

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

Los Angeles Times | BUSINESS

LOCAL U.S. WORLD BUSINESS SPORTS ENTERTAINMENT HEALTH STYLE TRAVEL

Science has lost its way, at a big cost to humanity

Researchers are rewarded for splashy findings, not for double-checking accuracy. So many scientists looking for cures to diseases have been building on ideas that aren't even true.

Science AAAS-ORG FEEDBACK HELP LIBRARAINS All Science Journals ▾| Search The
AAAS NEWS SCIENCE JOURNALS CAREERS MULTIMEDIA COLLECTIONS

Science Home Current Issue Previous Issues Science Express Science Products My Science About the Journal

Science Magazine > 22 MARCH 2012 > MONDAY, 26 (1660) : 229

Article Views Science 17 January 2014;
Vol. 343 no. 6168 p. 229
DOI: 10.1126/science.1250475

Summary Full Text Full Text (PDF)

EDITORIAL Reproducibility Marcia McNutt

Science advances on a foundation of trusted data, but the lack of reproducibility is threatening that approach that scientists used to gain confidence in their results. As the scientific community has learned, however, we are not alone in our challenges. In response, we are taking steps to address the problem. One recommendation of the U.S. National Institute of General Medical Sciences' recent study of ways to increase transparency,¹ authors will indicate handling (such as how to deal with outliers), whether they ensure a sufficient signal-to-noise ratio, whether the experimenter was blind to the conduct of the experiments, and whether the results were replicated.

Save to My Folders Download Citation Alert Me When Article Is Cited Post to CiteULike E-mail This Page Rights & Permissions Commercial Reprints and E-Prints View Publication Citation Related Content

Announcement: Reducing our irreproducibility: Nature News & Comment

nature.com Sitemap Log in Register

nature International weekly journal of science

Home News & Comment Research Careers & Jobs Current Issue Archive

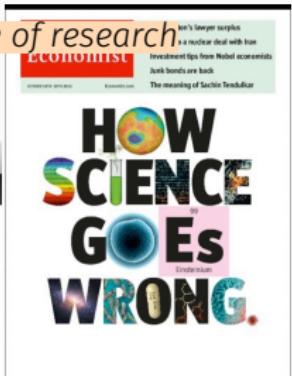
Archive > Volume 483 > issue 7591 > Editorial > Article

NATURE | EDITORIAL

Announcement: Reducing our irreproducibility

24 April 2013

PDF Rights & Permissions



nature International weekly journal of science

Menu Advanced search Search

archive > volume 483 - issue 7591 - editorials - article

NATURE | EDITORIAL

Must try harder

Nature 483, 509 (29 March 2012) doi:10.1038/483509a
Published online: 28 March 2012

PDF Cite this Reprints Rights & permissions Article metrics

Too many sloppy mistakes are creeping into scientific papers. Lab heads must look more rigorously at the data — and at themselves.

TheScientist

EXPLORING LIFE. INSPIRING INNOVATION

NIH Tackles Irreproducibility

The federal agency speaks out about how to improve the quality of scientific research.

By Jef Akst | January 28, 2014

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 \approx \$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 \approx \$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 \approx \$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!*

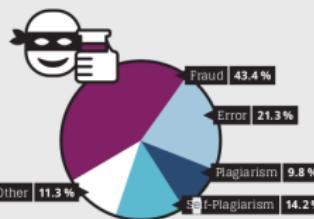
A CREDIBILITY CRISIS?

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

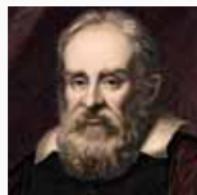
The Battle against Scientific Fraud in the CNRS International Magazine

Biomedical fraud in figures

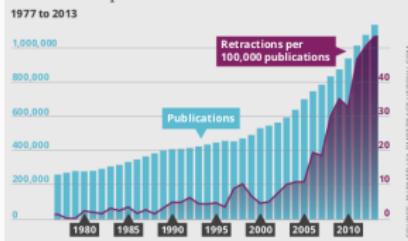
Cause of retraction 1977 to 2012



SOURCE: FAUCI ET AL. (2012) PNAS



Number of publications and retractions



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

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

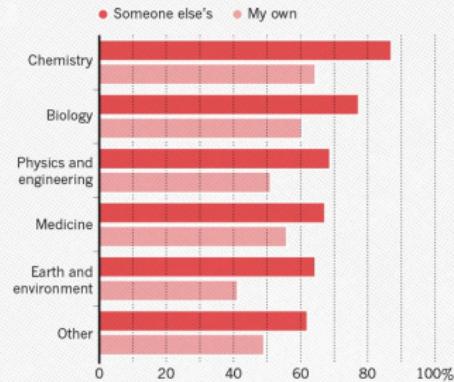
- Every domain has its black sheep

- The publish or perish pressure is a pain

A REPRODUCIBILITY CRISIS?

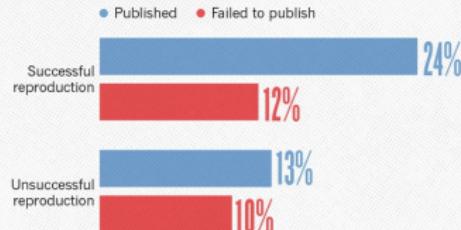
HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



HAVE YOU EVER TRIED TO PUBLISH A REPRODUCTION ATTEMPT?

Although only a small proportion of respondents tried to publish replication attempts, many had their papers accepted.



1,500 scientists lift the lid on reproducibility,

Nature, May 2016

Social causes

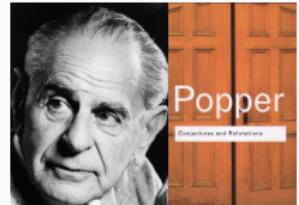
- Fraud, conflict of interest (pharmaceutic, ...)
- No incentive to reproduce/check our own work (afap), nor the work of others (big results!), nor to allow others to check (competition)
- Peer review does not scale: 1+ million articles per year!

