Good practices in the lab
Research documentation and electronic notebooks

Arnaud Legrand

Neurocampus Open Science Workshop
October 2023

What's expected from Science?

Goal

- Describe and explain **Nature** (human organizations and artificial objects as well)

What's expected from Science?

Goal

- Describe and explain **Nature** (human organizations and artificial objects as well)
- Find solutions to **society** needs and issues
  - Feed technological development and industry: health, energy, communications, ...
  - Enlighten public decisions: educate citizens, politics, regulation, ...

What's expected from Science?

Goal

- Describe and explain **Nature** (human organizations and artificial objects as well)
- Find solutions to **society** needs and issues
  - Feed technological development and industry: health, energy, communications, ...
  - Enlighten public decisions: educate citizens, politics, regulation, ...

What else distinguishes science from other human activities?

What's expected from Science?

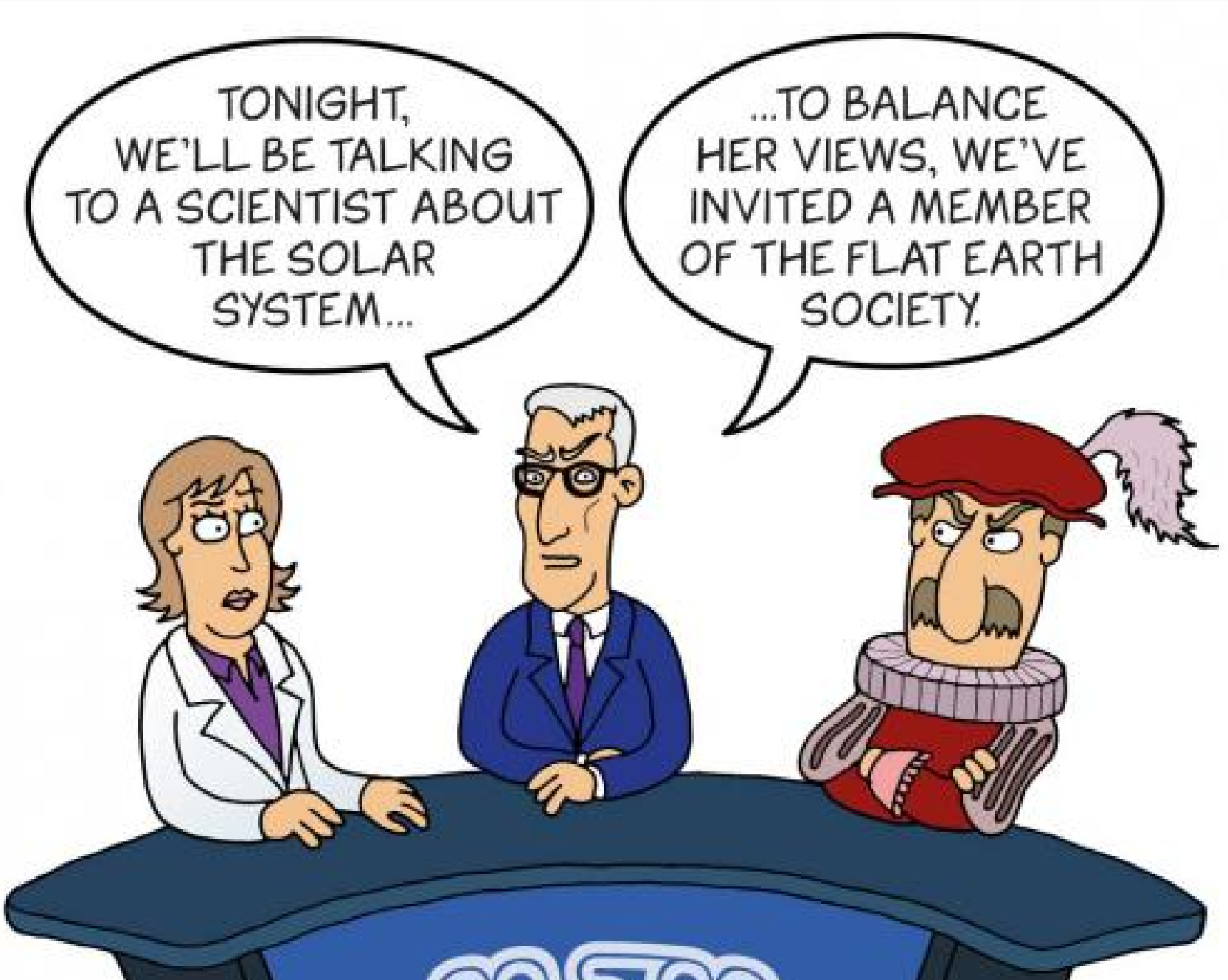
Goal

- Describe and explain **Nature** (human organizations and artificial objects as well)
- Find solutions to **society** needs and issues
  - Feed technological development and industry: health, energy, communications, ...
  - Enlighten public decisions: educate citizens, politics, regulation, ...

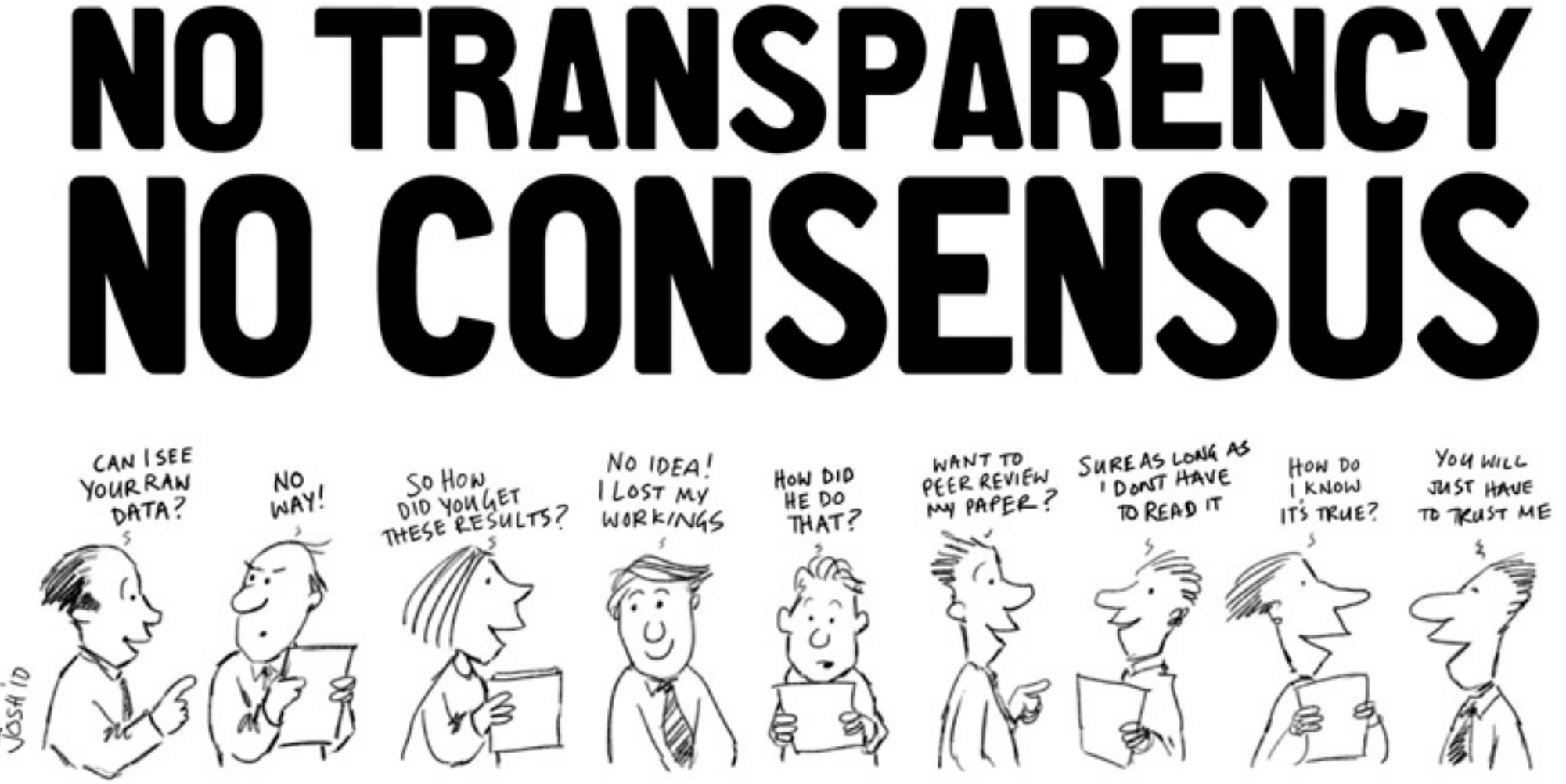
What else distinguishes science from other human activities?

