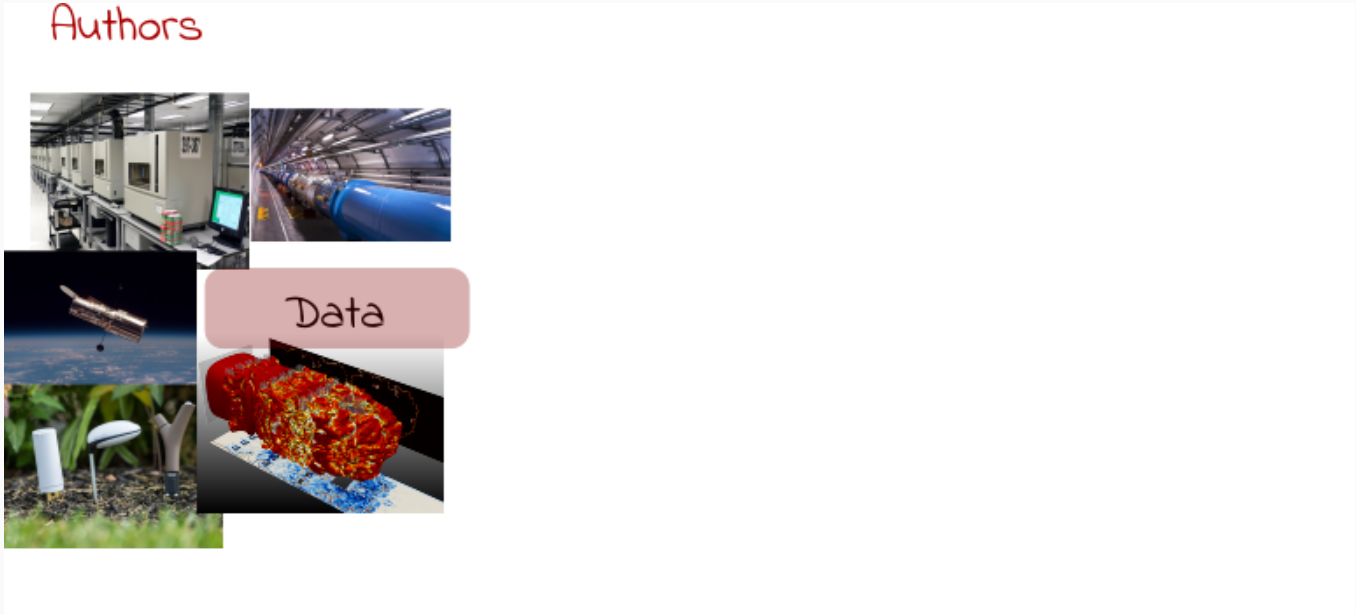
Social Sciences, Oncology, ... methodology, statistics, pre-registration