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

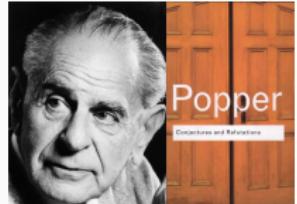
1934: Karl Popper puts the notions of **falsifiability** and **crucial experiment** as the **hallmark of science**



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

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

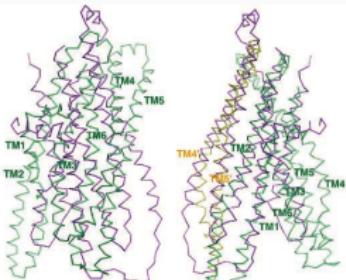
But, scientific practices have greatly evolved, in particular since we rely on computers



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 crystallography and studies the structure of cell membrane proteins.

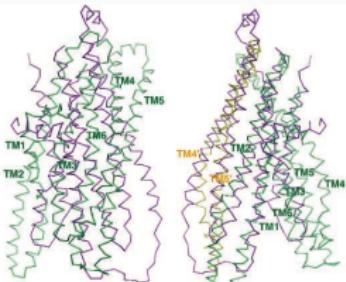
He specialized in structures of **multidrug resistant transporter proteins in bacteria**: MsbA de Escherichia Choli (Science, 2001), Vibrio cholera (Mol. Biology, 2003), Salmonella typhimurium (Science, 2005)

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 retractions that motivate **improved software engineering practices** in comp. biology

How COMPUTERS BROKE SCIENCE



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

He specialized in structures of **multidrug resistant transporter proteins in bacteria**: MsbA de Escherichia 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 retractions 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 CONCERN IN MODERN SCIENCE

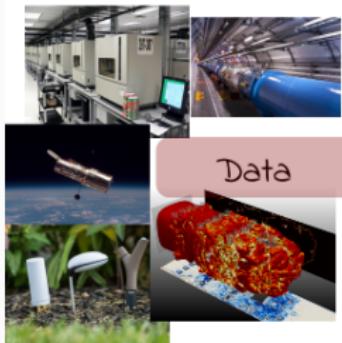
Social Sciences, Oncology, ... methodology, statistics, pre-registration

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

Authors



Data

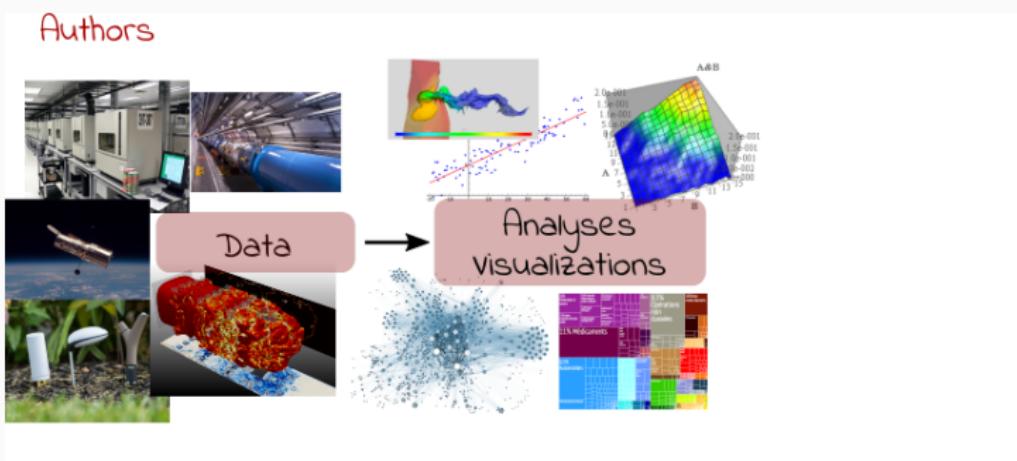
DIFFERENT REPRODUCIBILITY CONCERN IN MODERN SCIENCE