Method

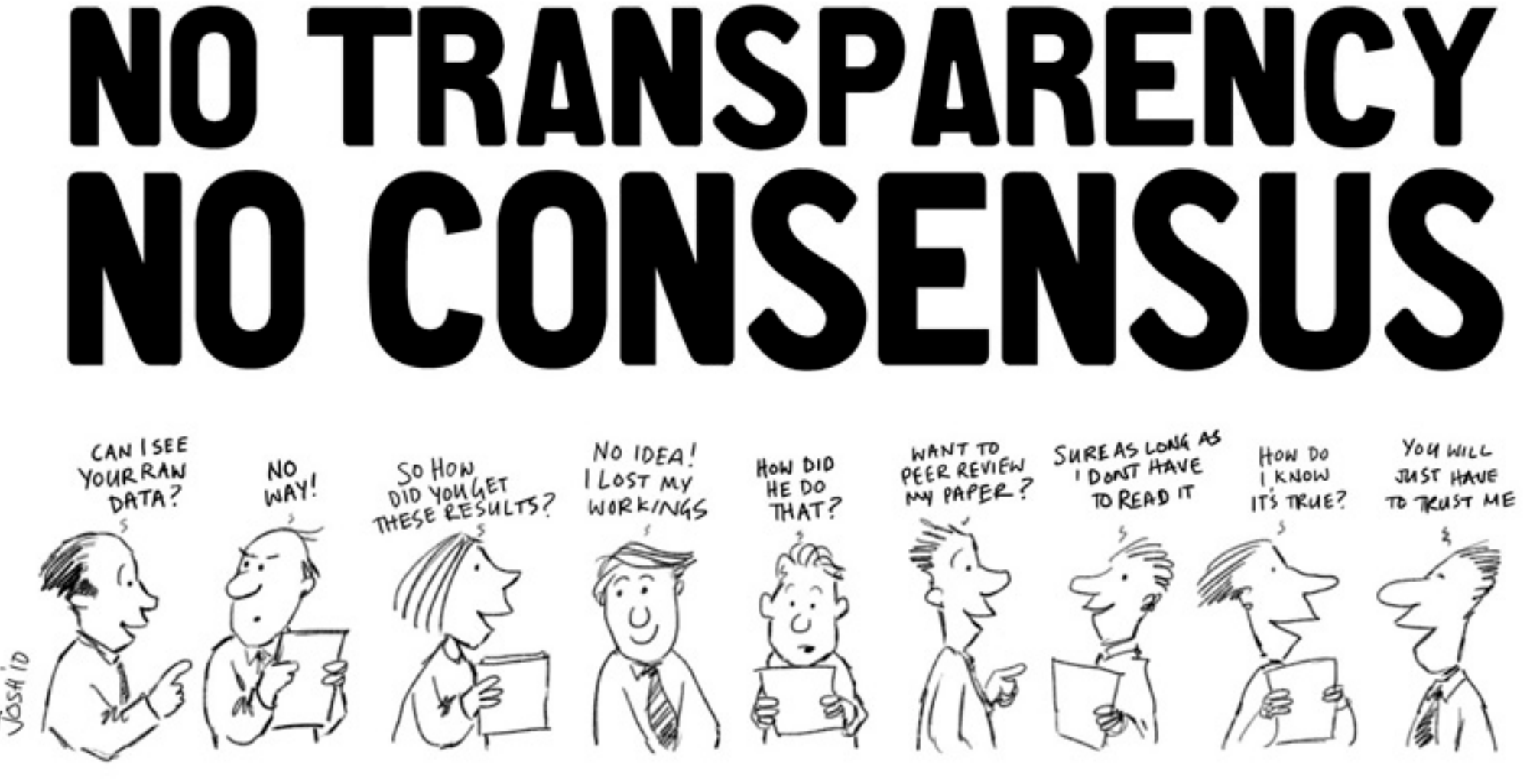
Scientific Consensus



Scientific Consensus



Scientific Consensus



Enlightening the society requires moral/methodological/technical warranties

Merton's four norms of science (1940)

Universalism Scientific validity is independent of the sociopolitical status/personal attributes of its participants (origin, gender, sexuality, religion, etc.)

Communism/communalism All scientists should have common ownership of scientific goods (intellectual property), to promote collective collaboration; secrecy is the opposite of this norm

- We say "Newton's law" to remember that Newton made the original discovery, but not because he has any property of this law

Disinterestedness Scientific institutions act for the benefit of a common scientific enterprise, rather than for the personal gain of individuals

Organized skepticism Scientific claims should be exposed to critical scrutiny (including reproduction) before being accepted

Merton's four norms of science (1940)

Universalism Scientific validity is independent of the sociopolitical status/personal attributes of its participants (origin, gender, sexuality, religion, etc.)

Communism/communalism All scientists should have common ownership of scientific goods (intellectual property), to promote collective collaboration; secrecy is the opposite of this norm

- We say "Newton's law" to remember that Newton made the original discovery, but not because he has any property of this law

Disinterestedness Scientific institutions act for the benefit of a common scientific enterprise, rather than for the personal gain of individuals

Organized skepticism Scientific claims should be exposed to critical scrutiny (including reproduction) before being accepted

Merton was not saying *"you should follow these norms"*, but rather

"this is the ~~behavior~~ belief I am observing in the scientific community"

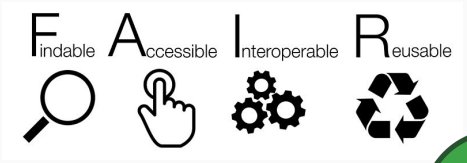
Transparent Research ∼ Open Science ?

Plan National pour la Science Ouverte (BSN ⇝ CoSO)

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

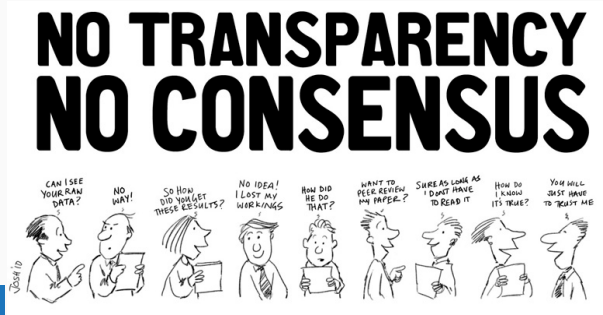
Main pillars:

1. Open access (started before 2000)
2. Open data



CC-BY-SA by default at CNRS!)

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



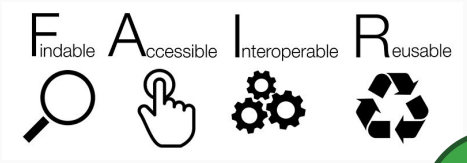
Transparent Research ∼ Open Science ?

Plan National pour la Science Ouverte (BSN ⇝ CoSO)

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

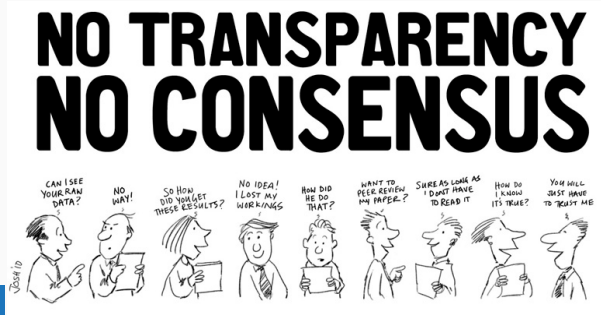
Main pillars:

1. Open access (started before 2000)
2. Open data



CC-BY-SA by default at CNRS!)