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

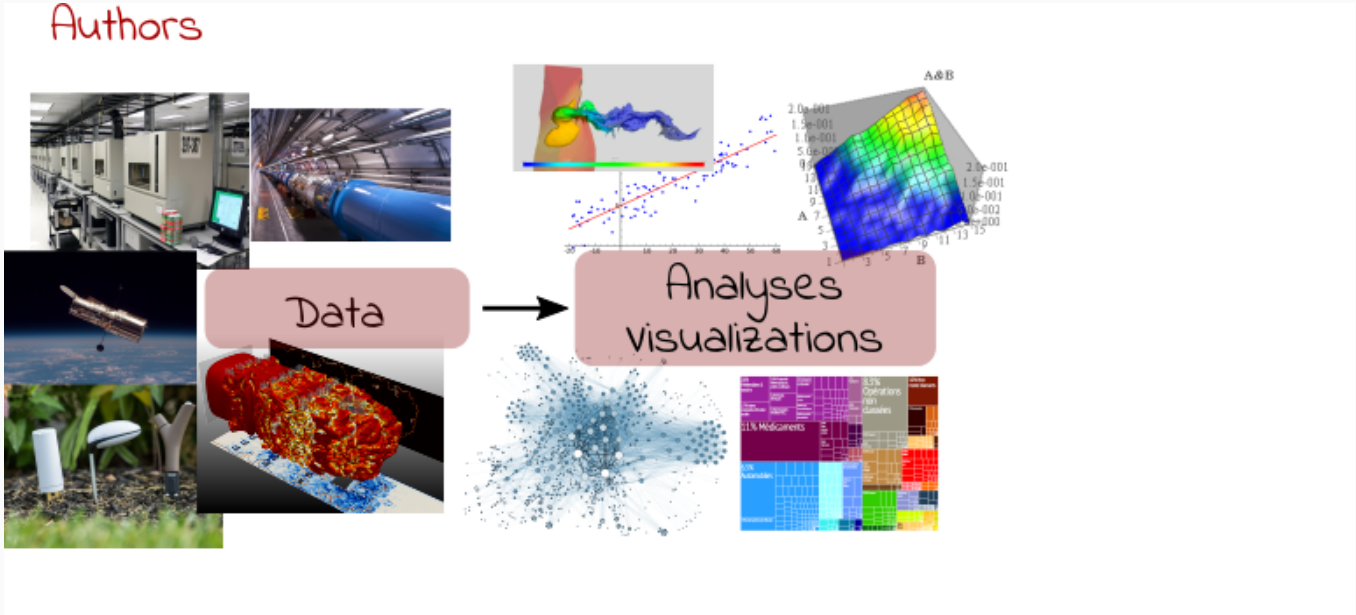
Social Sciences, Oncology, ... methodology, statistics, pre-registration

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

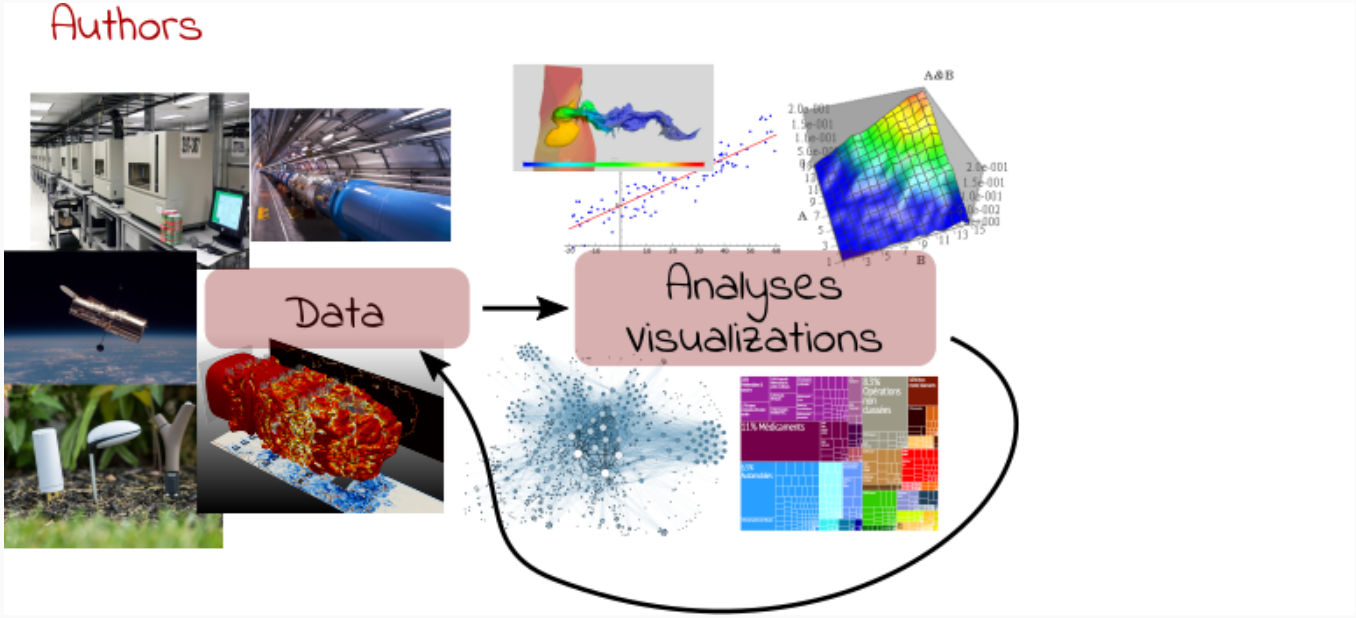
Social Sciences, Oncology, ... methodology, statistics, pre-registration