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



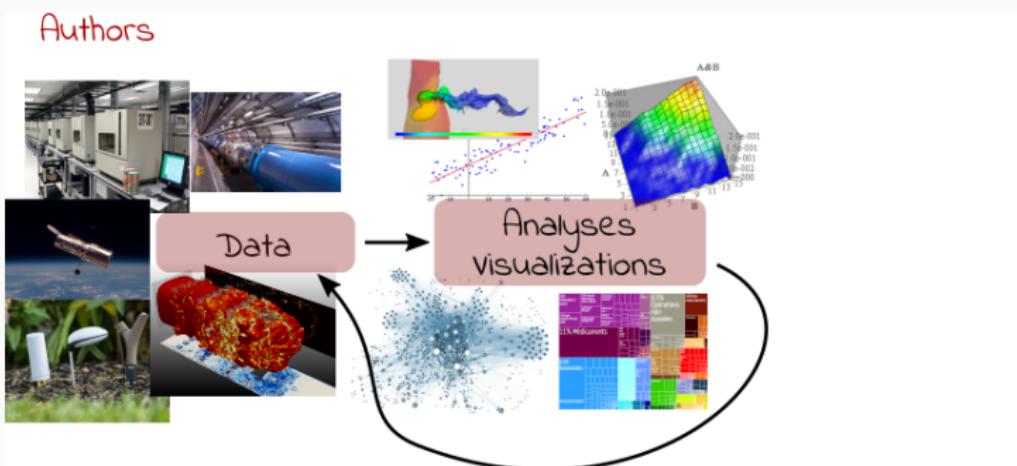
DIFFERENT REPRODUCIBILITY CONCERN IN MODERN SCIENCE

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



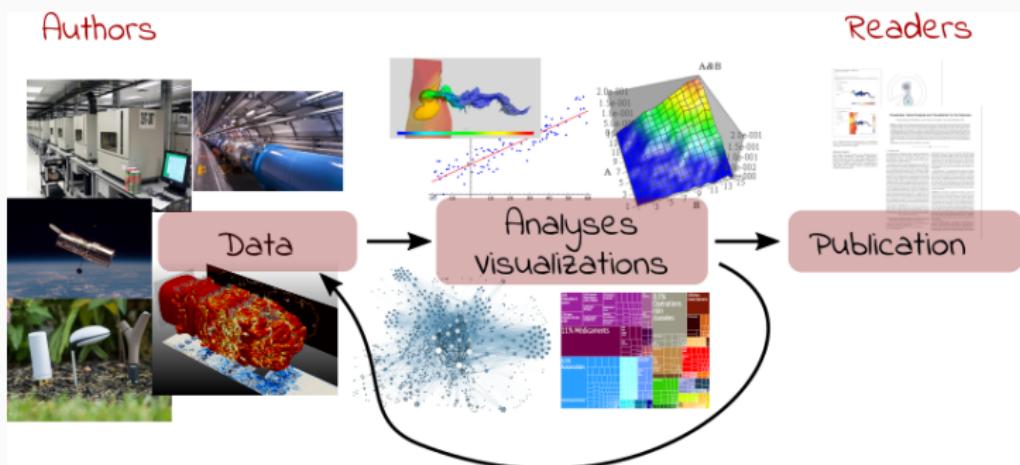
DIFFERENT REPRODUCIBILITY CONCERN IN MODERN SCIENCE

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



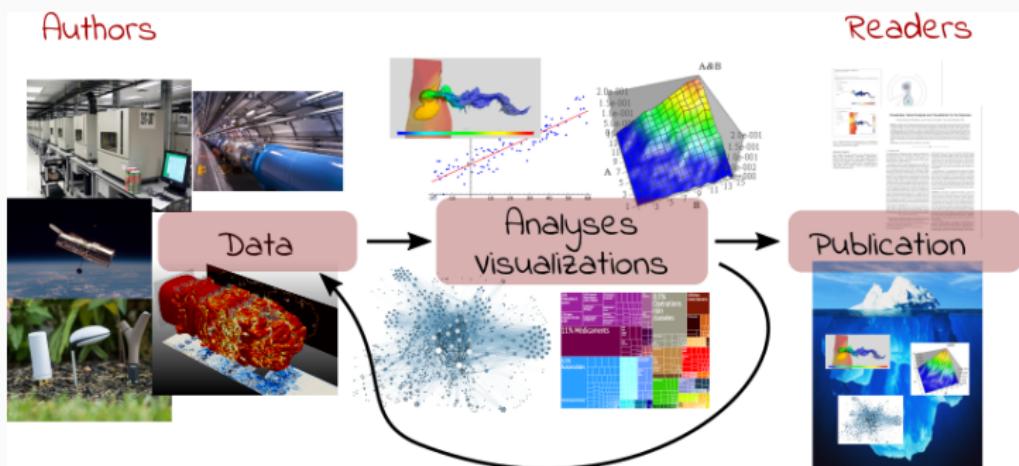
DIFFERENT REPRODUCIBILITY CONCERN IN MODERN SCIENCE

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



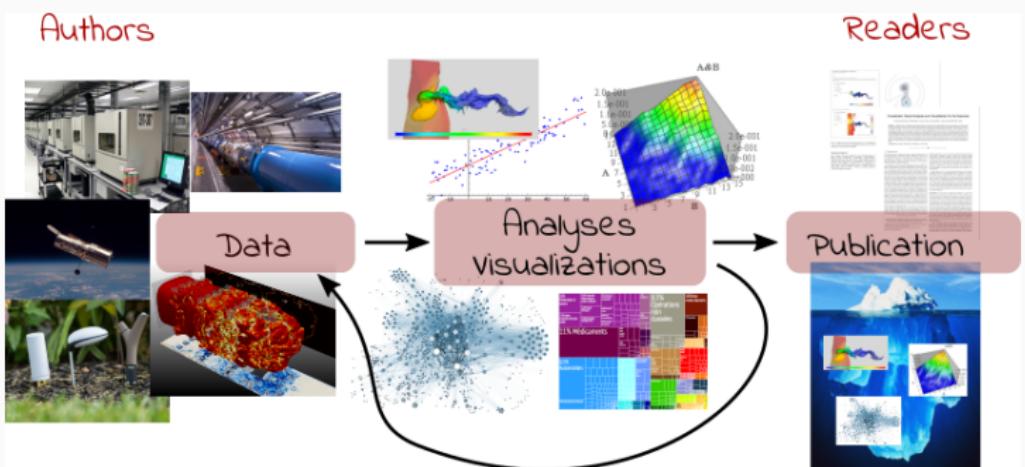
DIFFERENT REPRODUCIBILITY CONCERN IN MODERN SCIENCE

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 provides a new meaning of reproducibility in computer documents.

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).

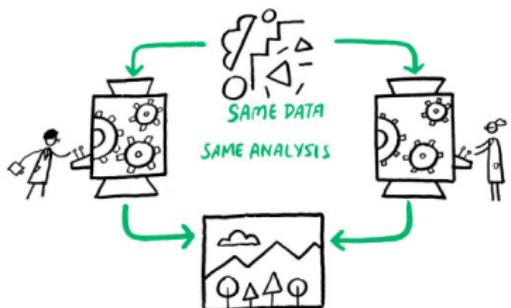
- make incremental improvements in electronic-document software
- seek partners for broadening standards (and making incremental improvements).

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.