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



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

The scientific Method

Claude Bernard (1865) and Karl Popper (1934)

Observations $\xrightarrow{\text{induction}}$ theory building/proposal

$\xrightarrow{\text{repeated experiments}}$ theory *confirmation*

Claude Bernard (1865) and Karl Popper (1934)

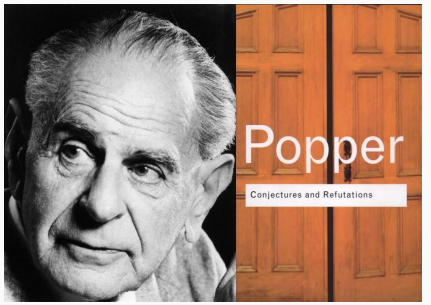
Observations $\xrightarrow{\text{induction}}$ theory building/proposal

$\xrightarrow{\text{repeated experiments}}$ theory *confirmation*

refutation attempts

(crucial experiment)

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

Claude Bernard (1865) and Karl Popper (1934)

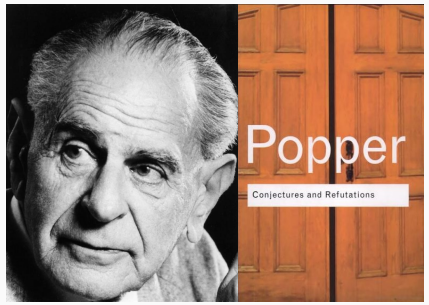
Observations $\xrightarrow{\text{induction}}$ theory building/proposal

$\xrightarrow{\text{repeated experiments}}$ theory *confirmation*

refutation attempts

(crucial experiment)

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, astrophysics, particle physics, biology, anthropology)

And in practice ?

Crucial experiment: what happens when an experiment whose results are incoherent with the theory?

And in practice ?

Crucial experiment: what happens when an experiment whose results are incoherent with the theory? Is it *real* ?

- Did I finally manage to break the *dominant theory*?
  - Others will want to confirm by repeating the experiment.
- Did I make a mistake?
  - Whoops

$\rightsquigarrow$ Careful inspection of the experimental process

And in practice ?

Crucial experiment: what happens when an experiment whose results are incoherent with the theory? Is it *real* ?

- Did I finally manage to break the *dominant theory*?
  - Others will want to confirm by repeating the experiment.
- Did I make a mistake?
  - Whoops

$\rightsquigarrow$ Careful inspection of the experimental process

The role of statistics: what are the chances that I observe this by chance (i.e., there is actually no relation but I observe one)?

- Observation $\neq$ experiments (HARKING)
- Selective reporting (p-hacking)

$\rightsquigarrow$ Objectivity? Failed/successful experiments are often carefully selected...

And in practice ?

Crucial experiment: what happens when an experiment whose results are incoherent with the theory? Is it *real* ?

- Did I finally manage to break the *dominant theory*?
  - Others will want to confirm by repeating the experiment.
- Did I make a mistake?
  - Whoops

$\rightsquigarrow$ Careful inspection of the experimental process

The role of statistics: what are the chances that I observe this by chance (i.e., there is actually no relation but I observe one)?

- Observation $\neq$ experiments (HARKING)
- Selective reporting (p-hacking)

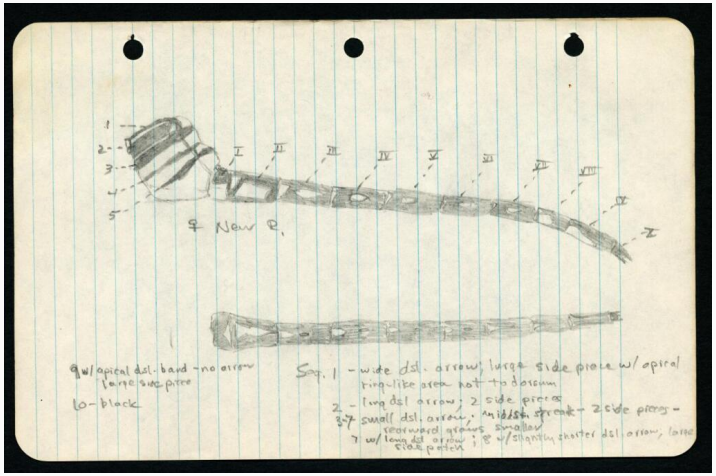
$\rightsquigarrow$ Objectivity? Failed/successful experiments are often carefully selected...

These mistakes require **more than public numbers and raw data**

Notebooks

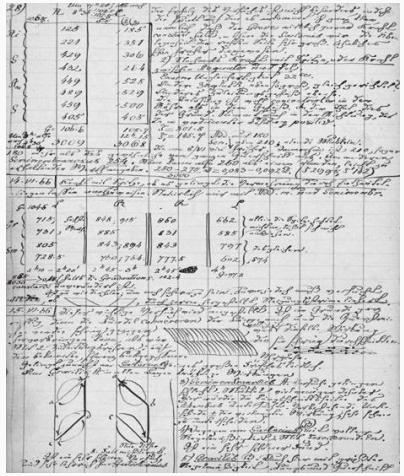
Field and Laboratory notebooks

Social Sciences, Ecology, Biology



Robert Henry Gibbs, Jr.,
ichthyologist (1929 – 1988)

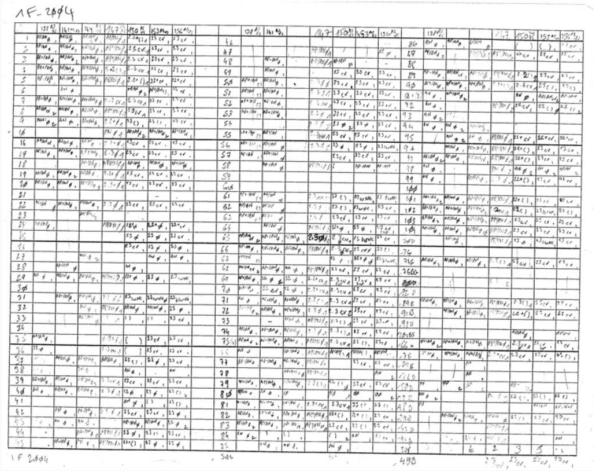
Chemistry, Physics, Biology



Emil Heinrich du Bois-Reymond,
electrophysiology (1818 – 1896)

Field and Laboratory notebooks

Social Sciences, Ecology, Biology



Marie-Claude Quidoz, Centre d'Ecologie Fonctionnelle et Evolutive

Chemistry, Physics, Biology

Contemporary laboratory notebook in Neurology

Essentially no evolution throughout the last century

Pros and Cons

Pros

- Intuitive
- Convenient (lightweight)
- Sketching, equations, thoughts
- *Difficult* to temper

Cons

- Attachments
- Large amount of information
- Indexing and Searching
- No sharing nor backup

Pros and Cons

Pros

- Intuitive
- Convenient (lightweight)
- Sketching, equations, thoughts
- *Difficult* to temper

Cons

- Attachments
- Large amount of information
- Indexing and Searching
- No sharing nor backup

