Reproducibility Crisis and Open Science

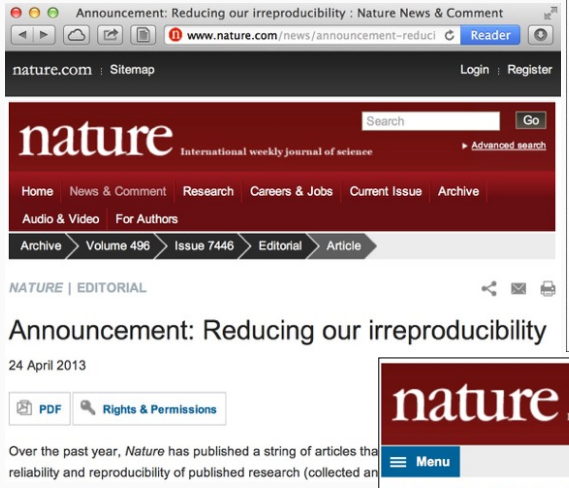
Arnaud Legrand

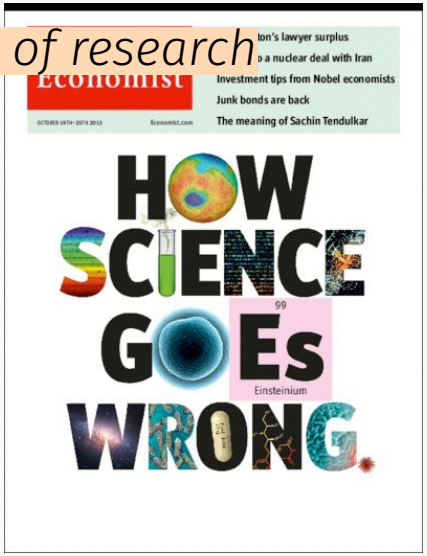
Sciences de l'information géographique reproductibles

June 2021

Public Evidence for a Lack of Reproducibility

- J.P. Ioannidis. *Why Most Published Research Findings Are False* PLoS Med. 2005.
- *Lies, Damned Lies, and Medical Science*, The Atlantic. Nov, 2010
- *Reproducibility: A tragedy of errors*, Nature, Feb 2016.
- Steen RG, *Retractions in the scientific literature: is the incidence of research fraud increasing?*, J. Med. Ethics 37, 2011





Courtesy V. Stodden, SC, 2015

Newsworthy Stories about Scientific Misconduct

Dong-Pyou Han Assistant professor, Biomedical sciences, Iowa State University, 2013

Falsified blood results to make it appear as though a vaccine exhibited anti-HIV activity

- Han and his team received ≈ $19 million from NIH
- 1 retracted publication and resignation of university. Sentenced in 2015 to 57 months imprisonment for fabricating and falsifying data in HIV vaccine trials. $7.2 million!

Newsworthy Stories about Scientific Misconduct

Dong-Pyou Han Assistant professor, Biomedical sciences, Iowa State University, 2013

Falsified blood results to make it appear as though a vaccine exhibited anti-HIV activity

- Han and his team received ≈ $19 million from NIH
- 1 retracted publication and resignation of university. Sentenced in 2015 to 57 months imprisonment for fabricating and falsifying data in HIV vaccine trials. $7.2 million!

Diederik Stapel Professor, Social Psychology, Univ. Tilburg, 2011

I failed as a scientist. I adapted research data and fabricated research. Not once, but several times, not for a short period, but over a longer period of time. [..] I am aware of the suffering and sorrow that I caused to my colleagues... I did not withstand the pressure to score, to publish, the pressure to get better in time. I wanted too much, too fast. In a system where there are few checks and balances, where people work alone, I took the wrong turn.

58 retracted publications

Newsworthy Stories about Scientific Misconduct

Dong-Pyou Han Assistant professor, Biomedical sciences, Iowa State University, 2013

Falsified blood results to make it appear as though a vaccine exhibited anti-HIV activity

- Han and his team received ≈ $19 million from NIH
- 1 retracted publication and resignation of university. Sentenced in 2015 to 57 months imprisonment for fabricating and falsifying data in HIV vaccine trials. $7.2 million!

Diederik Stapel Professor, Social Psychology, Univ. Tilburg, 2011

I failed as a scientist. I adapted research data and fabricated research. Not once, but several times, not for a short period, but over a longer period of time. [..] I am aware of the suffering and sorrow that I caused to my colleagues... I did not withstand the pressure to score, to publish, the pressure to get better in time. I wanted too much, too fast. In a system where there are few checks and balances, where people work alone, I took the wrong turn.

58 retracted publications

Brian Wansink Professor, Psychological Nutrition, Cornell, 2016

I gave her a data set of a self-funded, failed study which had null results. I said "This cost us a lot of time and our own money to collect. There's got to be something here we can salvage because it's a cool (rich & unique) data set." I told her what the analyses should be. [..] Every day she came back with puzzling new results, and every day we would scratch our heads, ask "Why," and come up with another way to reanalyze the data with yet another set of plausible hypotheses

17 retracted publications

Scientific Misconduct? What are the consequences ?

Reinhart and Rogoff Professors of Economics at Harvard

> *gross debt [..] exceeding 90 percent of the economy has a significant negative effect on economic growth* – *Growth in a Time of Debt (2010)*

> *While using RR's working spreadsheet, we identified coding errors, selective exclusion of available data, and unconventional weighting of summary statistics.* – *2013: Herndon, Ash and Pollin*

> *For 3 years, austerity was not presented as an option but as a necessity.* – *2013: Paul_Krugman*

At least, a scientific debate has been possible.

Scientific Misconduct? What are the consequences ?

Reinhart and Rogoff Professors of Economics at Harvard

> *gross debt [..] exceeding 90 percent of the economy has a significant negative effect on economic growth* – *Growth in a Time of Debt (2010)*
>
> *While using RR's working spreadsheet, we identified coding errors, selective exclusion of available data, and unconventional weighting of summary statistics.* – *2013: Herndon, Ash and Pollin*
>
> *For 3 years, austerity was not presented as an option but as a necessity.* – *2013: Paul_Krugman*

At least, a scientific debate has been possible.

Bad science is **deleterious**

- It is used to backup stupid politics, it affects people's life, ...
- It blurs the frontier between scientists and crooks

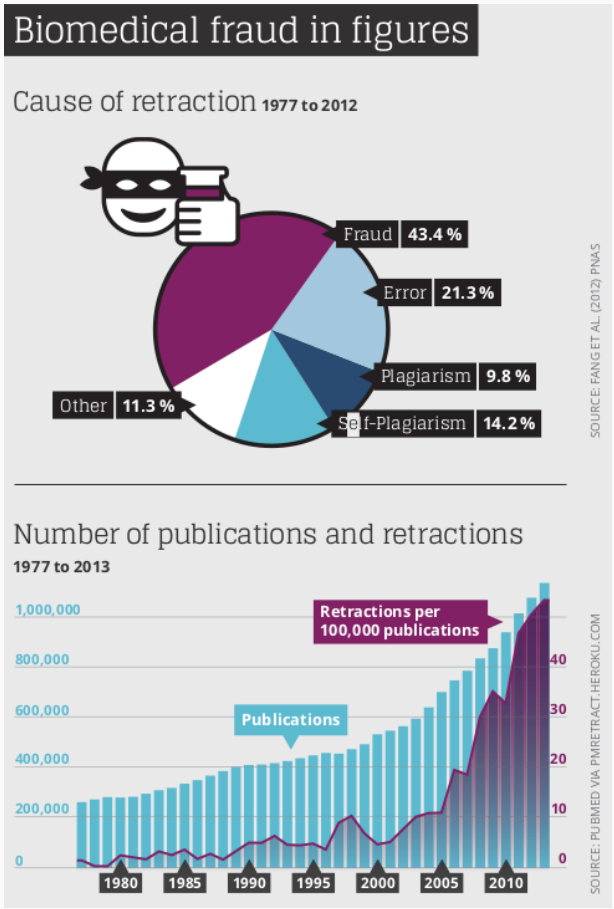
Media attention inflates **conspiracy** opinions ☹

- *Scientific result are worthless.*
- *Scientists can't even agree with each others on economy/climate/vaccine/5G/...*
- *Stop the scientific dictatorship/lobby!*

A Credibility Crisis?

How so? Why now? Why is this important? What can we do about it?

The Battle against Scientific Fraud in the CNRS International Magazine



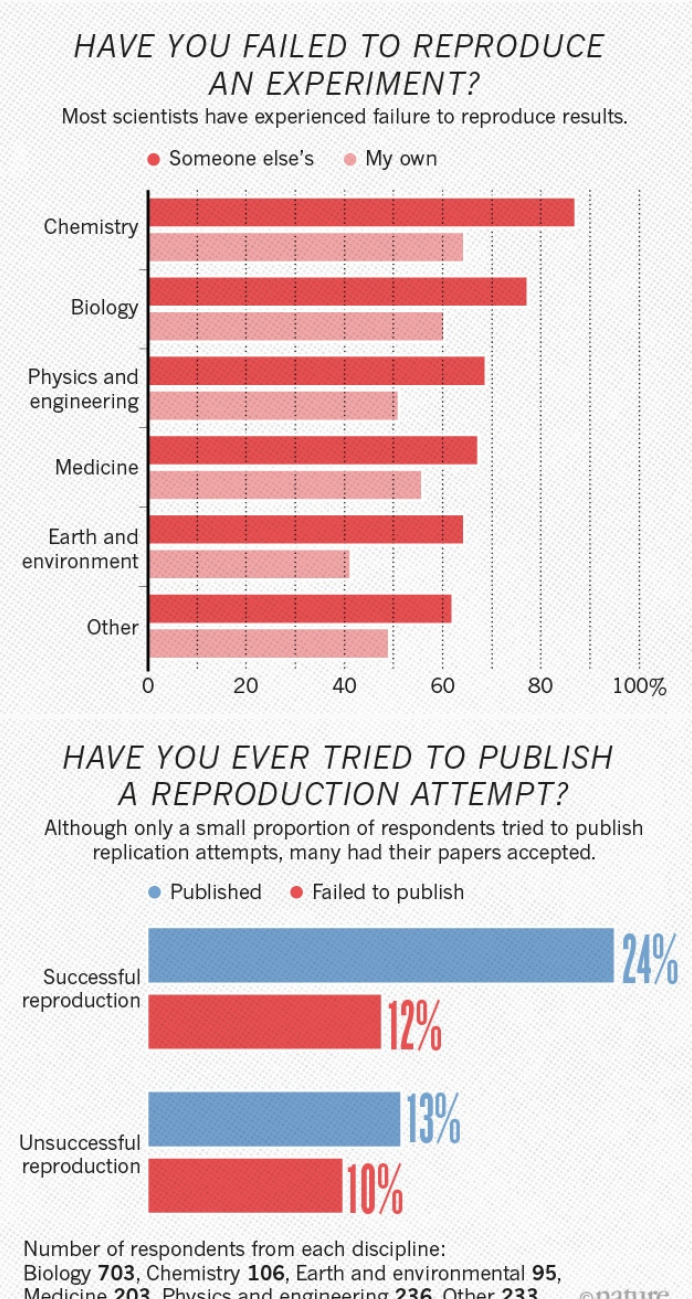



Galileo (data fabrication), **Ptolemy** (plagiarism), **Mendel** (data enhancement), **Pasteur** (rigorous but hid failures), ...

Scientific misconduct is obviously wrong but it's not new!

- Every domain has its black sheep
- The publish or perish pressure is a pain

A Reproducibility Crisis?



1,500 scientists lift the lid on reproducibility, Nature, May 2016

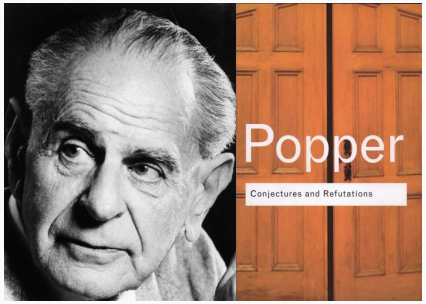
Social causes

- Fraud, conflict of interest (pharmaceutic, ...)
- No incentive to reproduce/check our own work (afap), nor the work of others (big results!), nor to allow others to check (competition)
- Peer review does not scale: 1+ million articles per year!