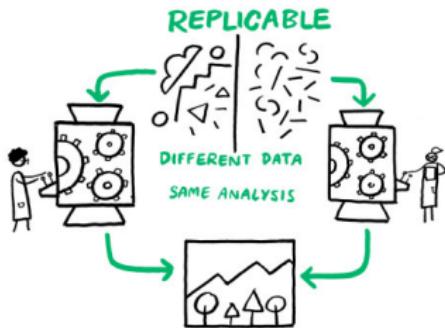
Now that we have begun using CD-ROM publication, we can go much further. Every figure caption contains a pushbutton that jumps to the appropriate science directory (folder) and initiates a figure rebuild command and then displays the figure, possibly as a movie or interactive program. We normally display 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 commercial television. We include all our plot software as well as freely available software from many sources, including compilers and the L^AT_EX word processing systems. Naturally we cannot include licensed software, but with the exception

REPRODUCIBILITY, REPLICABILITY, ROBUSTNESS, GENERALIZATION

REPRODUCIBLE



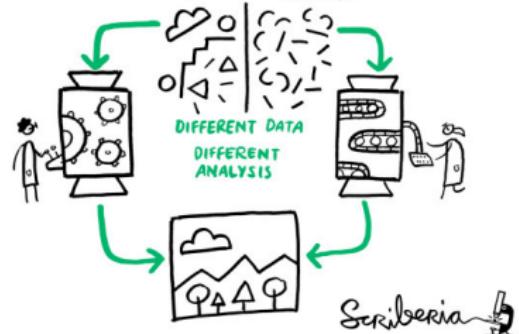
REPLICABLE



ROBUST



GENERALISABLE



REPRODUCIBILITY (GLOSSARY MAY VARY)

Many **definitions** (*replicability, repeatability, reproducibility*), sometimes conflicting
(*new data, same person, independent researcher*)

experimental reproducibility	similar input (data) + similar experimental protocol	→	similar results ¹
statistical reproducibility	different 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.

– Courtesy G. Durrif, 2021

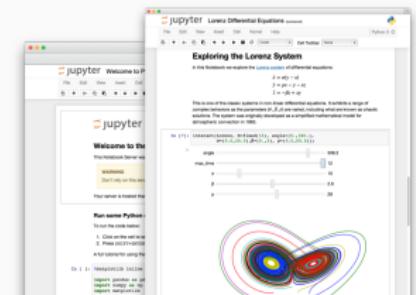
¹Up-to measurement variability and precision

²Independently from (random) sampling variability (fight bias)

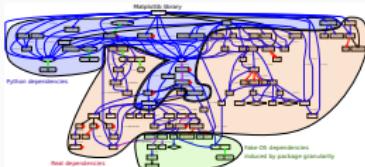
³Bitwise

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

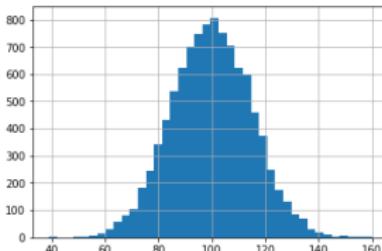
3.141592653589793

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

The screenshot shows a Jupyter Notebook interface with the following details:

- Title:** # Un document computationnel
- In [1]:**

```
from math import *  
print(pi)  
3.141592653589793
```
- Out [1]:** 3.141592653589793
- In [2]:**

```
import numpy as np  
N = 1000000  
x = np.random.uniform(size=N, low=0, high=pi/2)  
theta = np.random.uniform(size=N, low=0, high=pi/2)  
2/(sum((x+np.sin(theta))>1))/N
```
- Out [2]:** 3.14371986944998765
- In [3]:**

```
%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()
```
- Out [3]:** A histogram showing a normal distribution centered at approximately 100.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

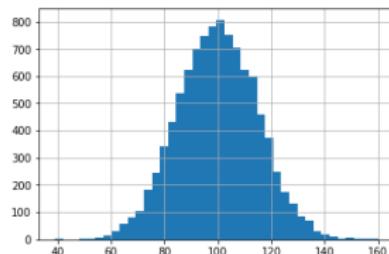
3.141592653589793

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
```

3.14371986944998765

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

The screenshot shows a Jupyter Notebook interface with three code cells:

- In [1]:** A cell containing Python code to print pi, followed by its value: 3.141592653589793. Above the code, a note says "Mon ordinateur m'indique que \$\pi\$ vaut "approximativement"
- In [2]:** A cell containing Python code to calculate pi using Buffon's needle method. It includes a note: "Mais calculé avec la [méthode des aiguilles de Buffon](#) (https://fr.wikipedia.org/wiki/Aiguille_de_Buffon), on obtient une approximation :". The output is 3.1437198694998765.
- In [3]:** A cell containing Python code to generate a histogram of random numbers. The output is a histogram plot showing a bell-shaped curve centered around 100.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

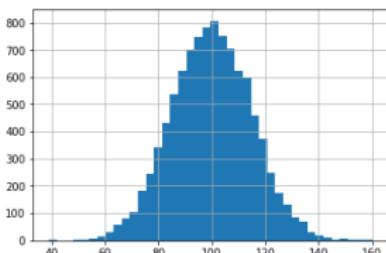
3.141592653589793

Mais calculé avec la [méthode des aiguilles de Buffon](#), on obtient 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
```

3.1437198694998765

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

A screenshot of a Jupyter Notebook interface. The top bar shows 'jupyter example_pi' and 'Python 3'. The notebook contains three code cells:

- In [1]:** `# Un document computationnel`. The output shows the value of pi as 3,141592653589793.
- In [2]:** A more complex calculation involving numpy and random modules, resulting in the value 3,1437198694098765.
- In [3]:** A histogram plot of a normal distribution with mu=100 and sigma=15, centered at approximately 100.