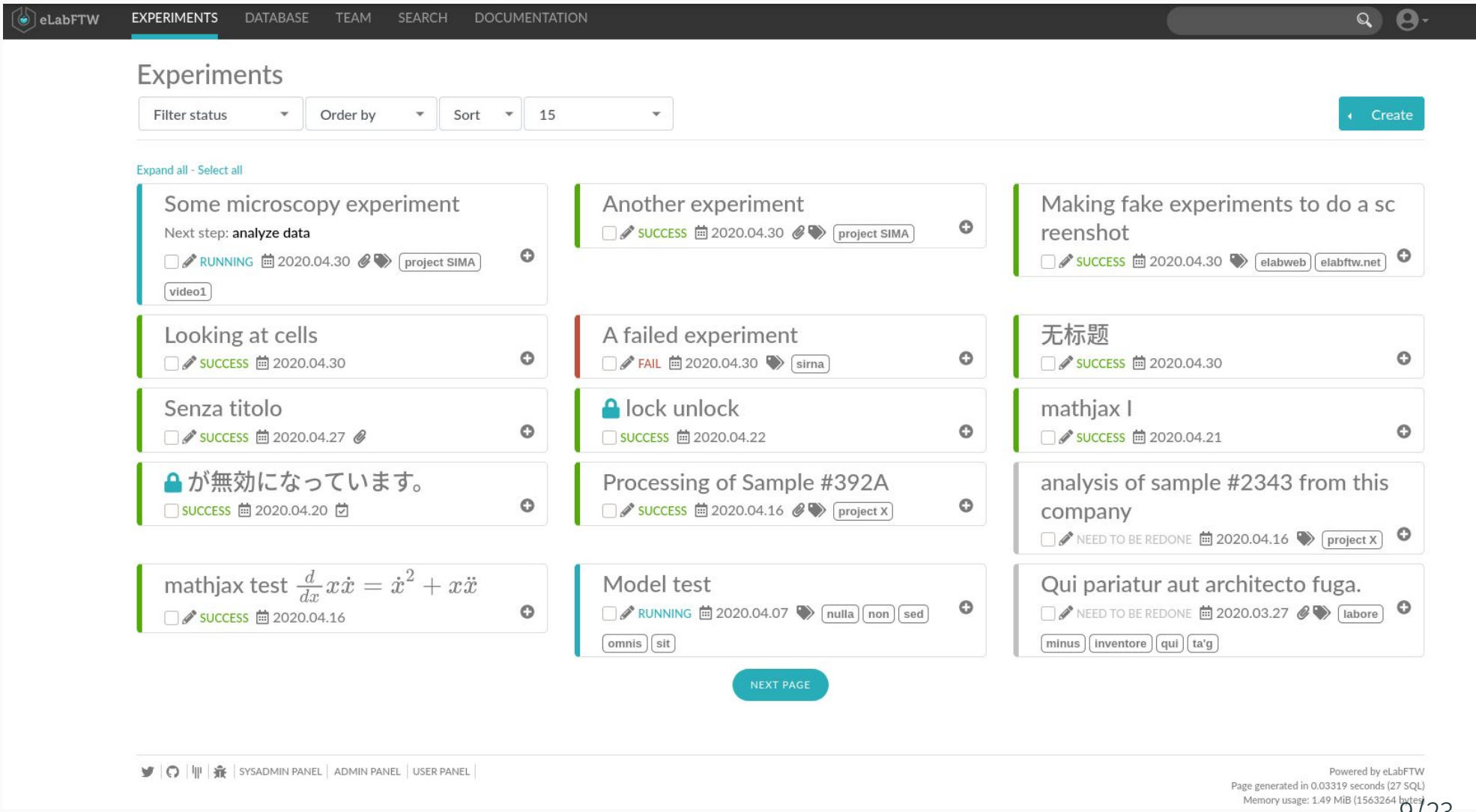
Toward digital notebooks

Electronic Laboratory Notebooks and Laboratory Information Management Systems

- Batch and Sample management
- Instrument integration
- Stock management (e.g., chemical substances)
- Native handling of digital objects

ELN examples 1/2: eLabFTW

eLabFTW: Chemistry @ Lyon (web server w. PHP)



ELN examples 1/2: eLabFTW

eLabFTW: Chemistry @ Lyon (web server w. PHP)

eLabFTW EXPERIMENTS DATABASE TEAM SEARCH DOCUMENTATION

Experiments

Back to listing

2020.04.30 Success Team User

project SIMA video1

Some microscopy experiment

Next step: analyze data

Goal:

Show what an experiment looks like for elabftw.net.

Sample #	Concentration	Absorbance	Density	Something
1	20 µM	0.239	3.4	102
2	30 µM	0.193	4.1	1339

You can have colors too.

And use **bold**, *italics*, underline like you would do in a normal text processing application. Have some mathjax: {x^y}^z is x^{y^z}, and x^{y^z} is x^{y^z}.

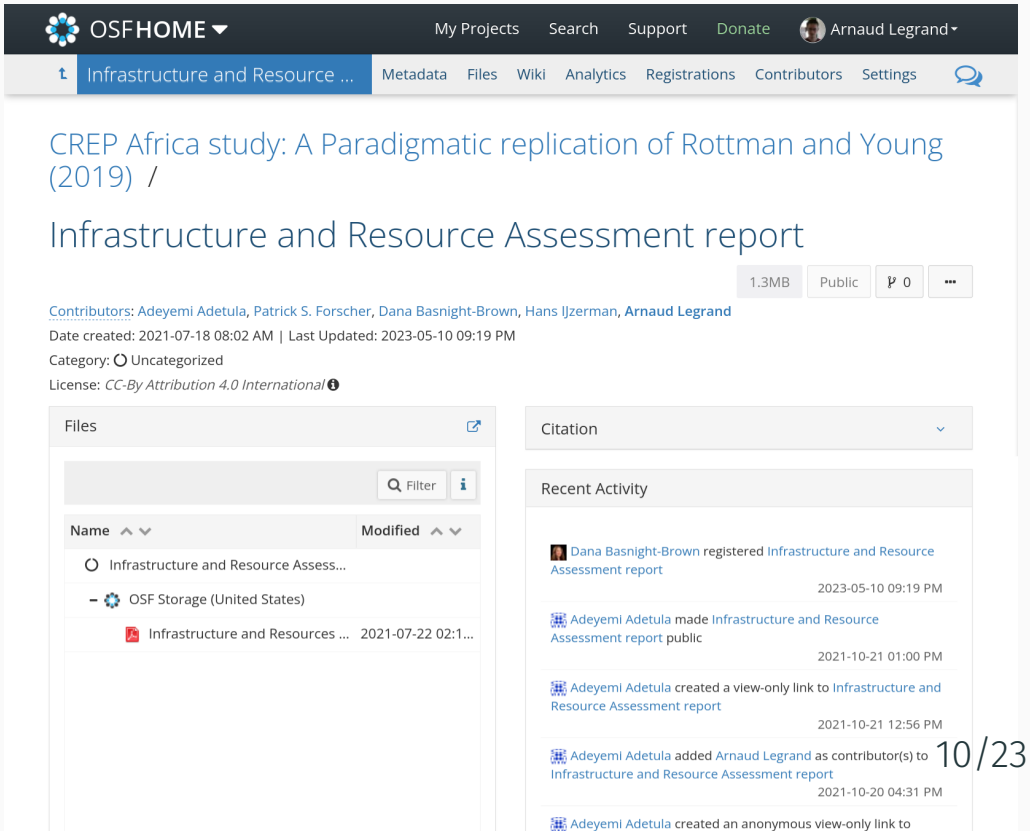
This is a list:

- List item 1
- List item 2
- Now show an image in the text

ELN examples 2/2: OSF.io

A cloud to ease collaboration at the team/lab level

- Shared documents (articles, data, spreadsheets ...)
- Built-in version control
- Integration with Zotero/Mendeley
- Preprints, DMP, Pre-registration, ...
- Social network