Methodological or technical causes

- The many biases (apophenia, confirmation, hindsight, experimenter, ...): bad designs
- Selective reporting, weak analysis (statistics, data manipulation mistakes, computational errors)
- Lack of information, code/raw data unavailable

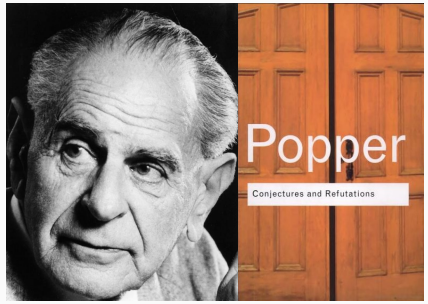
Reproducibility of experimental results: the *hallmark of science*



1934: **Karl Popper** puts the notions of falsifiability and crucial experiment as the hallmark of science

- If no experiment can be set up to disprove your theory, it is not science
- Good experiments discriminate good theories from bad ones
- Non-reproducible single occurrences are of no significance to science

Reproducibility of experimental results: the *hallmark of science*



1934: **Karl Popper** puts the notions of falsifiability and crucial experiment as the hallmark of science

- If no experiment can be set up to disprove your theory, it is not science
- Good experiments discriminate good theories from bad ones
- Non-reproducible single occurrences are of no significance to science

An ideal rather than the norm

Popper's proposal works well for Physics from the 18th century but is not so simple for many other domains:

- Theory of evolution
- Spotting a SuperNova
- Particle Physics (a single LHC)
- Biology (every animal does not behave in the same way)
- Anthropology (impact on people from a remote culture)

Reproducibility: a *core value of science*

1. Universality: Science aims for objective findings, accessible to anyone

   Reproducibility acts as a Universality/Robustness control

2. Incremental: We build on each others work but everybody makes mistakes

   Methods, biases, ... How to discriminate sound ~~theories~~ experiments from bad ones? 😉

   Reproducibility acts as a Quality control

Reproducibility: a *core value of science*

1. Universality: Science aims for objective findings, accessible to anyone

   Reproducibility acts as a Universality/Robustness control

2. Incremental: We build on each others work but everybody makes mistakes

   Methods, biases, ... How to discriminate sound ~~theories~~ experiments from bad ones? 😉

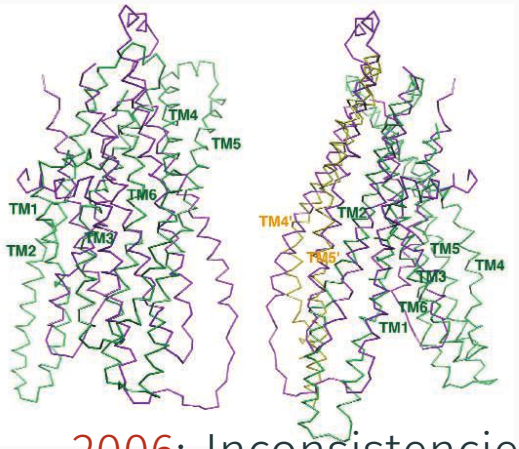
   Reproducibility acts as a Quality control

But, scientific practices have greatly evolved, in particular since we rely on computers

How computers broke science – and what we can do about it

– Ben Marwick, The conversation, 2015

How computers broke science



Geoffrey Chang (Scripps, UCSD) works on crystalography and studies the structure of cell membrane proteins.

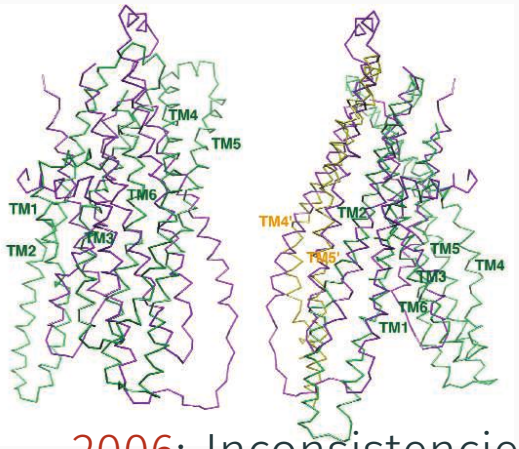
He specialized in structures of multidrug resistant transporter proteins in bacteria: MsbA de Escheria Choli (Science, 2001), Vibrio cholera (Mol. Biology, 2003), Salmonella typhimurium (Science, 2005)

2006: Inconsistencies reveal a programming mistake

> *A homemade data-analysis program had flipped two columns of data, inverting the electron-density map from which his team had derived the protein structure.*

5 retractations that motivate improved software engineering practices in comp. biology

How computers broke science



Geoffrey Chang (Scripps, UCSD) works on crystalography and studies the structure of cell membrane proteins.

He specialized in structures of multidrug resistant transporter proteins in bacteria: MsbA de Escheria Choli (Science, 2001), Vibrio cholera (Mol. Biology, 2003), Salmonella typhimurium (Science, 2005)

2006: Inconsistencies reveal a programming mistake

> *A homemade data-analysis program had flipped two columns of data, inverting the electron-density map from which his team had derived the protein structure.*

5 retractations that motivate improved software engineering practices in comp. biology

There is worse!

- The generalized and intensive use of spreadsheets (COVID tracing)
- Relying on black box statistical methods is infinitely easier than understanding them (Learning and Data Analytics frameworks = nuke)
- Numerical errors and software environment unawareness

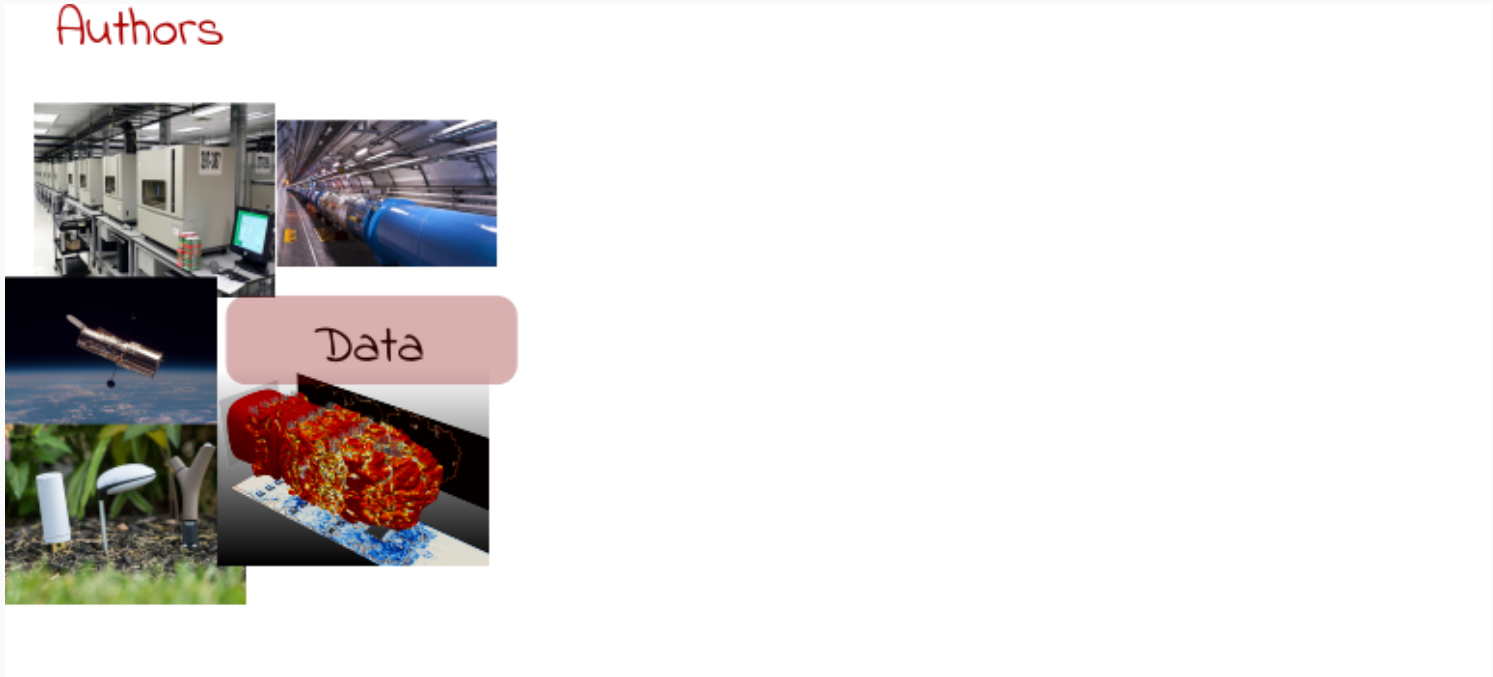
Different Reproducibility Concerns in Modern Science

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

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical issues

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



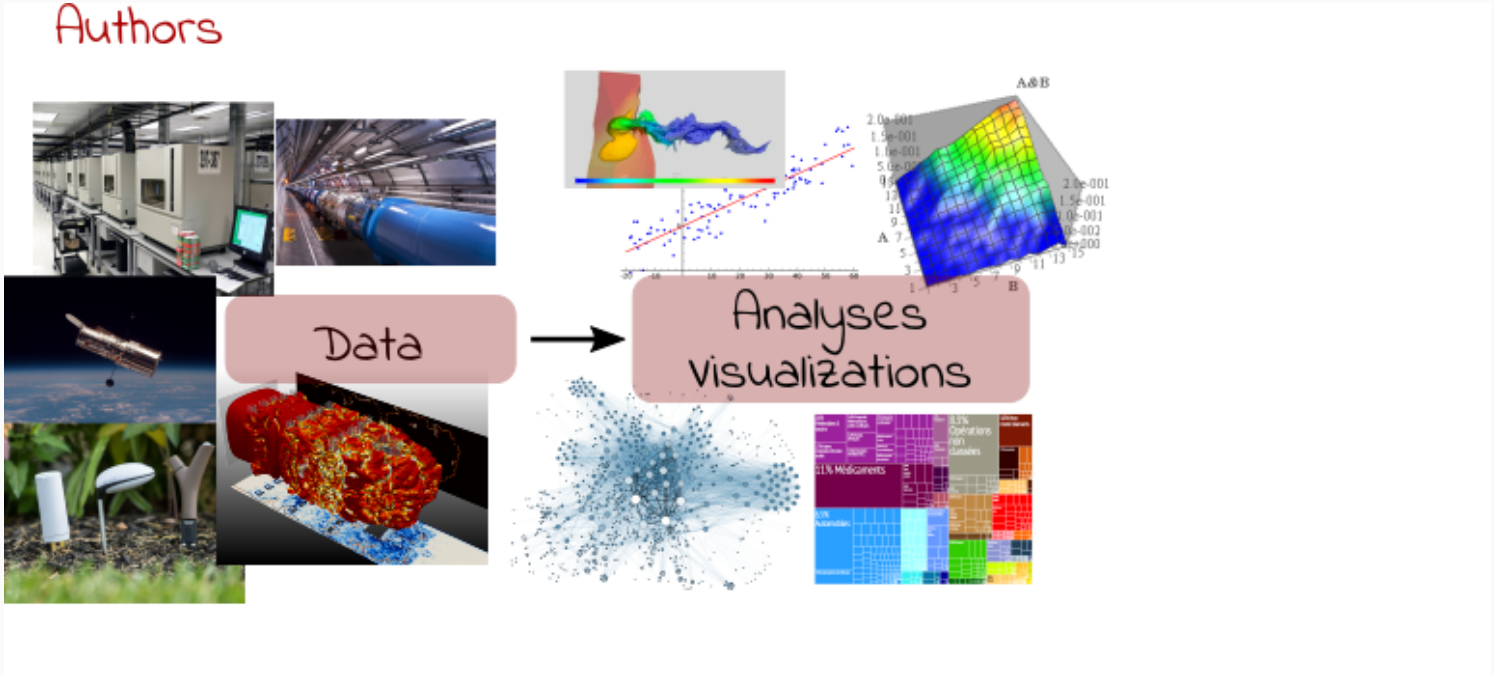
Different Reproducibility Concerns in Modern Science

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

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical issues

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



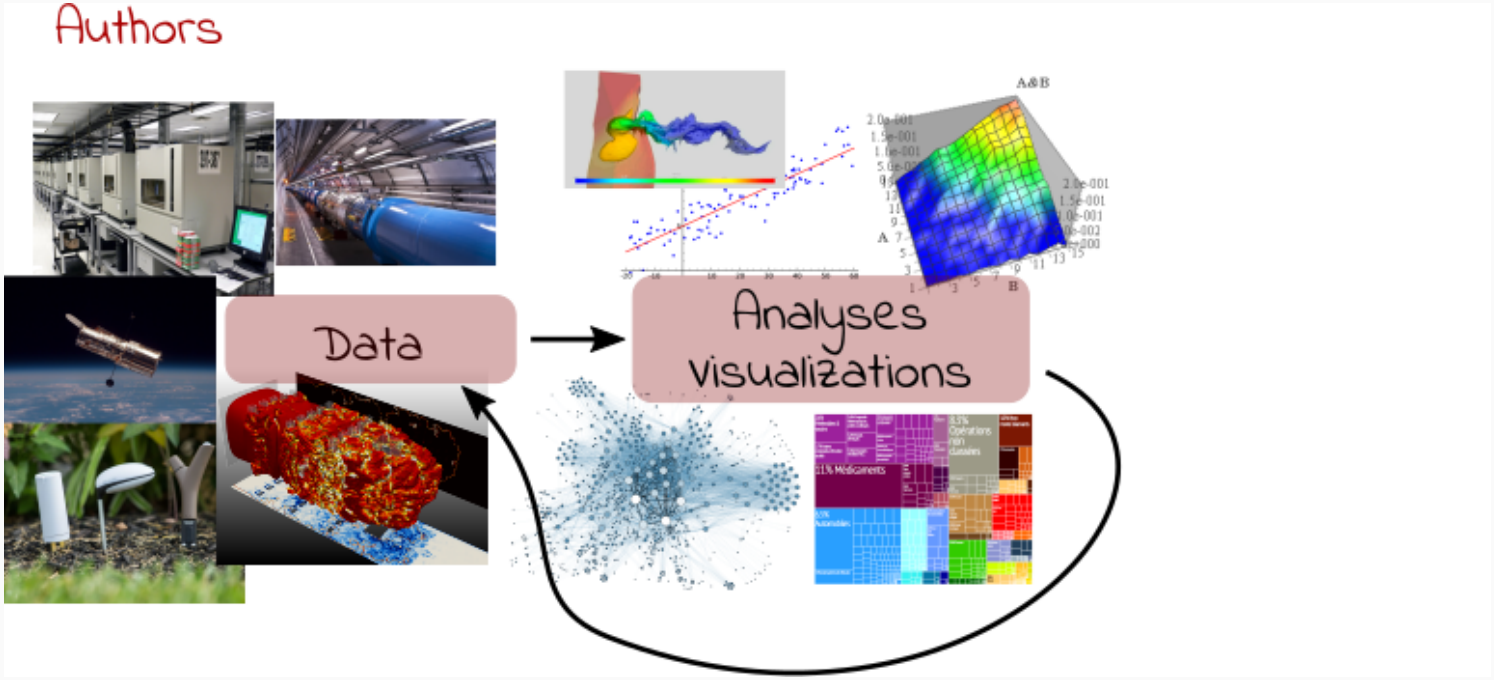
Different Reproducibility Concerns in Modern Science