Annotations with red arrows point from the text 'Code' to the code blocks in the notebook.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

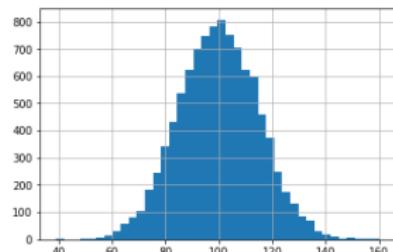
Mais calculé avec la **méthode des aiguilles de Buffon**, on obtiendrait comme approximation :

```
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

A screenshot of a Jupyter Notebook interface. The top bar shows 'jupyter example_pi' and 'Python 3'. The notebook contains three code cells:

- In [1]:** Prints the value of pi: `print(pi)` followed by `3.141592653589793`. A note below says "Mais calculé avec la `_approximation_` des éimpulles de Buffon (https://fr.wikipedia.org/wiki/Algille_de_Buffon), on obtiendrait comme `_approximation_`:
- In [2]:** Generates random points and calculates the ratio of points in a circle to total points to estimate pi.
- In [3]:** Plots a histogram of the generated data.

Annotations in red highlight the output of In [1] and In [2], and point to the histogram in In [3]. A large red arrow labeled "Résultats" points from the text annotations to the histogram.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

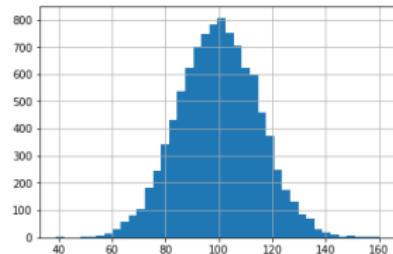
Mais calculé avec la **méthode des aiguilles de Buffon**, on obtiendrait comme approximation :

```
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
```

3.14371986949098765

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

The screenshot shows a Jupyter Notebook interface with three code cells:

- In [1]:** Prints the value of pi (3.141592653589793) and includes a note about calculating pi with Buffon's needle method.
- In [2]:** Generates random points (x, theta) and calculates an approximation of pi based on the ratio of points falling within a unit circle.
- In [3]:** Plots a histogram of x values, showing a bell-shaped distribution centered around 100.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

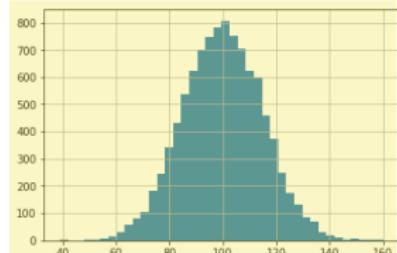
Mais calculé avec la méthode des [aiguilles de Buffon](#), on obtiendrait comme approximation :

```
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

Export

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

The screenshot shows a Jupyter Notebook interface with three code cells:

- In [1]:** Prints the value of pi (3.141592653589793) and includes a note about calculating pi with Buffon's needle method.
- In [2]:** Generates random points (x, y) and calculates the ratio of points below the unit circle to total points, which approximates pi.
- In [3]:** Plots a histogram of 100,000 random numbers between 0 and 1, showing a bell-shaped distribution centered at 0.5.

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut approximativement

3.141592653589793

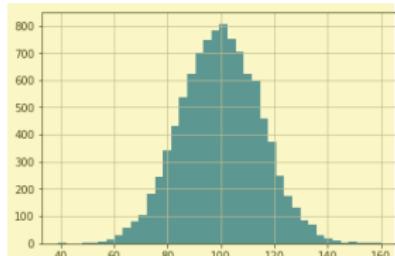
Mais calculé avec la **méthode des aiguilles de Buffon**, on obtiendrait comme approximation :

```
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

Export

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

The screenshot shows a Jupyter Notebook interface with several code cells and a plot.

- In [1]:**

```
# Un document computationnel
# Mon ordinateur m'indique que j'suis vers "approximativement"
from math import *
print(pi)
3.141592653589793
```

But calculé avec la `__method__` des (appelées de Buffet) `__str__(self)` (`http://hgweb.sagemath.org/wk2d/1/pickle_on_Buffer`, on estendrait comme `approximation_`)
- In [2]:**

```
import numpy as np
n = 1000000
x = np.random.uniform(0, low=0, high=1)
theta = np.random.uniform(0, low=0, high=np.pi/2)
if (x * np.sin(theta)) > 1 / n
```

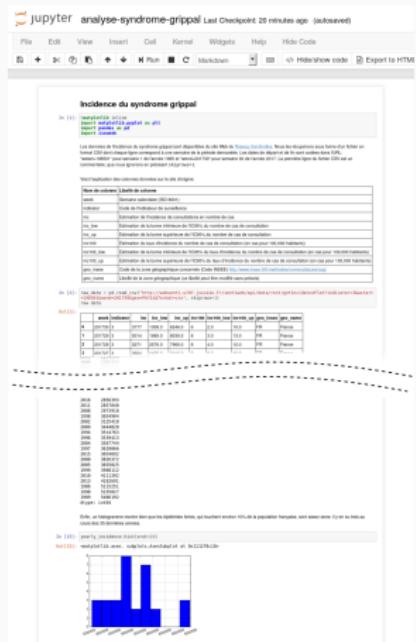
On peut inclure des formules mathématiques comme `$\frac{1}{\pi}\int_{-\pi}^{\pi} \frac{1}{1-x\cos(\theta)} d\theta$` et l'interpréteur Python nous donnera le résultat correspondant, dessiné ici il suffit de voir `!wolfram` (si ce n'est une constante de normalisation...).
- In [3]:**