ELN examples 2/2: OSF.io

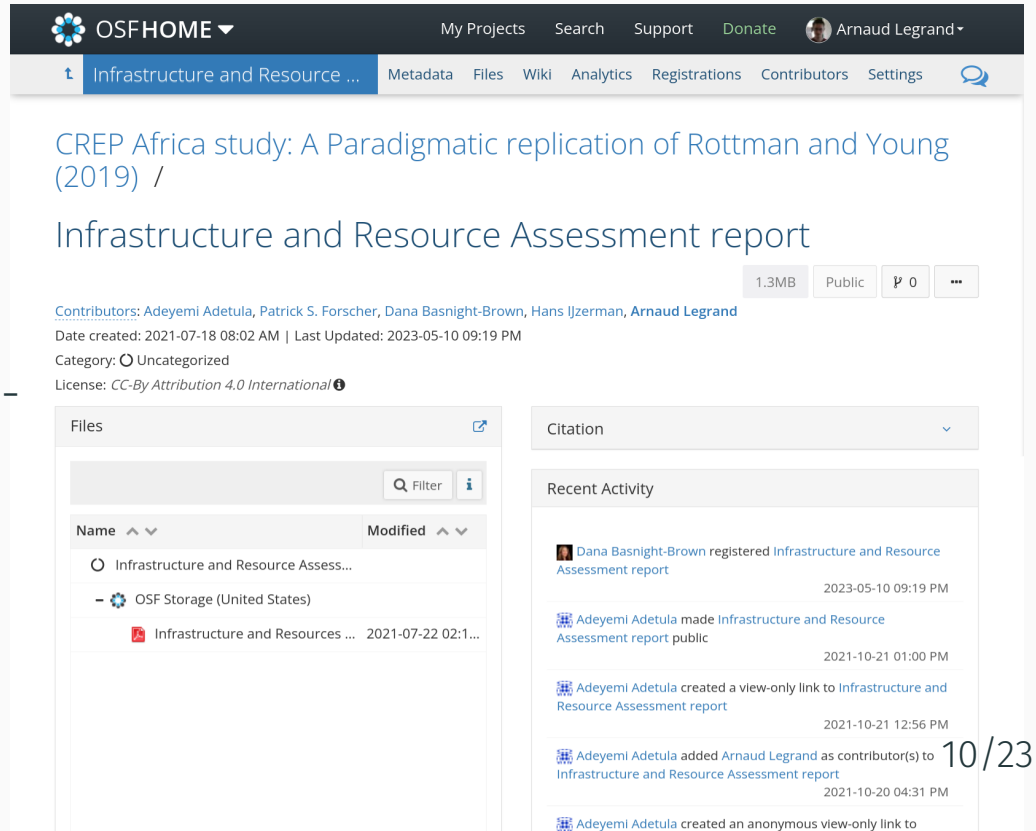
A cloud to ease collaboration at the team/lab level

- Shared documents (articles, data, spreadsheets ...)
- Built-in version control
- Integration with Zotero/Mendeley
- Preprints, DMP, Pre-registration, ...
- Social network

CENTER FOR OPEN SCIENCE

But also

- Dropbox, Google drive storage (privacy? sovereignty?)
- Google/Facebook/ORCID/Institution authentication
- **Locks you in** (huge machinery)
- Nothing done for computations



Damned computers...

How computers broke science –

and what we can do about it

> *Most modern science is so complicated, and most journal articles so brief, it's impossible for the article to include details of many important methods and decisions made by the researcher as he analyzed his data on his computer. How, then, can another researcher judge the reliability of the results, or reproduce the analysis?*

– Ben Marwick,
The conversation, 2015

Damned computers...

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

> *Most modern science is so complicated, and most journal articles so brief, it's impossible for the article to include details of many important methods and decisions made by the researcher as he analyzed his data on his computer. How, then, can another researcher judge the reliability of the results, or reproduce the analysis?*

– Ben Marwick,
The conversation, 2015

When moving to the digital world, the rigor is often completely lost.

- Point-and-click procedures are rampant but hinder reproducibility.
- Spreadsheets are generalized and intensively used in biology:
  - `Membrane-Associated Ring Finger (C3HC4) 1`,
    `E3 Ubiquitin Protein Ligase` → `MARCH1` → 2016-03-01 → 1456786800
  - `2310009E13` → 2.31E+19
- And more recently, we had the COVID tracing failure.

Not even mentioning non-mature software and blackbox machine learning procedures

Notes and Documenting

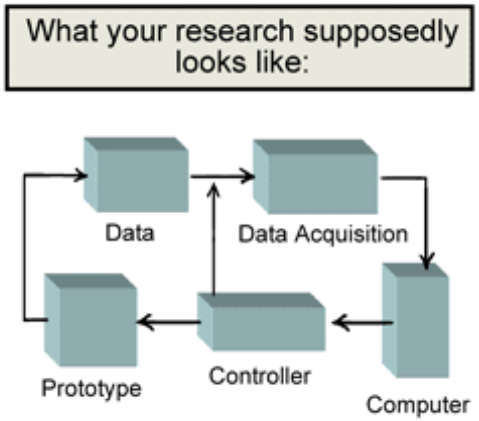


Figure 1. Experimental Diagram

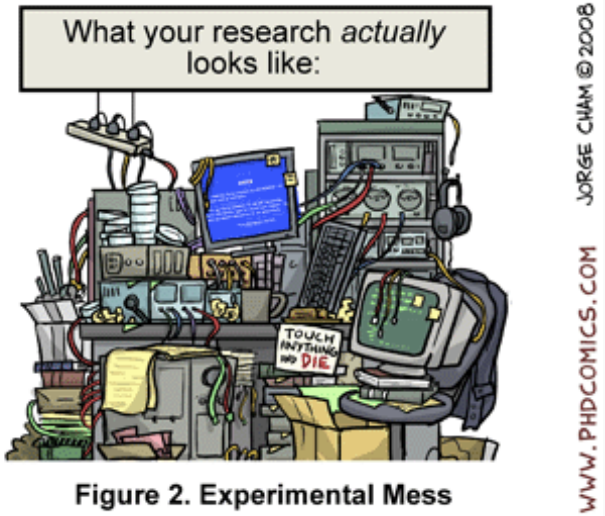


Figure 2. Experimental Mess

Main Challenges for a computational scientist

```
my_code --cfg=magical_param:0.94572 '*.dat' --output foo.csv
```

Tracking code version

- `my_code` is revision `21b95ecfa0911d6ca87668482b11ab9498edd8f3`

Main Challenges for a computational scientist

```
1  my_code --cfg=magical_param:0.94572 '*.dat' --output foo.csv
```

Tracking code version

- `my_code` is revision `21b95ecfa0911d6ca87668482b11ab9498edd8f3`

Tracking software environment

- `my_code` depends on a dozen of libraries, which depend on dozens of libraries
- `my_code` was compiled with `clang 1:9.0-49.1` and `-O3 -funroll-loops -fno-strict-aliasing -finline-functions ...`

Main Challenges for a computational scientist

```
my_code --cfg=magical_param:0.94572 '*.dat' --output foo.csv
```

Tracking code version

- `my_code` is revision `21b95ecfa0911d6ca87668482b11ab9498edd8f3`

Tracking software environment

- `my_code` depends on a dozen of libraries, which depend on dozens of libraries
- `my_code` was compiled with `clang 1:9.0-49.1` and `-O3 -funroll-loops -fno-strict-aliasing -finline-functions ...`

Tracking parameters and data

- `*.dat`? Ooh, you ran this in `data/2091293-AJXQ37`?
- Wasn't `mymap.dat` updated since then?
- That was for `foo.csv`. What about `bar.csv`? Is it reproducible?

Main Challenges for a computational scientist

```
my_code --cfg=magical_param:0.94572 '*.dat' --output foo.csv
```

Tracking code version

- `my_code` is revision `21b95ecfa0911d6ca87668482b11ab9498edd8f3`

Tracking software environment

- `my_code` depends on a dozen of libraries, which depend on dozens of libraries
- `my_code` was compiled with `clang 1:9.0-49.1` and `-O3 -funroll-loops -fno-strict-aliasing -finline-functions ...`

Tracking parameters and data

- `*.dat`? Ooh, you ran this in `data/2091293-AJXQ37`?
- Wasn't `mymap.dat` updated since then?
- That was for `foo.csv`. What about `bar.csv`? Is it reproducible?

Tracking the process (on short/long term)

- Why did I run this? What did I learn from it? I remember doing this but when?

Main Challenges for a computational scientist

```
my_code --cfg=magical_param:0.94572 '*.dat' --output foo.csv
```

Tracking code version

- `my_code` is revision `21b95ecfa0911d6ca87668482b11ab9498edd8f3`

Tracking software environment

- `my_code` depends on a dozen of libraries, which depend on dozens of libraries
- `my_code` was compiled with `clang 1:9.0-49.1` and `-O3 -funroll-loops -fno-strict-aliasing -finline-functions ...`

Tracking parameters and data

- `*.dat`? Ooh, you ran this in `data/2091293-AJXQ37`?
- Wasn't `mymap.dat` updated since then?
- That was for `foo.csv`. What about `bar.csv`? Is it reproducible?