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

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical issues

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



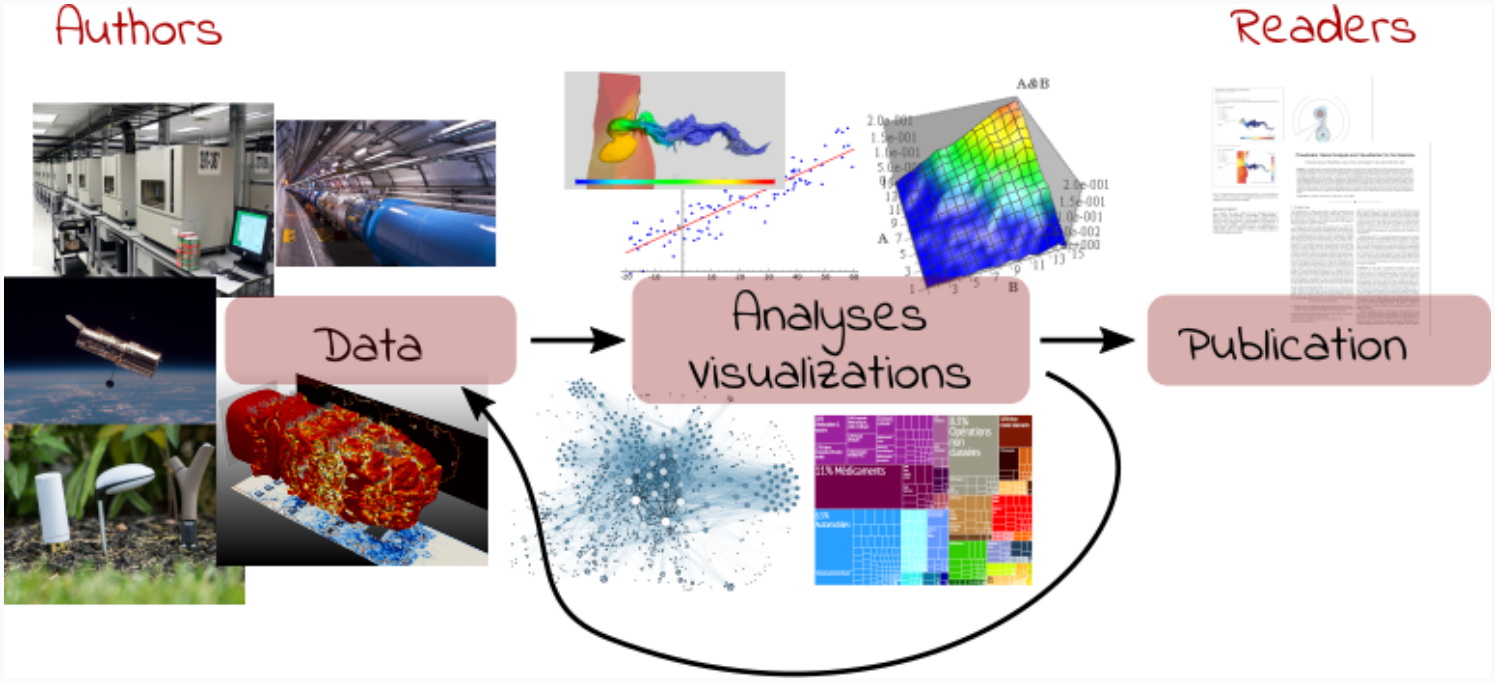
Different Reproducibility Concerns in Modern Science

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

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical issues

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



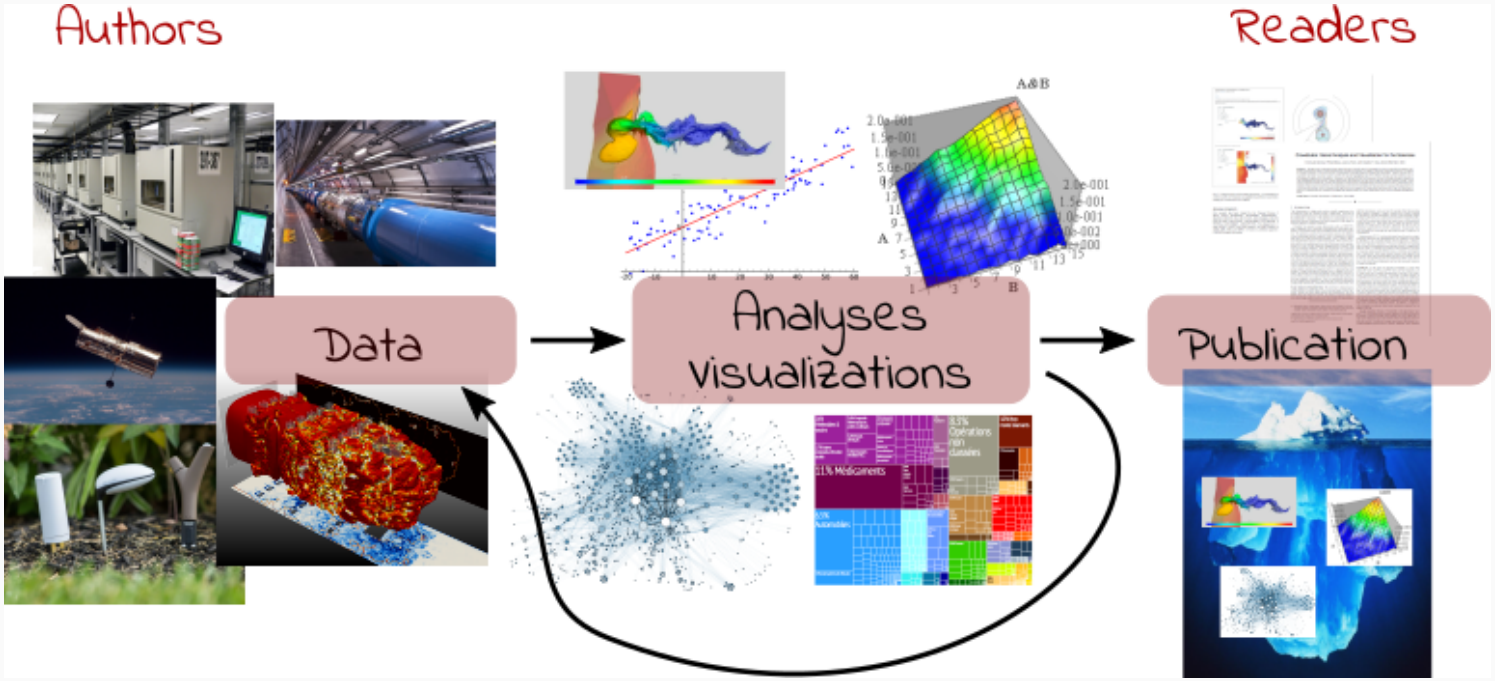
Different Reproducibility Concerns in Modern Science

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

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical issues

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



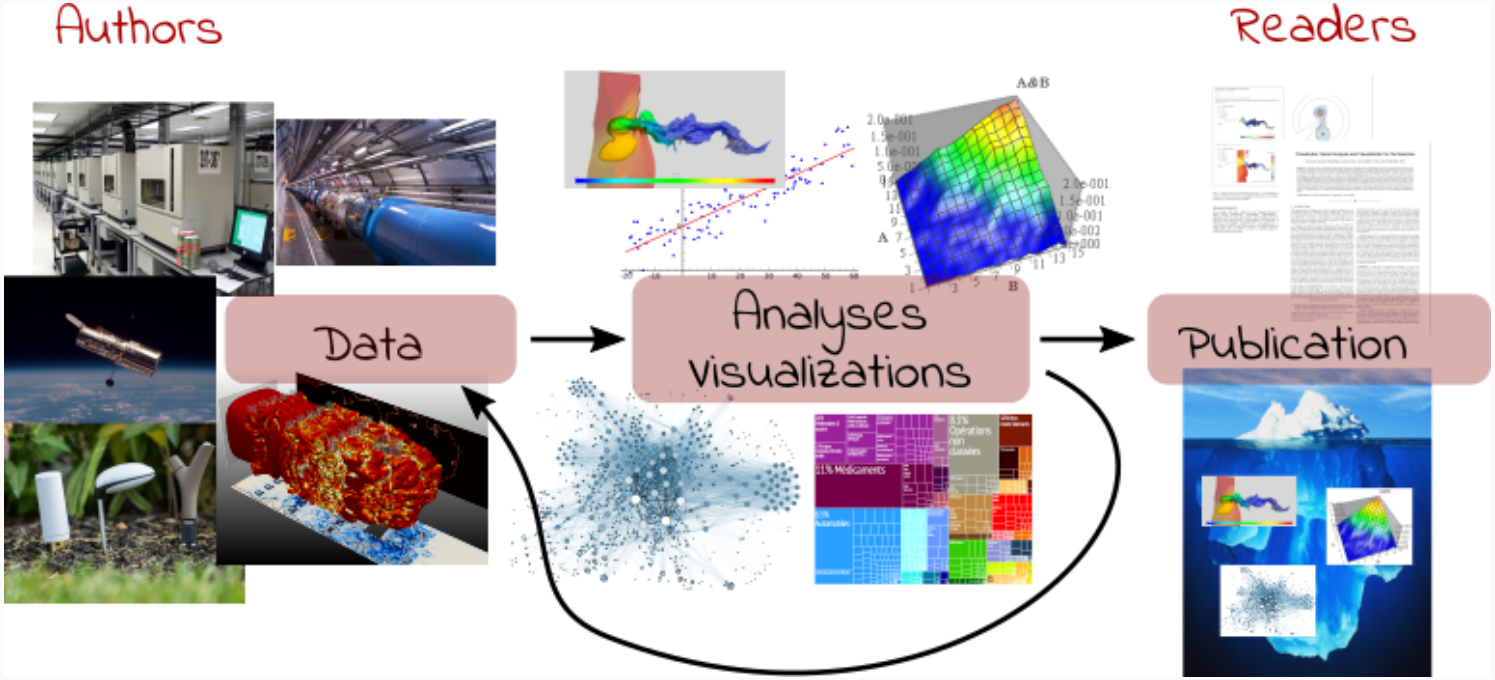
Different Reproducibility Concerns in Modern Science

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

Genomics software engineering, computational reproducibility, provenance

Computational fluid dynamics numerical issues

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



Reproducible Research = Bridging the Gap by working Transparently

Reproducible Research Practices

"Reproducible Research": First Appearance

Claerbout & Karrenbach, meeting of the Society of Exploration Geophysics, 1992

Electronic Documents Give Reproducible Research a New Meaning **RE1.3**

Jon F. Claerbout and Martin Karrenbach, Stanford Univ.