```
%matplotlib inline
import matplotlib.pyplot as plt
Nn, sigma = 100, 33
x = np.random.normal(0, sigma, Nn)
plt.hist(x, 40)
plt.title('Normal distribution')
plt.show()
```

A histogram showing a normal distribution curve with a peak at approximately 0, ranging from -30 to 30 on the x-axis and 0 to 800 on the y-axis.

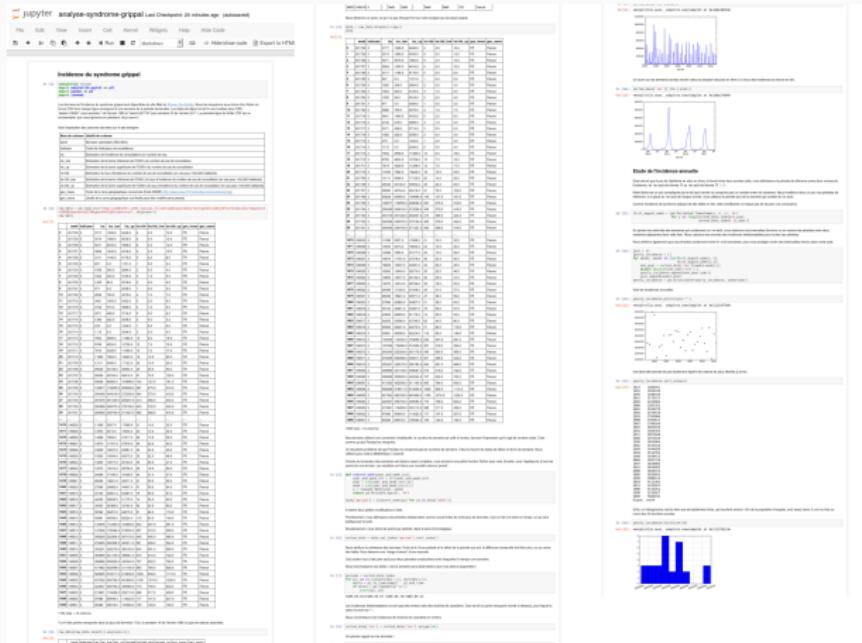
Tool 1 BIS: WORKFLOWS

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



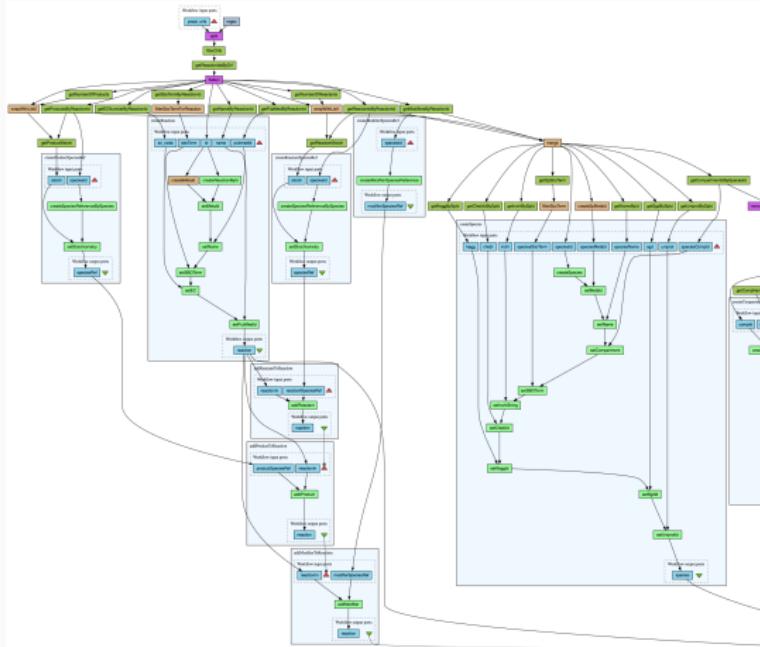
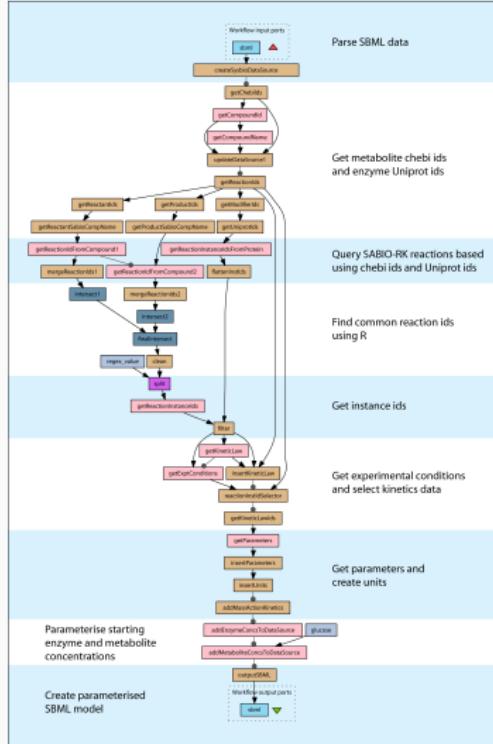
TOOL 1 BIS: WORKFLOWS

Notebooks are no panacea and do not help developing clean code

The image displays a 4x3 grid of Jupyter Notebook screenshots, each illustrating a different step or aspect of a workflow for analyzing color data and training a machine learning model.

- Row 1:**
 - Extracting Color Names by Web Image Search:** A cell showing Python code to extract color names from a web image using the `colornames` library.
 - Color names from chromaticity coordinates:** A cell showing Python code to convert chromaticity coordinates to color names using a pre-trained model.
 - Dimensionality reduction and model results:** A heatmap titled "Chromaticity plane and chromaticity model results, dimensionality reduced and filtered data" showing a complex pattern of colored regions.
- Row 2:**
 - Model building:** A cell showing the construction of a linear regression model to predict color names based on chromaticity coordinates.
 - Training the model:** A cell showing the training of the linear regression model using the `train_test_split` function.
 - Analysis:** A cell showing the results of the trained model's predictions on different test sets.
- Row 3:**