Tracking the process (on short/long term)

- Why did I run this? What did I learn from it? I remember doing this but when?

Handle complex sequences and reuse results (leverage supercomputers)

Tool 1: Computational Notebooks (Litterate Programming)

Un document computationnel

Mon ordinateur m'indique que π vaut *approximativement*

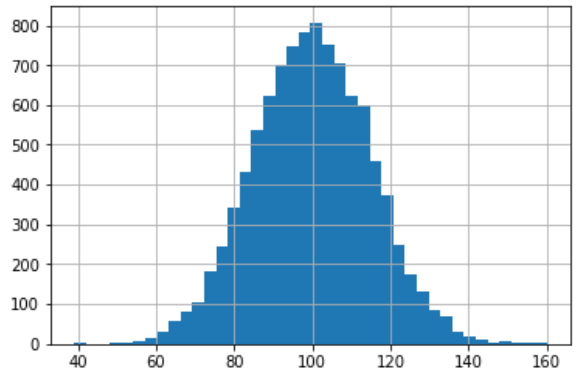
```
3.141592653589793
```

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... ☺).
```

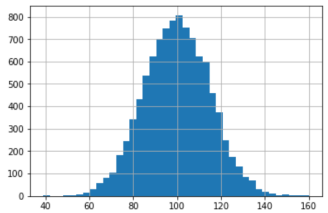
In [3]:

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*

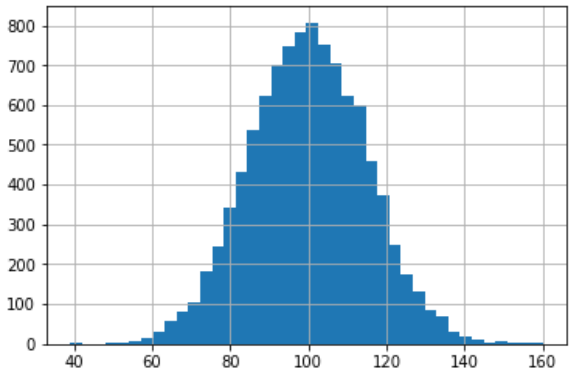
```
3.141592653589793
```

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
```

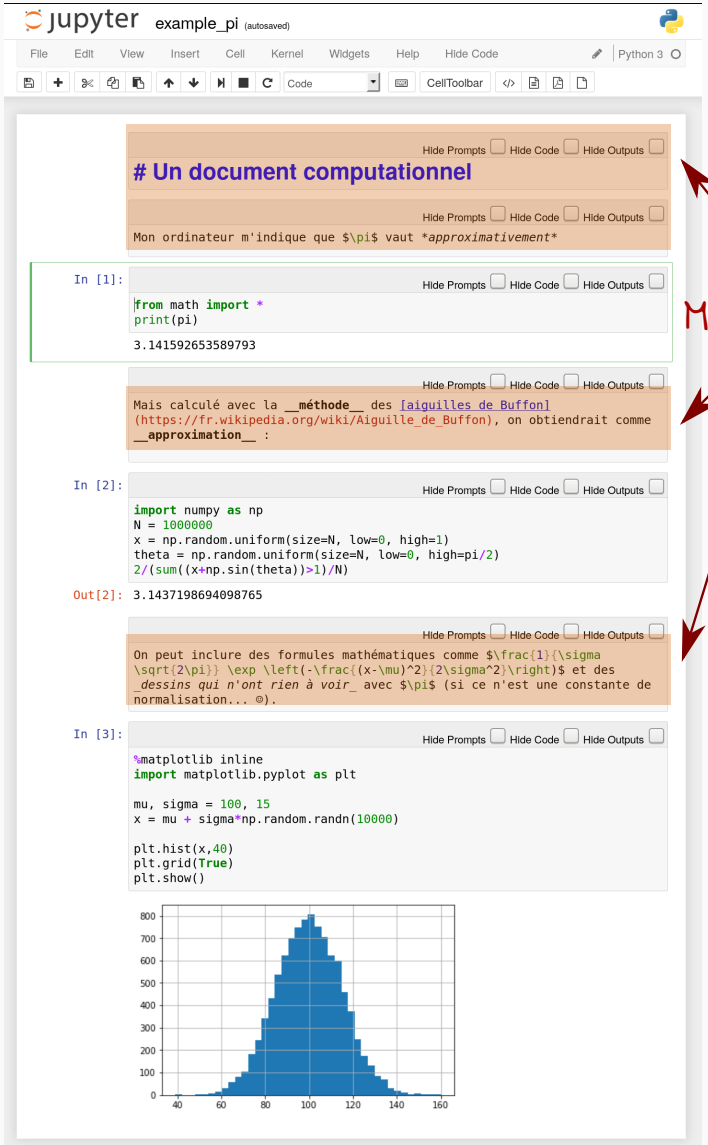
On peut inclure des formules mathématiques comme $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des *dessins qui n'ont rien à voir* avec π (si ce n'est une constante de normalisation... ☺).



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*

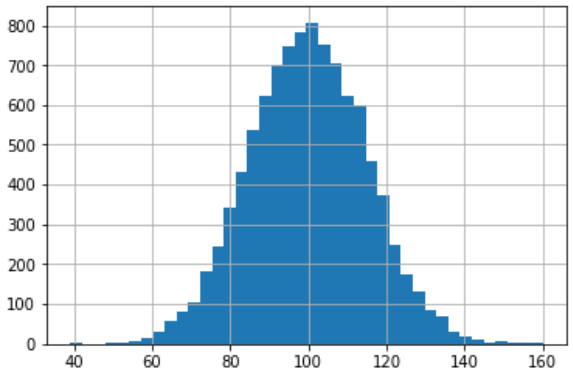
```
3.141592653589793
```

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)
```

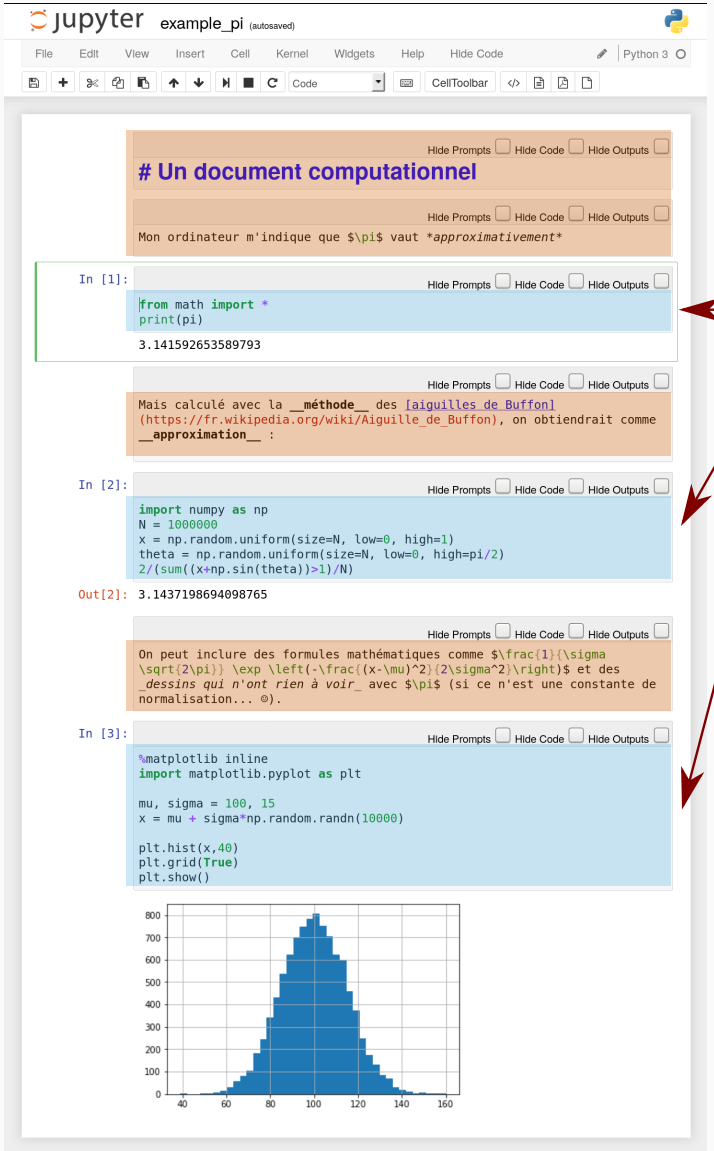
```
3.1437198694098765
```

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*

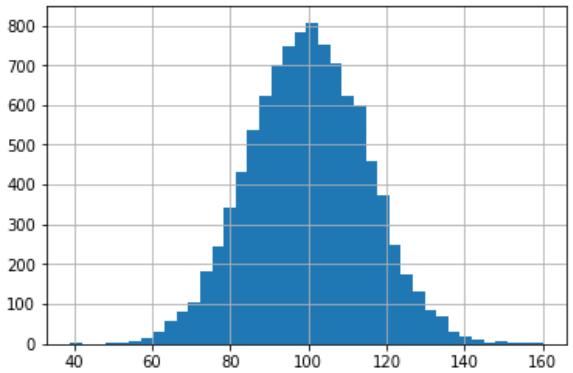
```
3.141592653589793
```

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

Document final

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

Mon ordinateur m'indique que π vaut *approximativement*

In [1]:

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