SUMMARY

A revolution in education and technology transfer follows from the marriage of word processing and software command scripts. In this marriage an author attaches to every figure caption a pushbutton or a name tag usable to recalculate the figure from all its data, parameters, and programs. This provid[illegible] of reproducibility in compute[illegible]

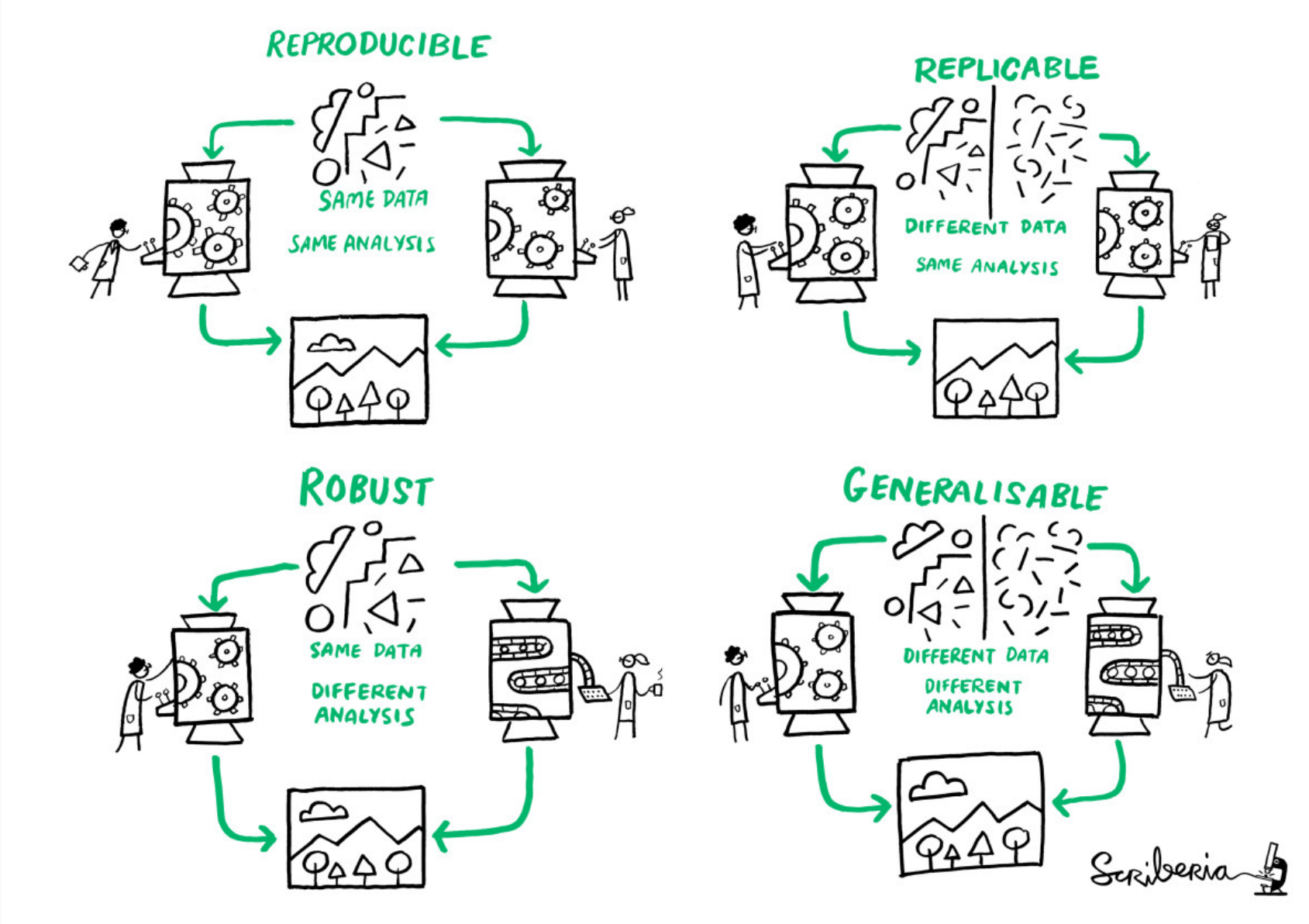
In 1990, we set this sequence of goals:

- Learn how to merge a publication with its underlying computational analysis.
- Teach researchers how to prepare a document in a form where they themselves can reproduce their own research results a year or more later by "pressing a single button".
- Learn how to leave finished work in a condition where coworkers can reproduce the calculation including the final illustration by pressing a button in its caption.
- Prepare a complete copy of our local software environment so that graduating students can take their work away with them to other sites, press a button, and reproduce their Stanford work.
- Merge electronic documents written by multiple authors (SEP reports).
- make incremental improvements in electronic-document software
- seek partners for broadening standards (and making incremental improvements).

Our basic goal is reproducible research. The electronic document is our means to this end. In principle, reproducibility in research can be achieved without electronic documents and that is how we started. Our first nonelectronic reproducible document was a textbook in which the paper document contained the name of a program script in every figure caption. The program scripts were organized by book chapter and section so they could be correlated to an accompanying magnetic tape dump of the file system. The magnetic tape also contained all the necessary data to feed the program script.

Now that we have begun using CD-ROM publication, we can go much further. Every figure caption contains a pushbutton that jumps to the appropriate science directory (folder) and initiates a figure rebuild command and then displays the figure, possibly as a movie or interactive program. We normally display seismic images of the earth's interior. but to reach wider audiences, Figure 1 shows a satellite weather picture which the pushbutton will animate as seen on commerical television. We include all our plot software as well as freely available software from many sources, including compilers and the LaTeX word processing system. Naturally

Reproducibility, Replicability, Robustness, Generalization



Reproducibility (glossary may vary)

Many definitions (*replicability, repeatability, reproducbility*), sometimes conflicting (*new data, same person, independant researcher*)

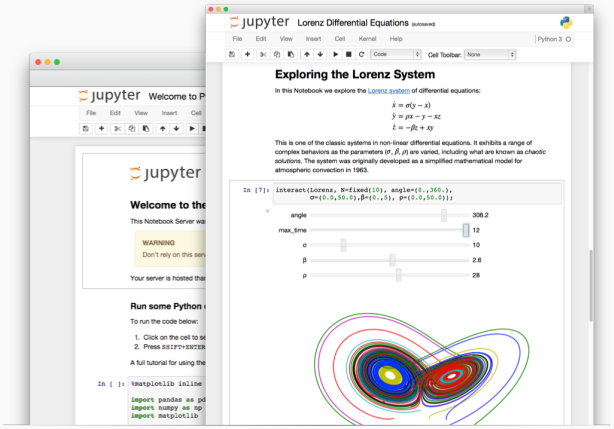
experimental reproducibility	similar input (data) + similar experimental protocol	→	similar results[1]
statistical reproducibility	same input (data) + same analysis	→	same conclusions[2]
computational reproducibility	similar input (data) + same code/software + same software environment	→	exact same results[3]

Reproducible Research = A way of doing science so that scientific experiments, discoveries, results, etc. can be easily reproduced (done again), to be confirmed, or to be built on for the next study.

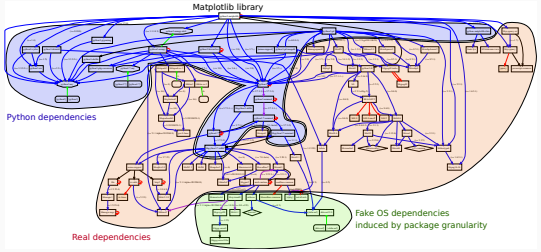
[1] Up-to measurement variability and precision
[2] Independently from (random) sampling variability (fight bias)
[3] Bitwise

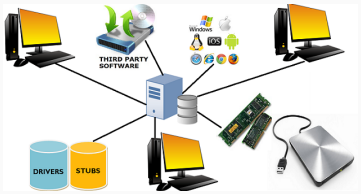
Existing Tools, Emerging Standards

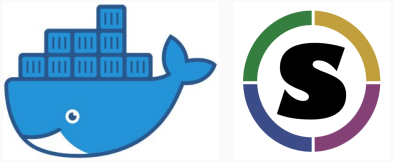
Notebooks and workflows



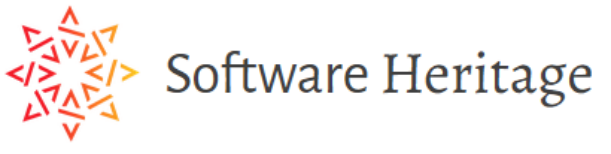
Software environments





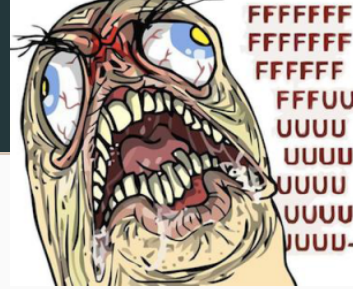


Sharing platforms



Good Practice #1
Taking Notes and Documenting

Frustration as an Author/Reviewer



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

Tool 1: Computational Notebooks/Litterate Programming

Un document computationnel

Mon ordinateur m'indique que π vaut *approximativement*

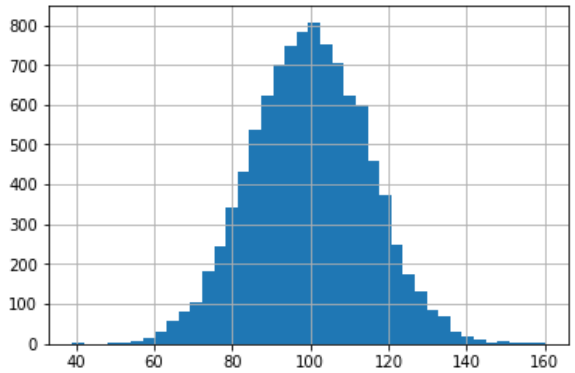
```
3.141592653589793
```

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... ☺).



Tool 1: Computational Notebooks/Litterate Programming

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... ☺).
```

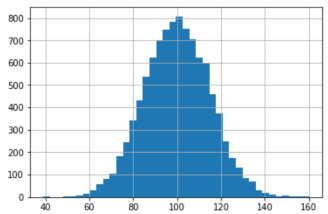
In [3]:

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()
```



Document final

Un document computationnel

Mon ordinateur m'indique que π vaut *approximativement*

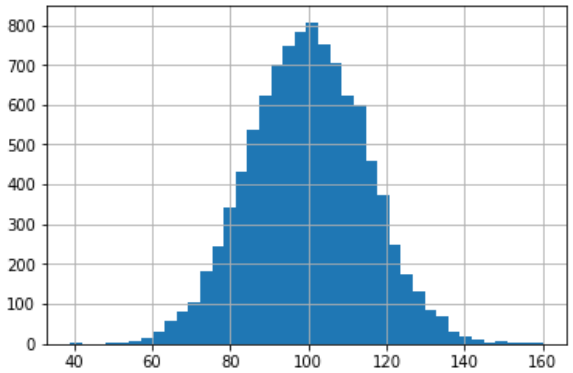
```
3.141592653589793
```

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
```

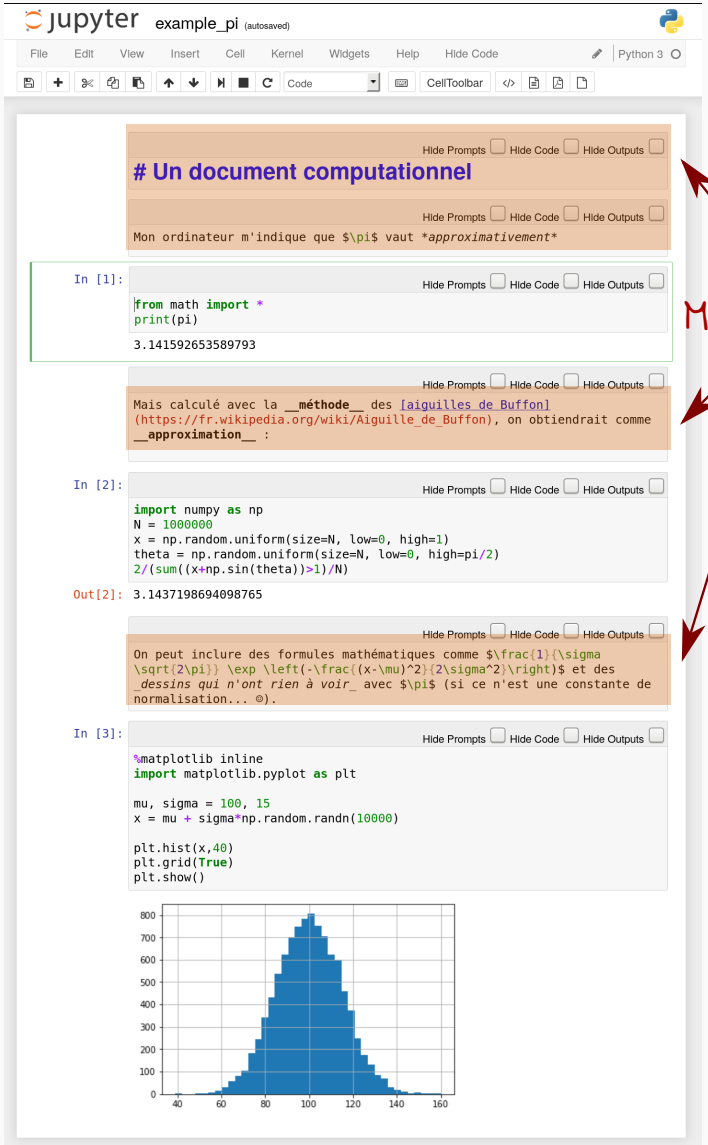
On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des *dessins qui n'ont rien à voir* avec π (si ce n'est une constante de normalisation... ☺).



Tool 1: Computational Notebooks/Litterate Programming

Document initial dans son environnement

Document final



Mark Down

Un document computationnel

Mon ordinateur m'indique que π vaut *approximativement*

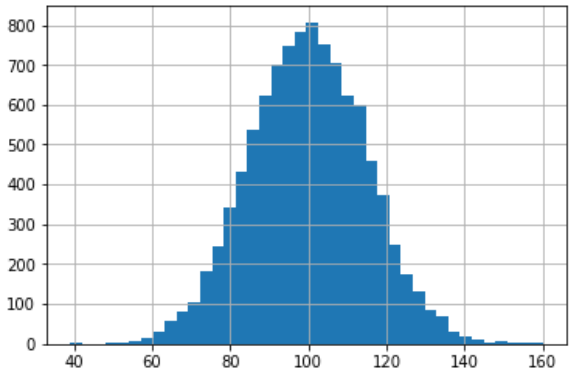
```
3.141592653589793
```

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)
```