 - Plotting the training data:** A scatter plot titled "Chromaticity distribution of training data" showing points colored by their predicted names.
 - Plotting the test data:** A scatter plot titled "Chromaticity distribution of test data" showing points colored by their predicted names.
 - Prediction error vs. Training sample variance:** A scatter plot titled "Prediction error vs. Training sample variance" showing the relationship between prediction error and training sample variance.
- Row 4:**
 - Extracting raw data:** A cell showing the extraction of raw data from a CSV file.
 - Resampling images:** A cell showing the resampling of images to a uniform resolution of 100x100 pixels.
 - Conclusion:** A cell concluding the project and encouraging the reader to explore further.

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

CONTROLLING SOFTWARE ENVIRONMENT

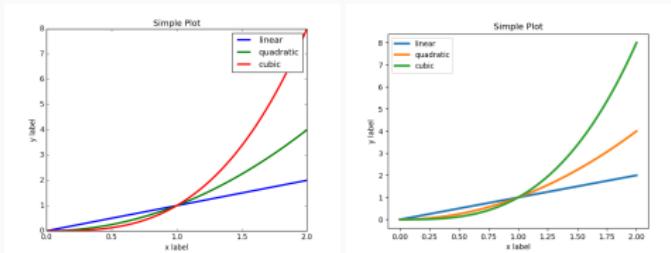
ARGH... DAMNED COMPUTERS

- 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)

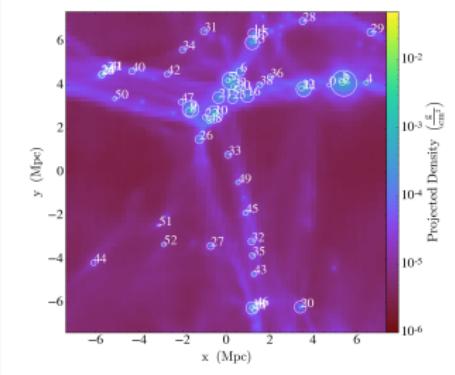
Significant differences in volume and cortical thickness were revealed across FreeSurfer versions. In addition, less pronounced differences were found between the Mac and HP workstations and between Mac OSX 10.5 and OSX 10.6.

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

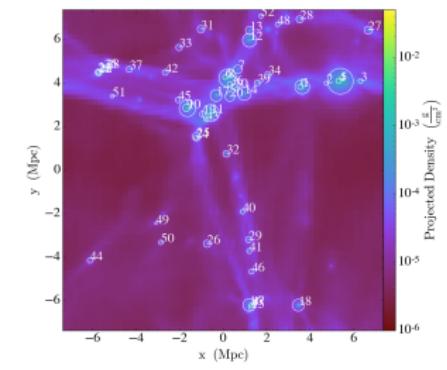


BACKWARDS COMPATIBILITY

- 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

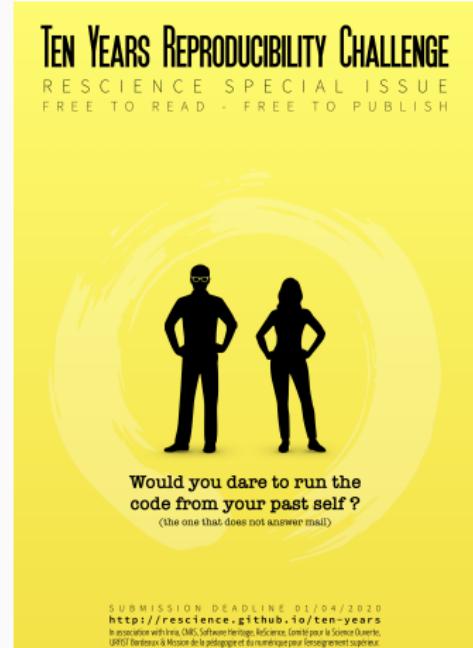


BACKWARDS COMPATIBILITY

- 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://rescence.github.io/ten-years/>

COMPLEX ECOSYSTEMS

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

3.1.2

COMPLEX ECOSYSTEMS

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

3.1.2