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

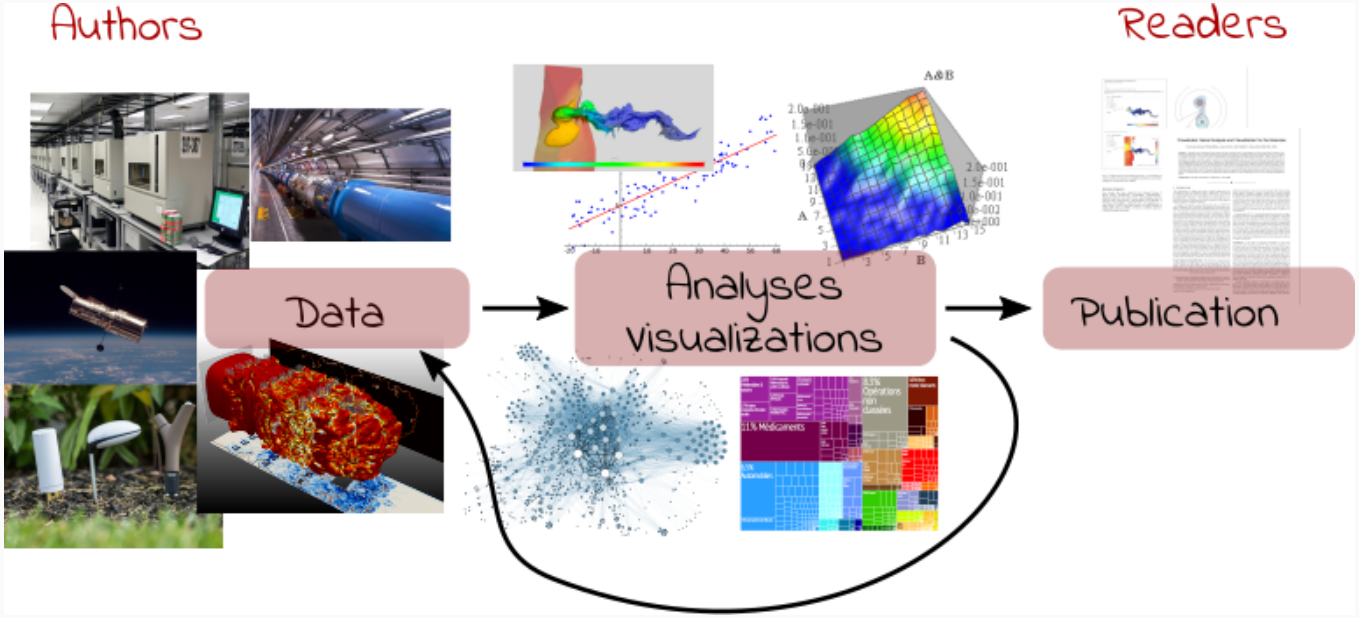
Social Sciences, Oncology, ... methodology, statistics, pre-registration