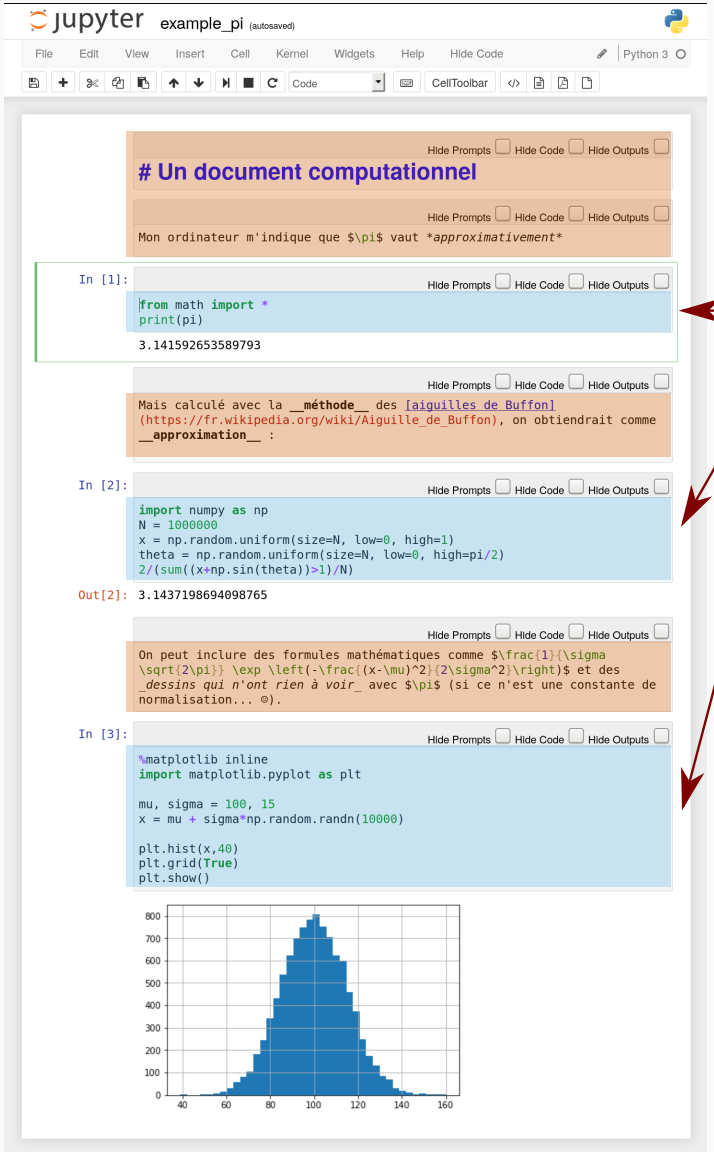
```
3.1437198694098765
```

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... ☺).



Tool 1: Computational Notebooks/Litterate Programming

Document initial dans son environnement



Code

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut *approximativement*

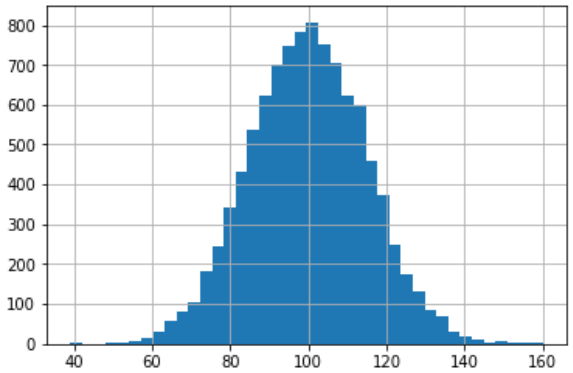
```
3.141592653589793
```

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... ☺).



Tool 1: Computational Notebooks/Litterate Programming

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... ☺).
```

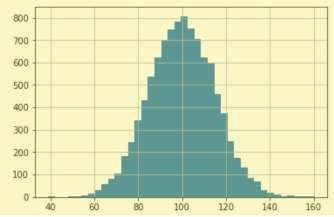
In [3]:

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()
```



Résultats

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut *approximativement*

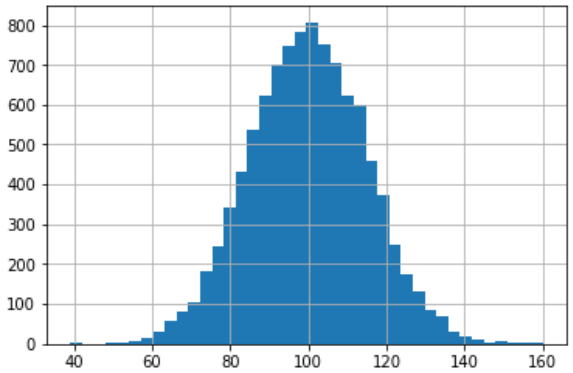
```
3.141592653589793
```

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... ☺).



Tool 1: Computational Notebooks/Litterate Programming

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... ☺).
```

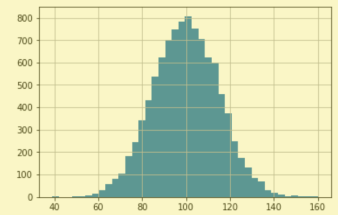
In [3]:

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*

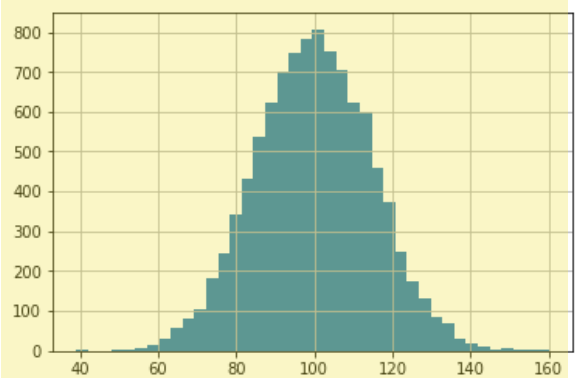
```
3.141592653589793
```

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... ☺).



Tool 1: Computational Notebooks/Litterate Programming

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... ☺).
```

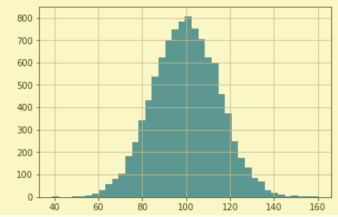
In [3]:

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*

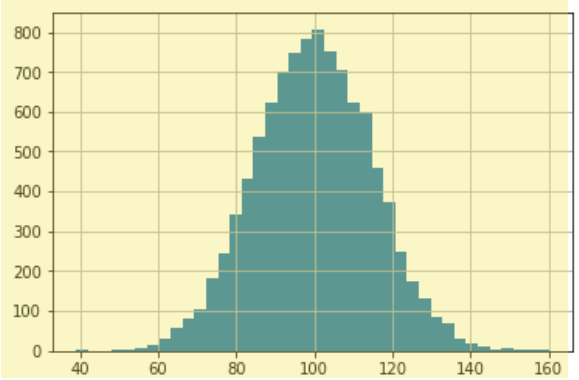
```
3.141592653589793
```

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... ☺).



Journal and Reproducible Article Demo

Document your:

- Hypotheses: keep track of your ideas/line of thoughts
- Experiments: details on how and why an experiment was run, including failed or ambiguous attempts
- Initial analysis or interpretation of these experiments: was the outcome conform to the expectation or not? does it (in)validate the hypothesis? why did you do this or that ?
- Organization: keep track of things to do/fix/test/improve

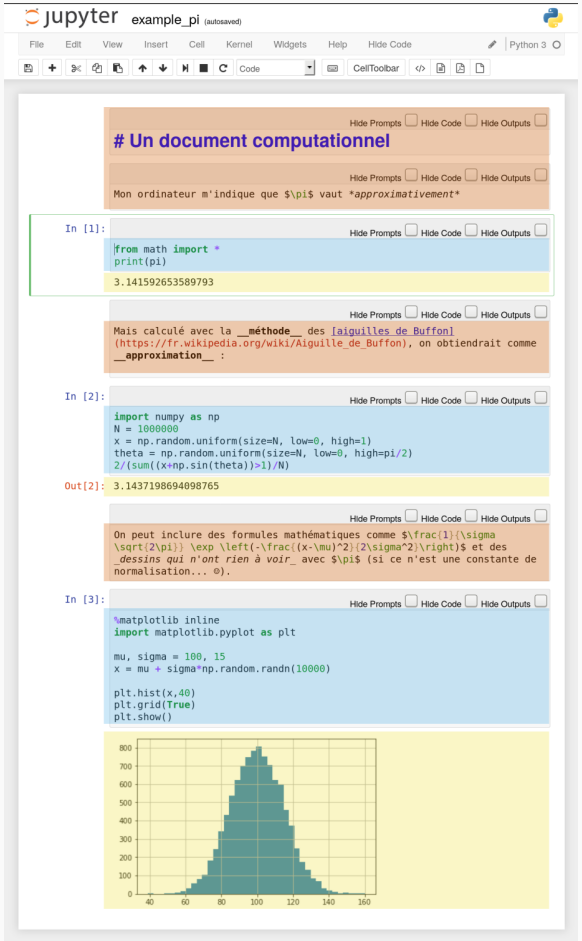
Write for the future you

I have a very intense usage of my journal and I can demo this today

- Experiment results are better structured by dates (add tags)
- Final rendering of results (figures, tables, article, presentation) should be reproducible
- Use plain text and lightweight markup languages (e.g., LaTeX or Markdown)