```
apt show python3-matplotlib
```

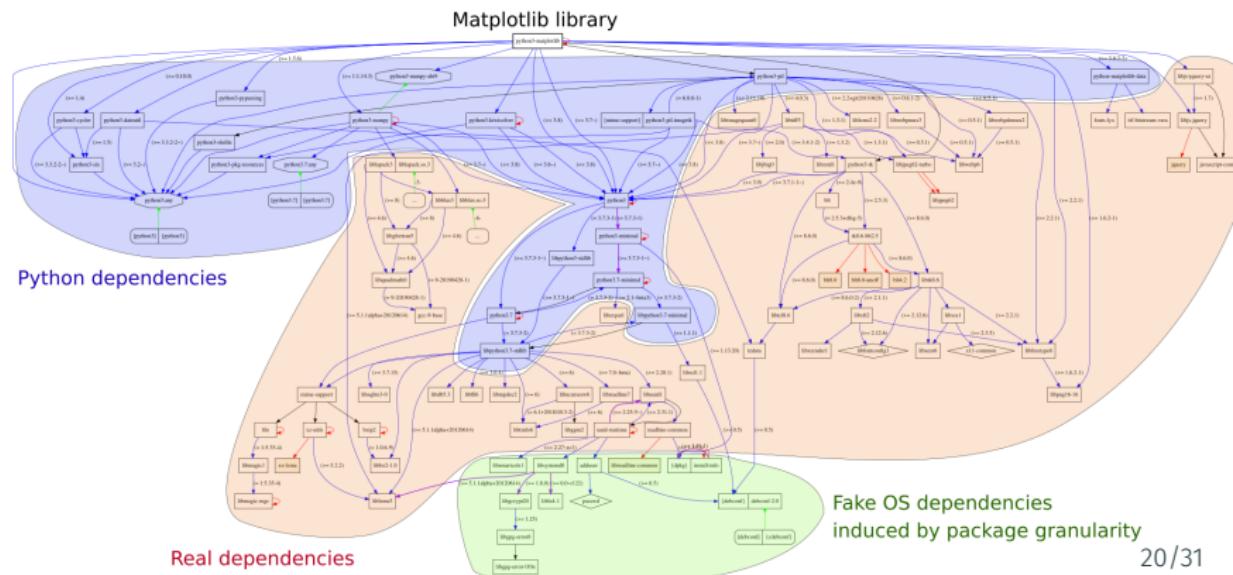
Package: python3-matplotlib
Version: 3.1.2-2
Priority: optional
Section: python
Source: matplotlib
Maintainer: Sandro Tosi <morph@debian.org>
Installed-Size: 15.3 MB
Depends: python3-dateutil, python-matplotlib-data (>= 3.1.2-2), python3-pyparsing,
six (>= 1.4), libjs-jquery, libjs-jquery-ui, python3-numpy (>= 1:1.16.0~rc1), py-
numpy-abi9, python3 (<< 3.9), python3 (>= 3.7~), python3-cycler (>= 0.10.0), py-
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, python3-20/31
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   methods

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      dplyr_1.0.2
 [5] assertthat_0.2.1  grid_3.6.3       R6_2.4.1         lifecycle_1.0.0
 [9] gtable_0.3.0      magrittr_1.5     scales_1.1.0      pillar_1.4.3
[13] rlang_0.4.4       lazyeval_0.2.2    glue_1.3.1       purrr_0.3.4
[17] munsell_0.5.0     compiler_3.6.3   pkgconfig_2.0.3   tibble_2.1.3
[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 ugly



Automatic tracking

TOOL 2: CONTAINERS AND PACKAGE MANAGERS

The good



The bad



The ugly



Automatic tracking

Containers

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

TOOL 2: CONTAINERS AND PACKAGE MANAGERS

The good



The bad



The ugly



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 ugly



Automatic tracking

Containers

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

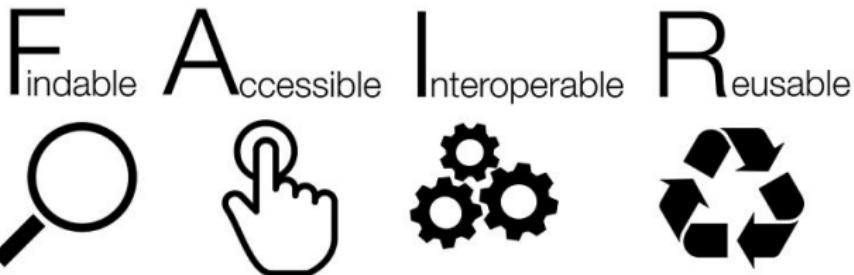
Package managers

- 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/>

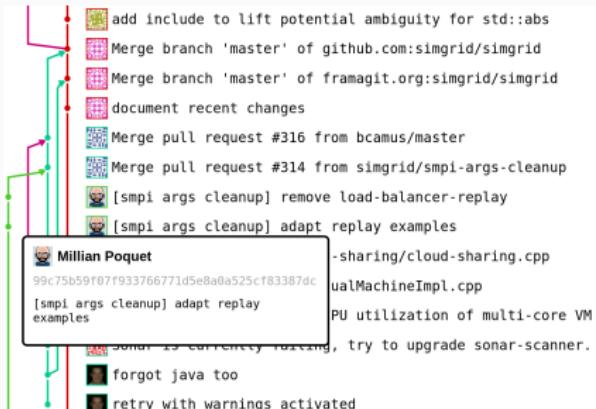
- "*Open as much as possible and close as much as necessary*"
- Management, publication, annotation (metadata), archiving
- Source code = specific data with specific consideration

Let's go beyond general principles!

TOOL 3: VERSION CONTROL AND FORGE

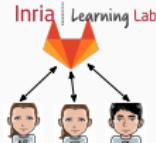
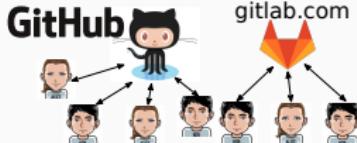
Git = version control

- 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
- Discontinued 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
- Discontinued forges: Code Space, Gitorious, Google code, Inria Gforge

Article archives



Data archives



figshare



Software Archive



Software Heritage

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](#), ...

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



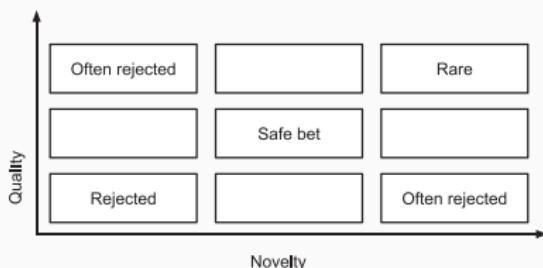
Joelle Pineau @ NeurIPS'18

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

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

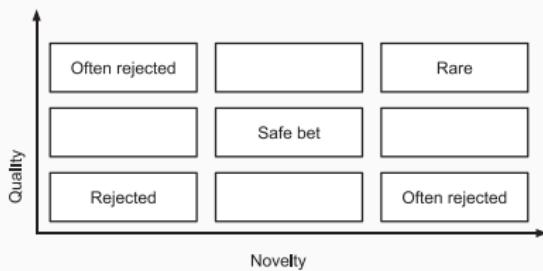
CHANGING ACADEMIC PRACTICES (PUBLISH OR PERISH)

- 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 ($\times 10$ 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



- 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**)



5. Open peer review (avoid collusion)

6. Open educational resources



**NO TRANSPARENCY
NO CONSENSUS**



RESOURCES AND ACKNOWLEDGMENTS



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

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

RESOURCES AND ACKNOWLEDGMENTS



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

- Loïc Desquillet, 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