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

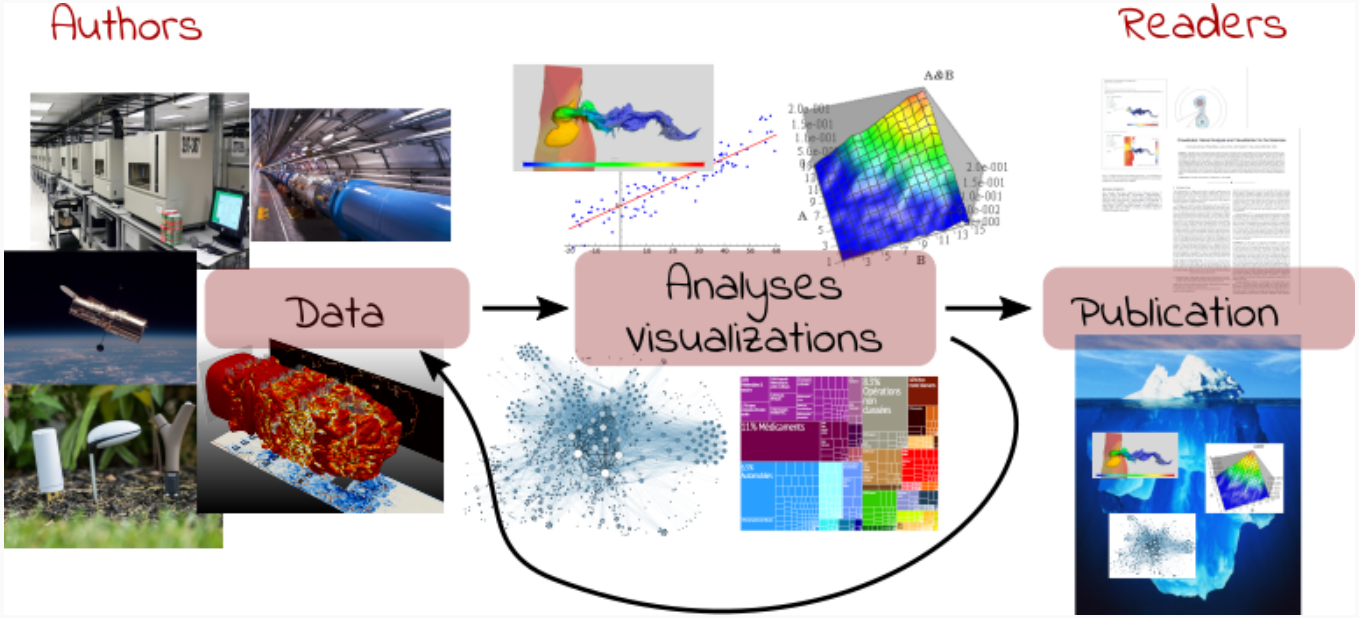
Social Sciences, Oncology, ... methodology, statistics, pre-registration

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical issues

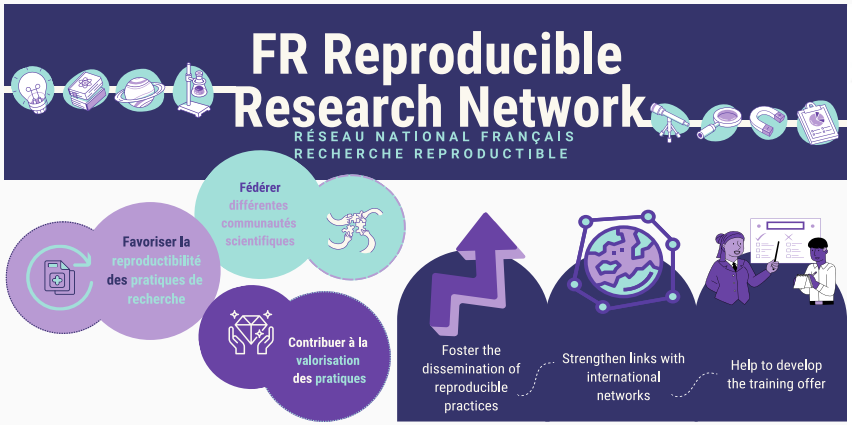
Artificial Intelligence most of the above