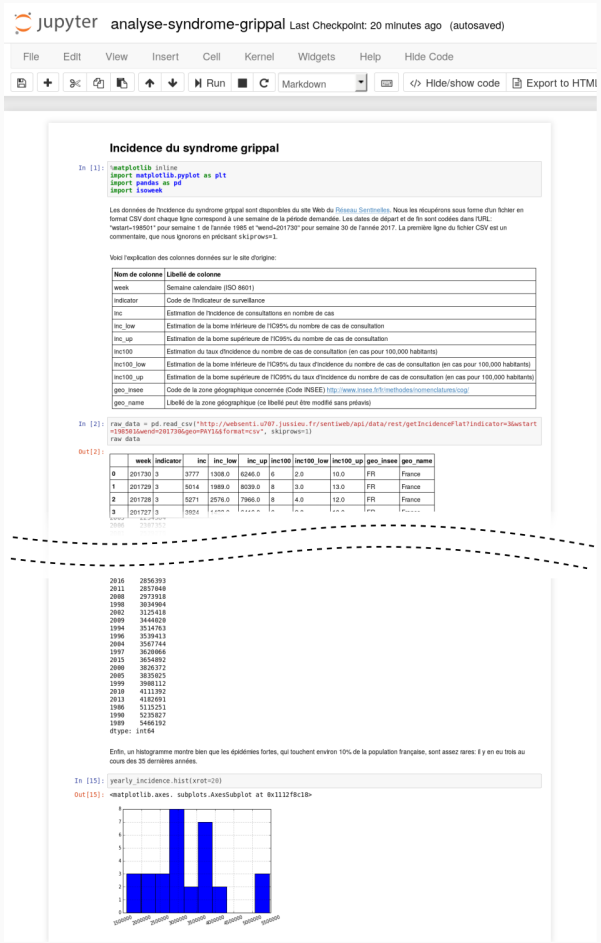
Tool 1 bis: Workflows

Notebooks are no panacea and do not help developing clean code



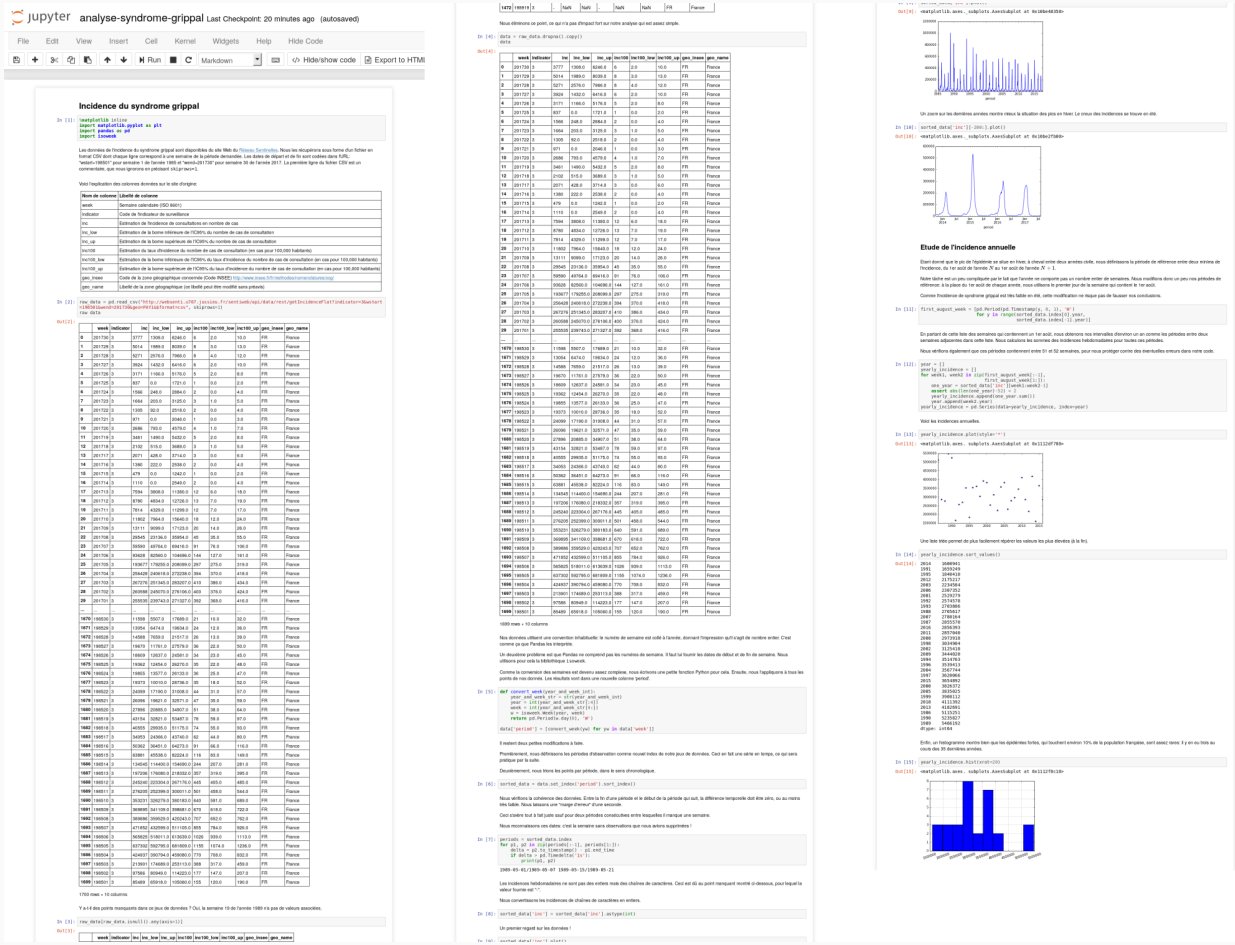
Tool 1 bis: Workflows

Notebooks are no panacea and do not help developing clean code



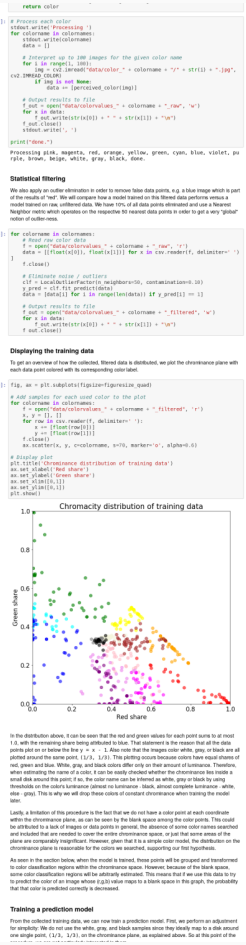
Tool 1 bis: Workflows

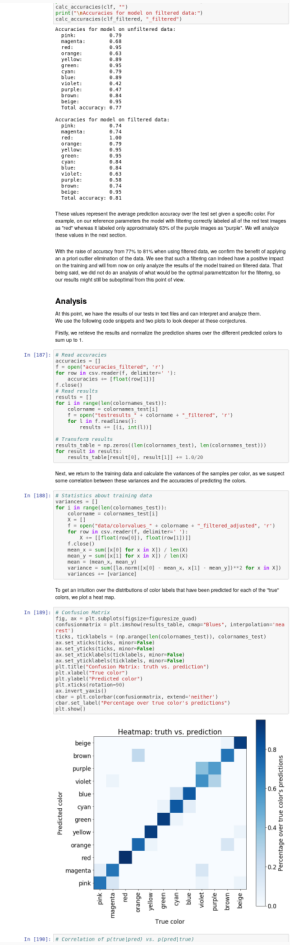
Notebooks are no panacea and do not help developing clean code



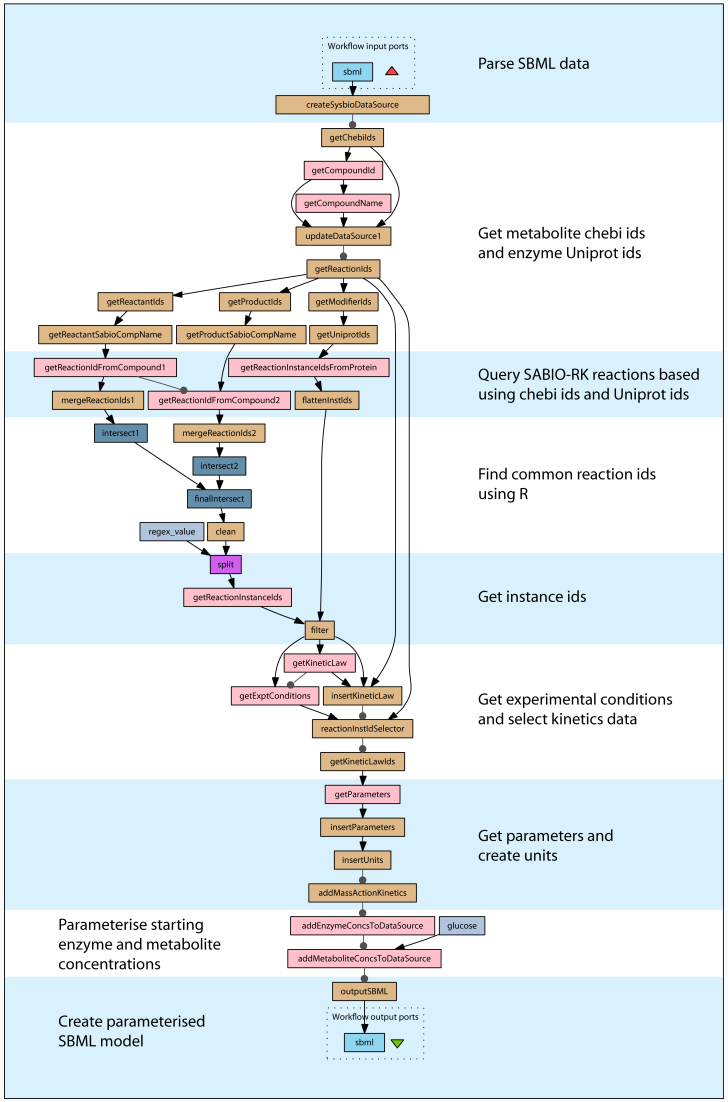
Tool 1 bis: Workflows

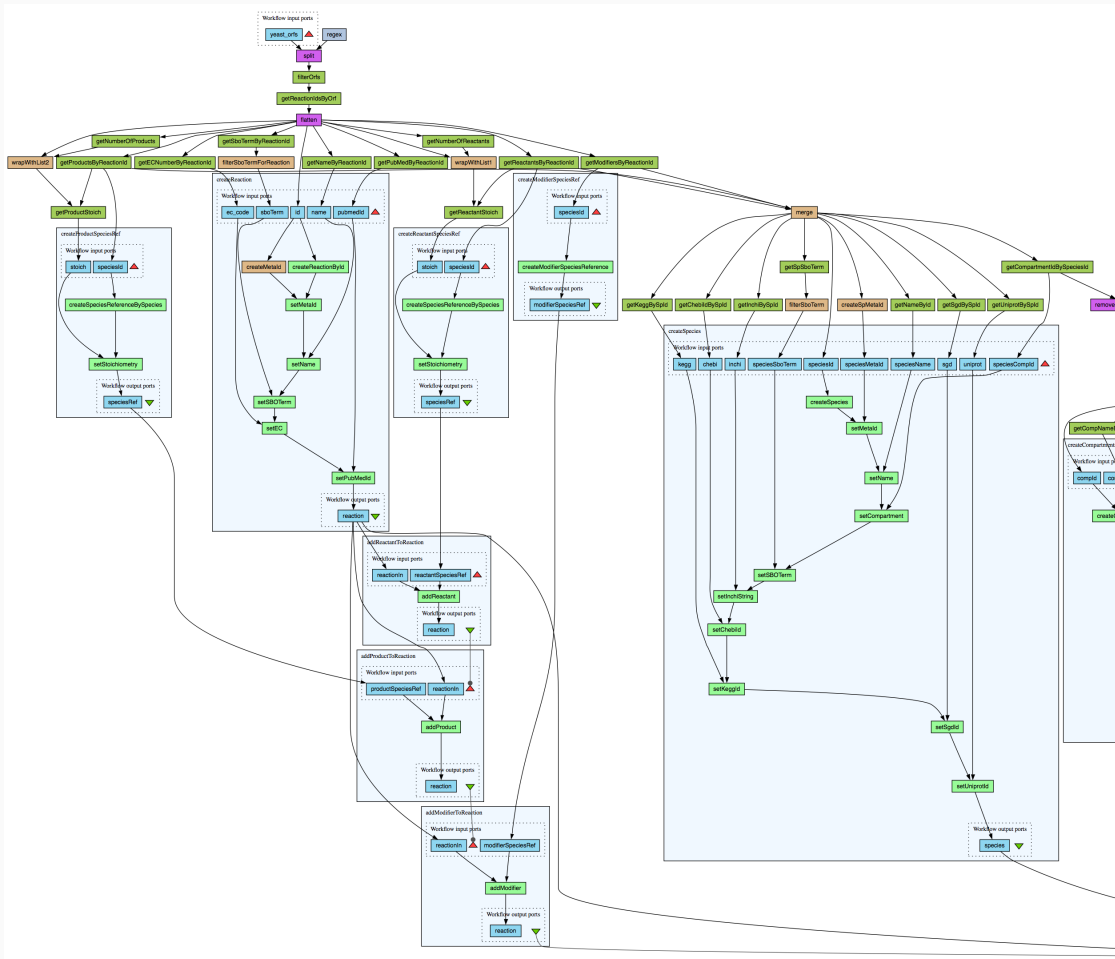
Notebooks are no panacea and do not help developing clean code





Tool 1 bis: Workflows





Tool 1 bis: Workflows

Workflows:

- Clearer high-level view
- Composition of codes and data movement made explicit
- Safer sharing, reusing, and execution
- Notebooks are a variant that is both impoverished and richer
- No simple/mature path from a notebook to a workflow

Examples:

- Galaxy, Kepler, Taverna, Pegasus, Collective Knowledge, VisTrails
- Light-weight: dask, drake, swift, snakemake, ...
- Hybrids: SOS-notebook, ...

Good Practice #2
Controling Software Environment

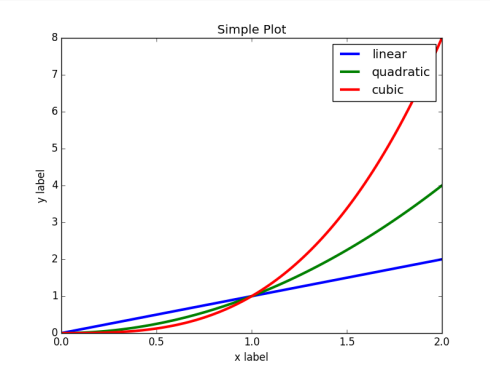
Argh... damned computers

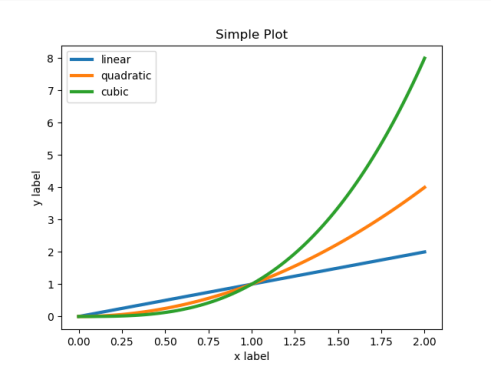
- **Alice**: I got 3.123123 **Bob**: I got segfault
- Damned! It used to work!!! Whenever I upgrade my computer, things break so I try to stay away from this ☹
- Anyway, I don't have the root password The what?...
- Whenever trying the code of my colleague, I had to install Foo but I broke everything and now neither his code nor mine works! ☹
- But hey! Here is my code, feel free to play with it! I'm doing open science 😉

Seriously ? How come all this is so painful ?