```
3.141592653589793
```

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

In [2]:

```
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

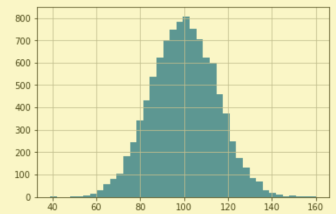
On peut inclure des formules mathématiques comme $\frac{1}{\sigma \sqrt{2\pi}} \exp \left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ et des *_dessins qui n'ont rien à voir_* avec π (si ce n'est une constante de normalisation... ☺).

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()
```



Résultats

Un document computationnel

Mon ordinateur m'indique que π vaut *approximativement*

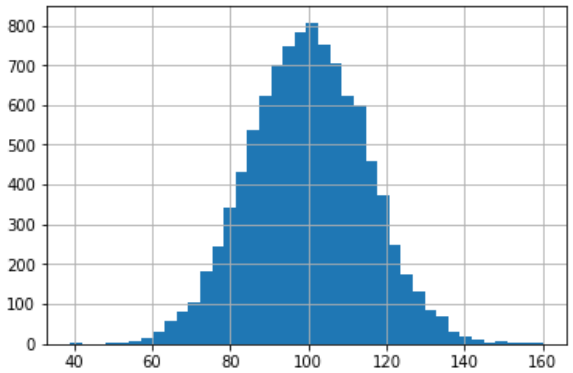
```
3.141592653589793
```

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)
```