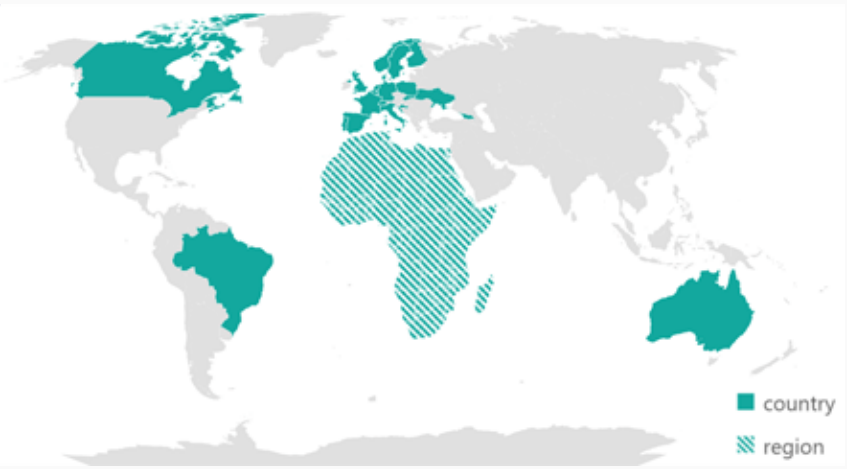
The processing steps between raw observations and findings have gotten increasingly numerous and complex



Reproducible Research = Bridging the Gap by working Transparently

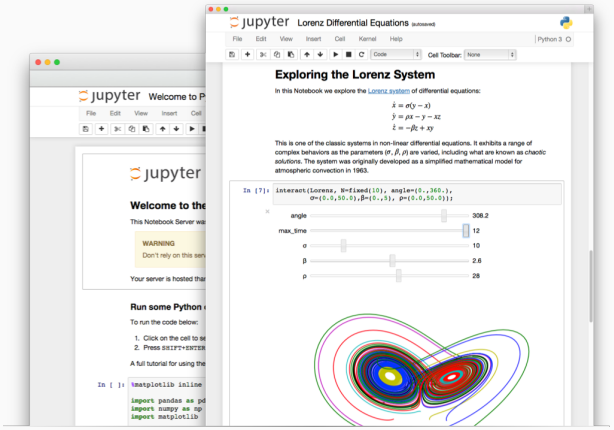
Good research requires time, resources, and friends





Good research requires time, resources, and friends

Computation provenance: notebooks and workflows

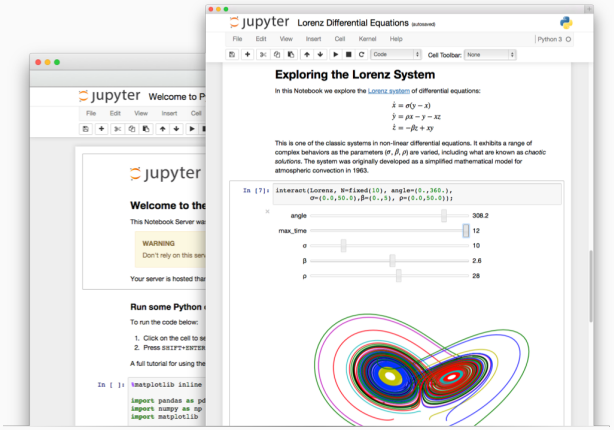


MOOC RR 1: Methodological principles for a transparent science

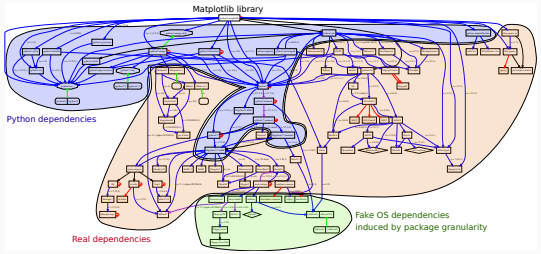
3rd Edition: March 2020 – ... (25,000+)

Good research requires time, resources, and friends

Computation provenance: notebooks



Software environments



Sharing and Archiving

arXiv.org

HAL archives-ouvertes.fr

zenodo

Software Heritage

GNU Make

snakemake

MOOC RR 1: Methodological principles for a transparent science

3rd Edition: March 2020 – … (25,000+)

MOOC RR 2: Practices and tools for managing computations and data

3rd Edition: May 2026 – … (5,000)