Backwards compatibility

- Software environment evolution





Backwards compatibility

- Software environment evolution
- Software evolution and OS heterogeneity

The Effects of FreeSurfer Version, Workstation Type, and Macintosh Operating System Version on Anatomical Volume and Cortical Thickness Measurements (PLOS ONE, 2012)

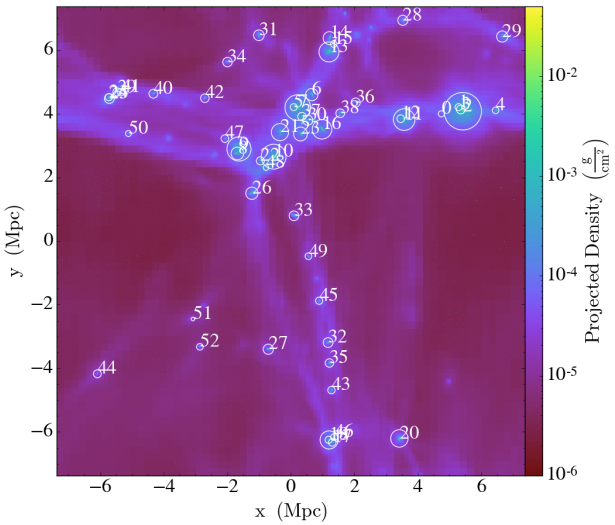
> *Significant differences in volume and cortical thickness were revealed across FreeSurfer versions. In addition, less pronounced differences were found between the Mac and HP workstations and between Mac OSX 10.5 and OSX 10.6.*

Backwards compatibility

- Software environment evolution
- Software evolution and OS heterogeneity
- Impact of the compiler

Assessing Reproducibility: An Astrophysical Example of Computational Uncertainty in the HPC Context (ResCuE-HPC, 2018)

Compiler	Optim.	Largest Halo Avg Mass.	Std. Err	Walltime
gcc@6.2.0	None	2.273E46	1.069E44	22h
gcc@6.2.0	Normal	2.266E46	1.218E44	10h
gcc@6.2.0	High	2.275E46	1.199E44	9h
intel@16.0.3	None	2.271E45	1.587E44	39h
intel@16.0.3	Normal	4.330(45)	1.248E44	7h
intel@16.0.3	High	2.268E46	1.414E44	6h
cce@8.5.5	Low	4.311(45)	1.353E44	16h
cce@8.5.5	Normal	2.271E46	1.261E44	6h
cce@8.5.5	High	2.272E46	1.341E44	5h

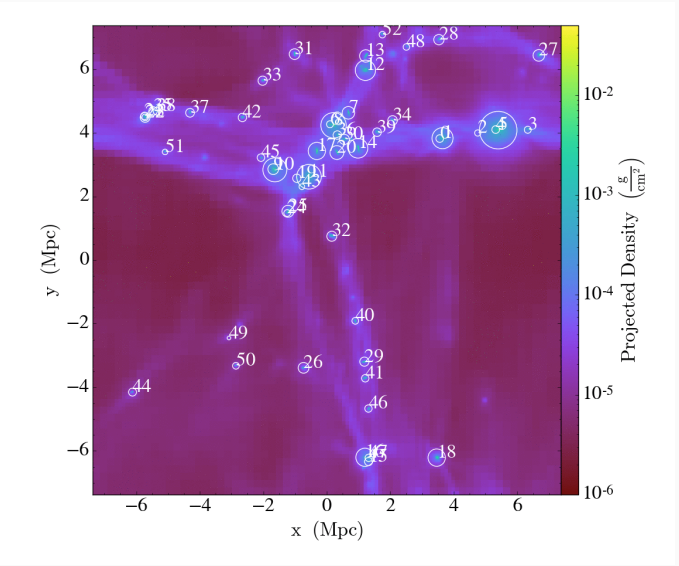


Backwards compatibility

- Software environment evolution
- Software evolution and OS heterogeneity
- Impact of the compiler

Assessing Reproducibility: An Astrophysical Example of Computational Uncertainty in the HPC Context (ResCuE-HPC, 2018)

Compiler	Optim.	Largest Halo Avg Mass.	Std. Err	Walltime
gcc@6.2.0	None	2.273E46	1.069E44	22h
gcc@6.2.0	Normal	2.266E46	1.218E44	10h
gcc@6.2.0	High	2.275E46	1.199E44	9h
intel@16.0.3	None	2.271E45	1.587E44	39h
intel@16.0.3	Normal	4.330(45)	1.248E44	7h
intel@16.0.3	High	2.268E46	1.414E44	6h
cce@8.5.5	Low	4.311(45)	1.353E44	16h
cce@8.5.5	Normal	2.271E46	1.261E44	6h
cce@8.5.5	High	2.272E46	1.341E44	5h

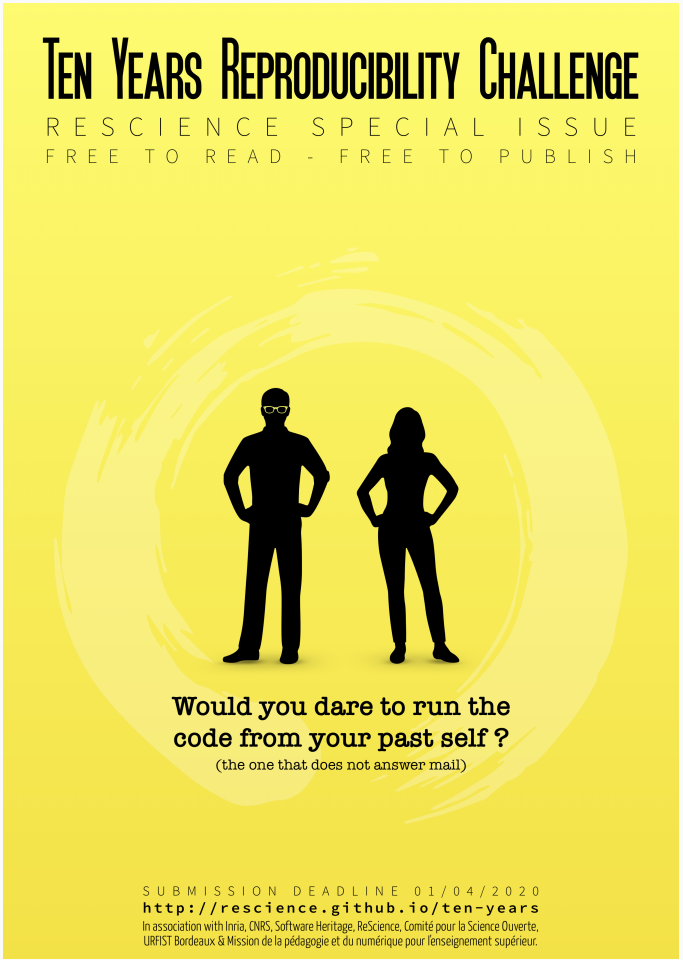


Backwards compatibility

- Software environment evolution
- Software evolution and OS heterogeneity
- Impact of the compiler

Assessing Reproducibility: An Astrophysical Example of Computational Uncertainty in the HPC Context (ResCuE-HPC, 2018)

Compiler	Optim.	Largest Halo Avg Mass.	Std. Err	Walltime
gcc@6.2.0	None	2.273E46	1.069E44	22h
gcc@6.2.0	Normal	2.266E46	1.218E44	10h
gcc@6.2.0	High	2.275E46	1.199E44	9h
intel@16.0.3	None	2.271E45	1.587E44	39h
intel@16.0.3	Normal	4.330(45)	1.248E44	7h
intel@16.0.3	High	2.268E46	1.414E44	6h
cce@8.5.5	Low	4.311(45)	1.353E44	16h
cce@8.5.5	Normal	2.271E46	1.261E44	6h
cce@8.5.5	High	2.272E46	1.341E44	5h



http://rescience.github.io/ten-years/

```
import matplotlib
print(matplotlib.__version__)
```

```
3.1.2
```

Complex ecosystems

```
import matplotlib
print(matplotlib.__version__)
```

```
3.1.2
```

```
apt show python3-matplotlib
```

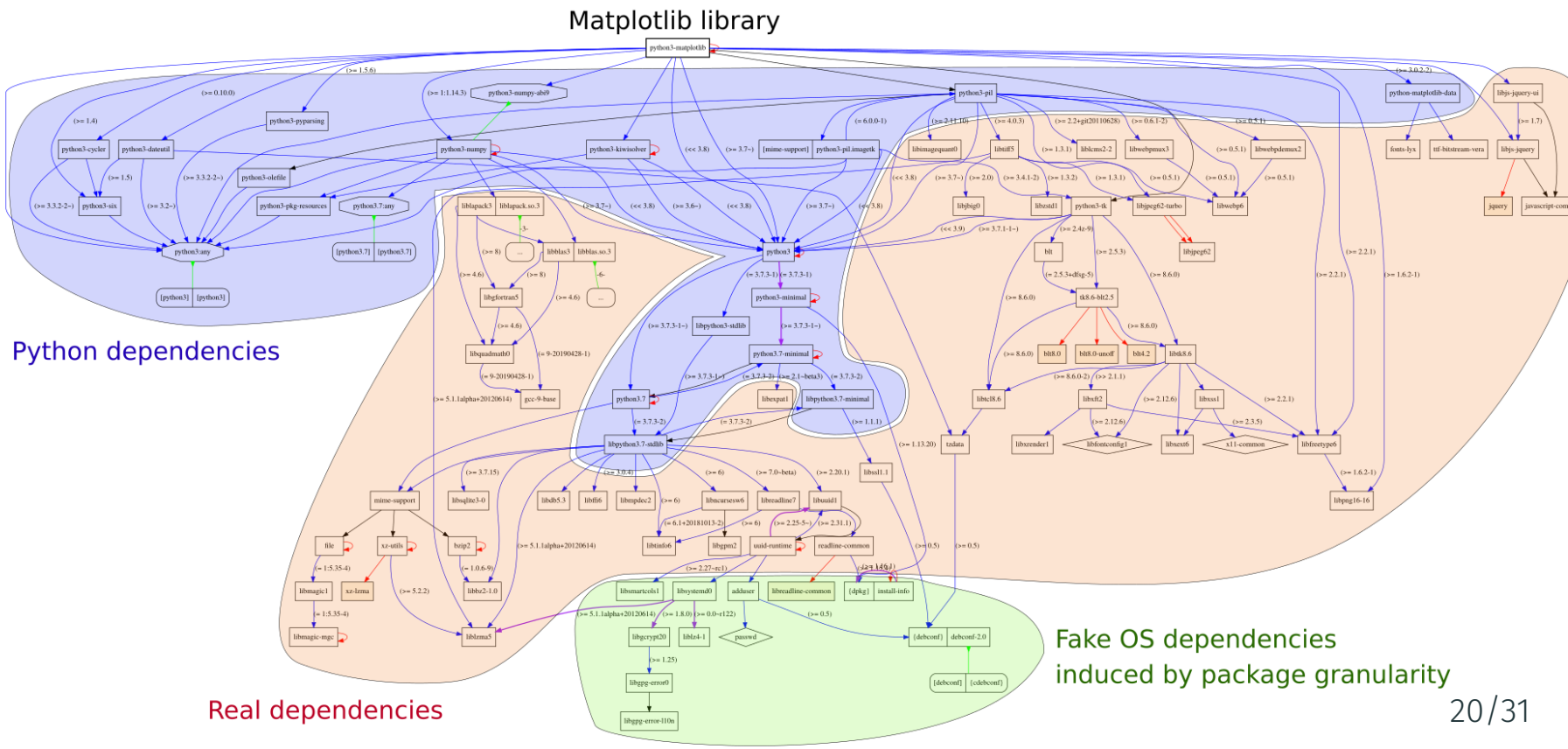
```
Package: python3-matplotlib
Version: 3.1.2-2
Priority: optional
Section: python
Source: matplotlib
Maintainer: Sandro Tosi <morph@debian.org>
Installed-Size: 15.3 MB
Depends: python3-dateutil, python-matplotlib-data (>= 3.1.2-2), python3-pyparsin
six (>= 1.4), libjs-jquery, libjs-jquery-ui, python3-numpy (>= 1:1.16.0~rc1), py
numpy-abi9, python3 (<< 3.9), python3 (>= 3.7~), python3-cycler (>= 0.10.0), pyt
kiwisolver, python3:any, libc6 (>= 2.29), libfreetype6 (>= 2.2.1), libgcc-
s1 (>= 3.0), libpng16-16 (>= 1.6.2-1), libstdc++6 (>= 5.2)
Recommends: python3-pil, python3-tk
Suggests: dvipng, ffmpeg, gir1.2-gtk-3.0, ghostscript, inkscape, ipython3, librs
common, python-matplotlib-doc, python3-cairocffi, python3-gi, python3-gi-cairo,
gobject, python3-nose, python3-pyqt5, python3-scipy, python3-sip, python3-
tornado, texlive-extra-utils, texlive-latex-extra, ttf-staypuft
```

Complex ecosystems

```
import matplotlib
print(matplotlib.__version__)
```

```
3.1.2
```

```
apt show python3-matplotlib
```



Non standard ecosystems

No standard

- Linux (`apt`, `rpm`, `yum`), MacOS X (`brew`, `MacPorts`, `Fink`), Windows (?)
- Neither for installation nor for retrieving the information... ☹

```
import sys
print(sys.version)
import matplotlib
print(matplotlib.__version__)
import pandas as pd
print(pd.__version__)
```

```
3.7.6 (default, Jan 19 2020, 22:34:52)
[GCC 9.2.1 20200117]
3.1.2
0.25.3
```

```
library(ggplot2)
sessionInfo()
```

```
R version 3.6.3 RC (2020-02-21 r77847)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Debian GNU/Linux bullseye/sid

Matrix products: default
BLAS:   /usr/lib/x86_64-linux-gnu/atlas/libblas.so.3.10.3
LAPACK: /usr/lib/x86_64-linux-gnu/atlas/liblapack.so.3.10.3

locale:
[1] C

attached base packages:
[1] stats     graphics  grDevices utils     datasets  method

other attached packages:
[1] ggplot2_3.2.1

loaded via a namespace (and not attached):
 [1] Rcpp_1.0.3       withr_2.1.2      crayon_1.3.4     dply
 [5] assertthat_0.2.1 grid_3.6.3       R6_2.4.1         life
 [9] gtable_0.3.0     magrittr_1.5     scales_1.1.0     pill
[13] rlang_0.4.4      lazyeval_0.2.2   glue_1.3.1       purr
[17] munsell_0.5.0    compiler_3.6.3   pkgconfig_2.0.3  colo
[21] tidyselect_1.0.0 tibble_2.1.3
```

Argh... damned computers

- 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, feel free to play with it! I'm doing open science $😉 $

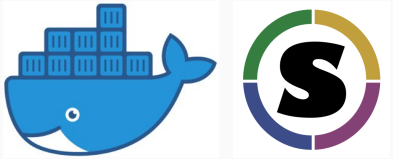
  Are you really aware of your dependencies ?

  - No one will ever run/use your code if it isn't easy to install
  - No one will ever manage to run your code if you don't document how to run it
  - Others (even you) are unlikely to get the same results unless you automate the execution

Tool 2: Containers and Package Managers

The good

The bad



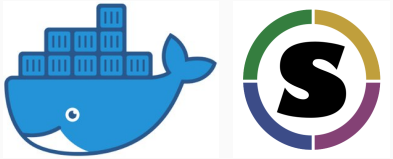
The uggly

Automatic tracking

Tool 2: Containers and Package Managers

The good

The bad



The uggly

Automatic tracking

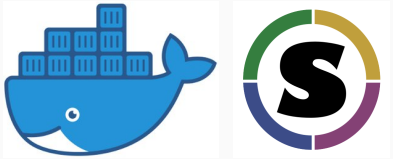
Containers

- 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>`

Tool 2: 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
  - Recipes rarely follow *reproducible good practices*

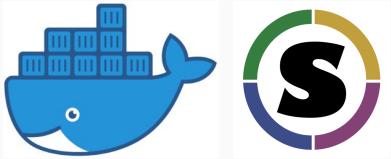
```
FROM ubuntu:20.04
RUN apt-get update
      && apt-get upgrade -y
      && apt-get install -y ...
```

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

Tool 2: Containers and Package Managers

The good

The bad



The uggly

Automatic tracking

Containers

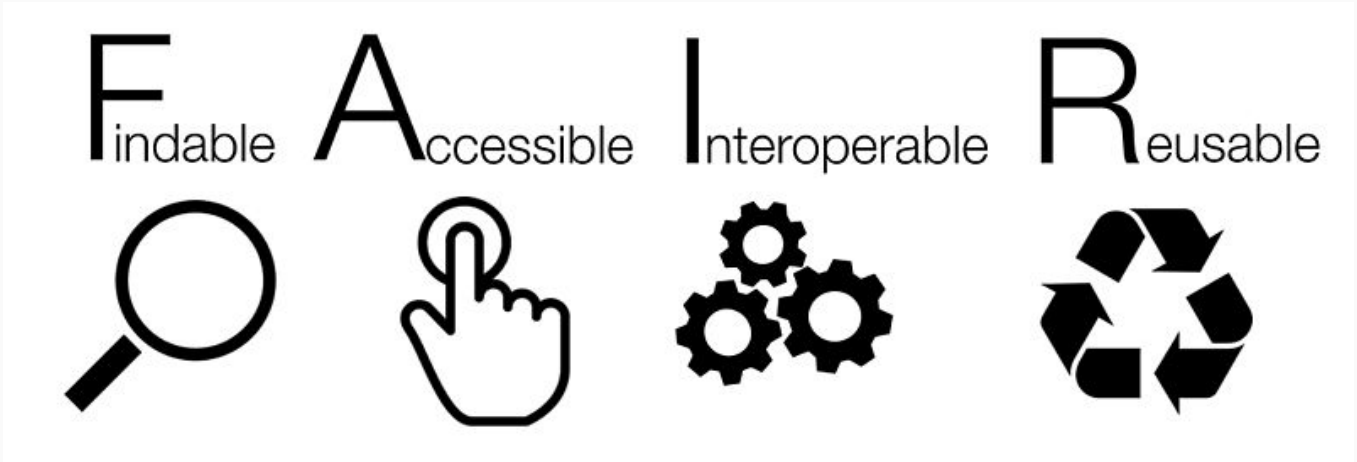
- Pros: Lightweight, Good isolation, Easy to use
- Cons: Opaque, Container build is generally not reproducible

Package managers

- 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)

Good Practice #3
Version Control and Archiving

FAIR principles



https://www.go-fair.org/fair-principles/

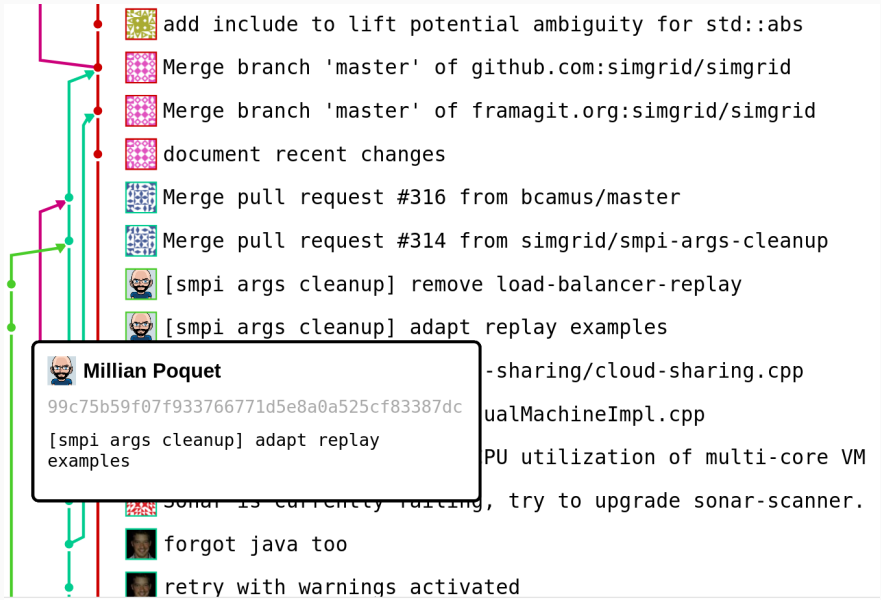
- *"Open as much as possible and close as much as necessary"*
- Management, publication, annotation (metadata), archiving
- Source code = specific data with specific consideration

Let's go beyond general principles!

Tool 3: Version Control and Forge

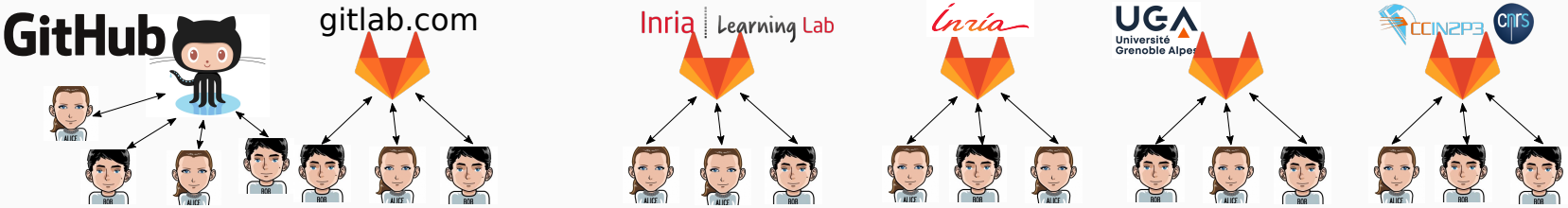
Git = version control

- Developed in 2005 by Linus Torvalds for the kernel development
- Local and efficient rollbacks
- Distributed: everyone has a full copy of the history



GitHub, GitLab, and Co

- Free hosting of public projects, social network
- Web interfaces (browsing, preview, online editing)
- User management (read/write, public/private)
- Issues, Continuous Integration, ...



Tool 3bis: 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 from its publication date.
- 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

Tool 3bis: 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 from its publication date.
- 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

INTERNET ARCHIVE

Collect/Preserve/Share

What Will it Take ?

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

Major conferences

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

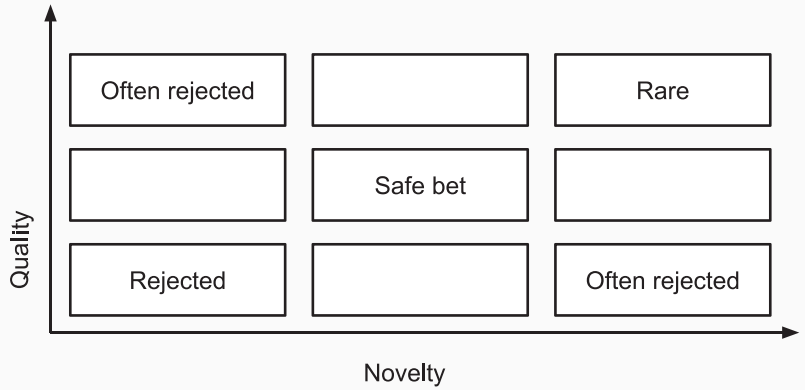
Joelle Pineau @ NeurIPS'18

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

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

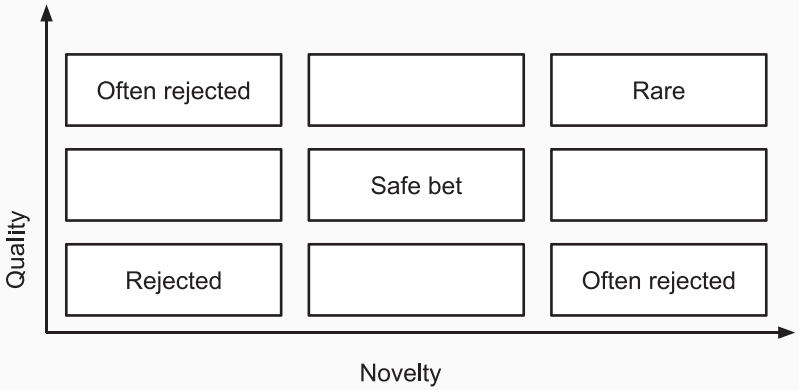
Changing Academic Practices (~~Publish or Perish~~)

- Goodhart's Law: Are Academic Metrics Being Gamed?, M. Fire 2019
  - AI: over 1,000 ranked journals ($\times$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



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



- 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*

What about Open Science ?

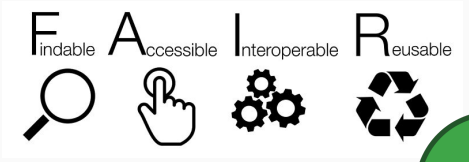
Plan National pour la Science Ouverte (BSN ⇝ CoSO)

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

Main pillars:

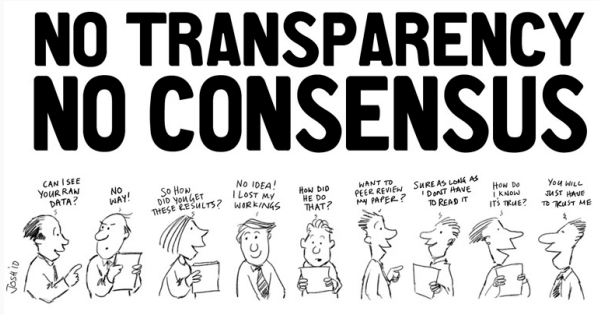
1. Open access

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

Resources and Acknowledgments

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

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

Resources and Acknowledgments

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 2022
- MOOC RR "Advanced" planned for ~~2021~~ 2022
  - Software environment control
  - Scientific workflow
  - Managing data