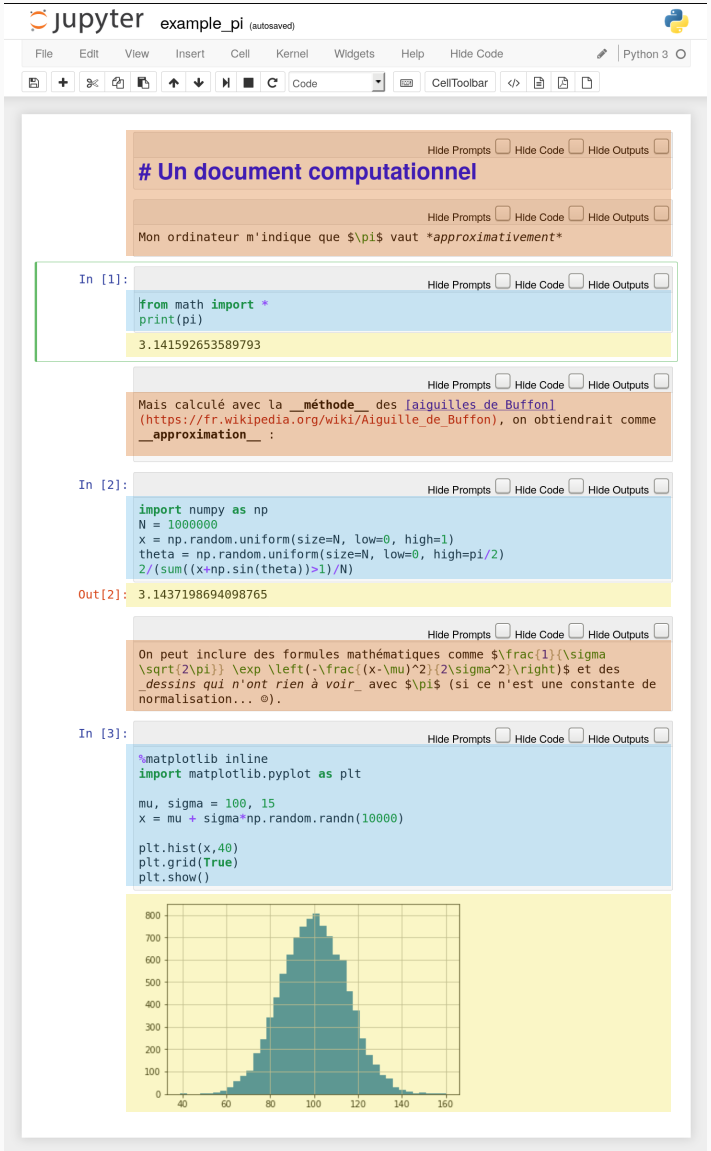
```
3.1437198694098765
```

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



Export

Document final

Un document computationnel

Mon ordinateur m'indique que π vaut *approximativement*

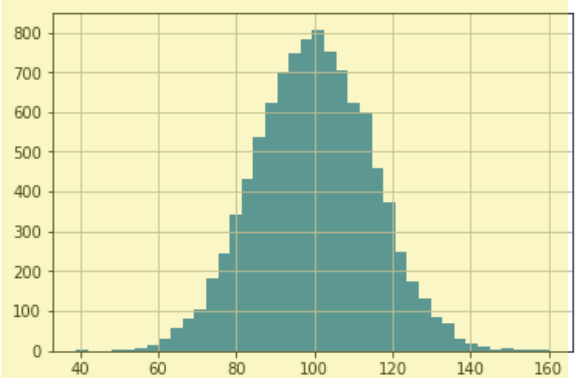
```
3.141592653589793
```

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... ☺).
```

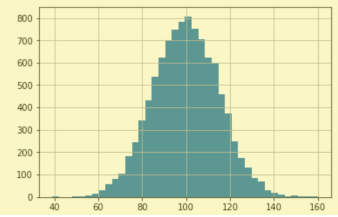
In [3]:

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*

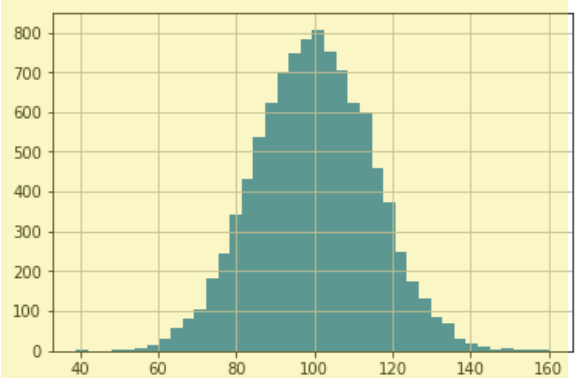
```
3.141592653589793
```

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... ☺).



Quick demo of Jupyter, Rstudio, and Org-Mode

(python)

Pros Python/R, friendly, portable (web browser, client/server)

Cons
- Installation, software dependencies (minimal-notebook $\approx$ 440Mb)
- Limited control on typography (unless using *Rube Goldberg* machines like quarto)

Quick demo of Jupyter, Rstudio, and Org-Mode

(python)

Pros Python/R, friendly, portable (web browser, client/server)

Cons
- Installation, software dependencies (minimal-notebook $\approx$ 440Mb)
- Limited control on typography (unless using *Rube Goldberg* machines like quarto)

(java/R)

Pros R/Python, friendly, portable, real IDE, Good typography control

Cons
- Installation, software dependencies (rocker/rstudio $\approx$ 550Mb)

Quick demo of Jupyter, Rstudio, and Org-Mode

(python)

Pros Python/R, friendly, portable (web browser, client/server)

Cons
- Installation, software dependencies (minimal-notebook ≈ 440Mb)
- Limited control on typography (unless using *Rube Goldberg* machines like quarto)

(java/R)

Pros R/Python, friendly, portable, real IDE, Good typography control

Cons
- Installation, software dependencies (rocker/rstudio ≈ 550Mb)

(emacs-lisp)

Pros R, Python, Perl, Ruby, C, Java, ...
- Pure text, Good typography control, "Lightweight", Programmable editor

Cons
- No default configuration ⇝rough and steep learning curve
- Big machinery: silex/emacs-alpine-ci ≈ 240Mb even though flycheck/emacs-cask ≈ 80Mb

Different reproducibility philosophies

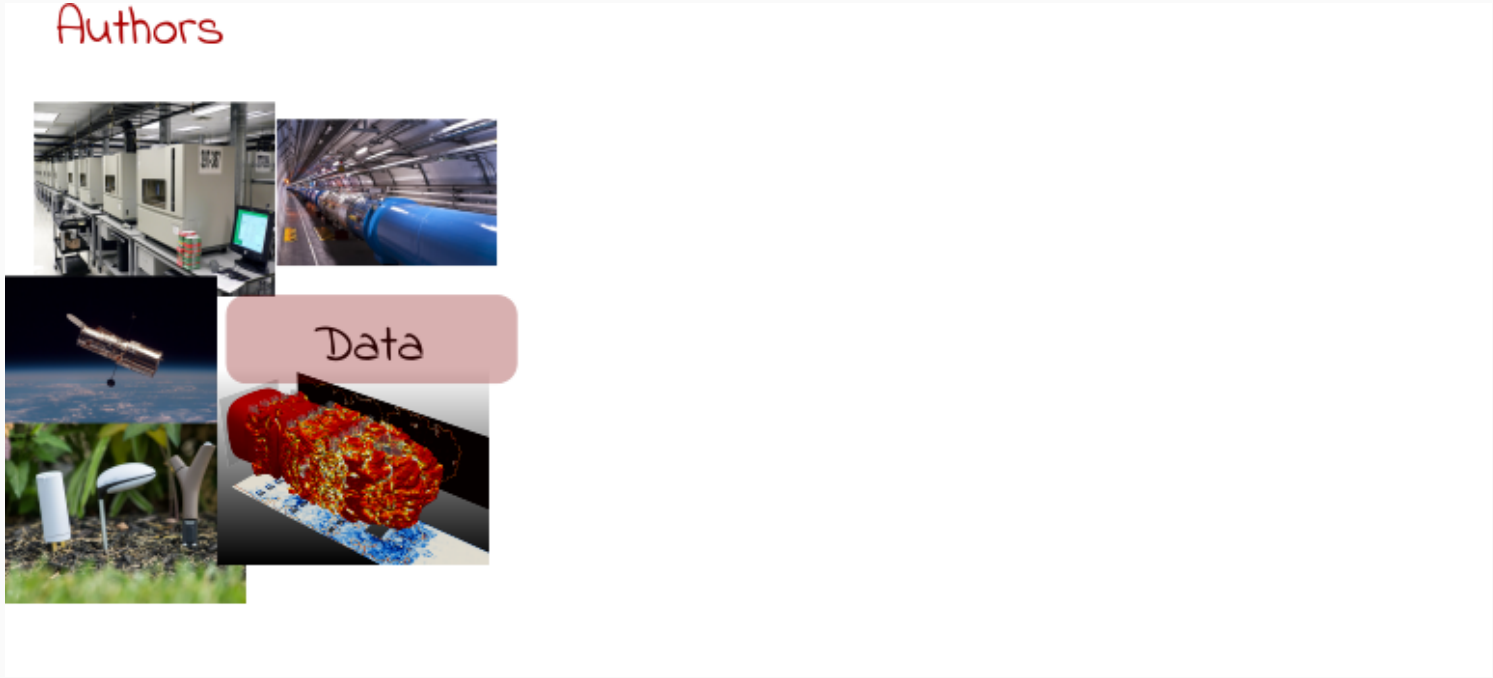
Why and how to use a computational notebook?

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

- A reproducible document links raw data with final figures
- What's at stake: verification and reuse by others

Why and how to use a computational notebook?

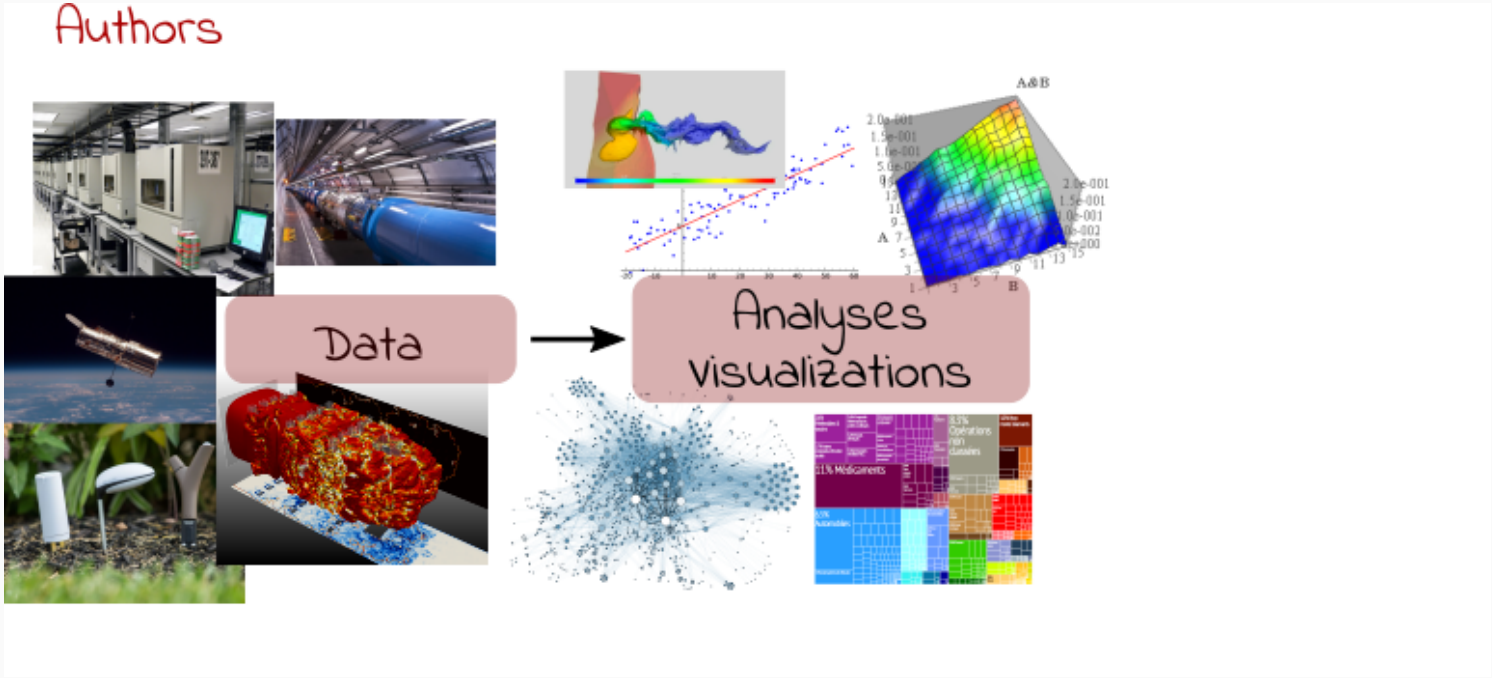
The processing steps between raw observations and findings have gotten increasingly numerous and complex



- A reproducible document links raw data with final figures
- What's at stake: verification and reuse by others

Why and how to use a computational notebook?

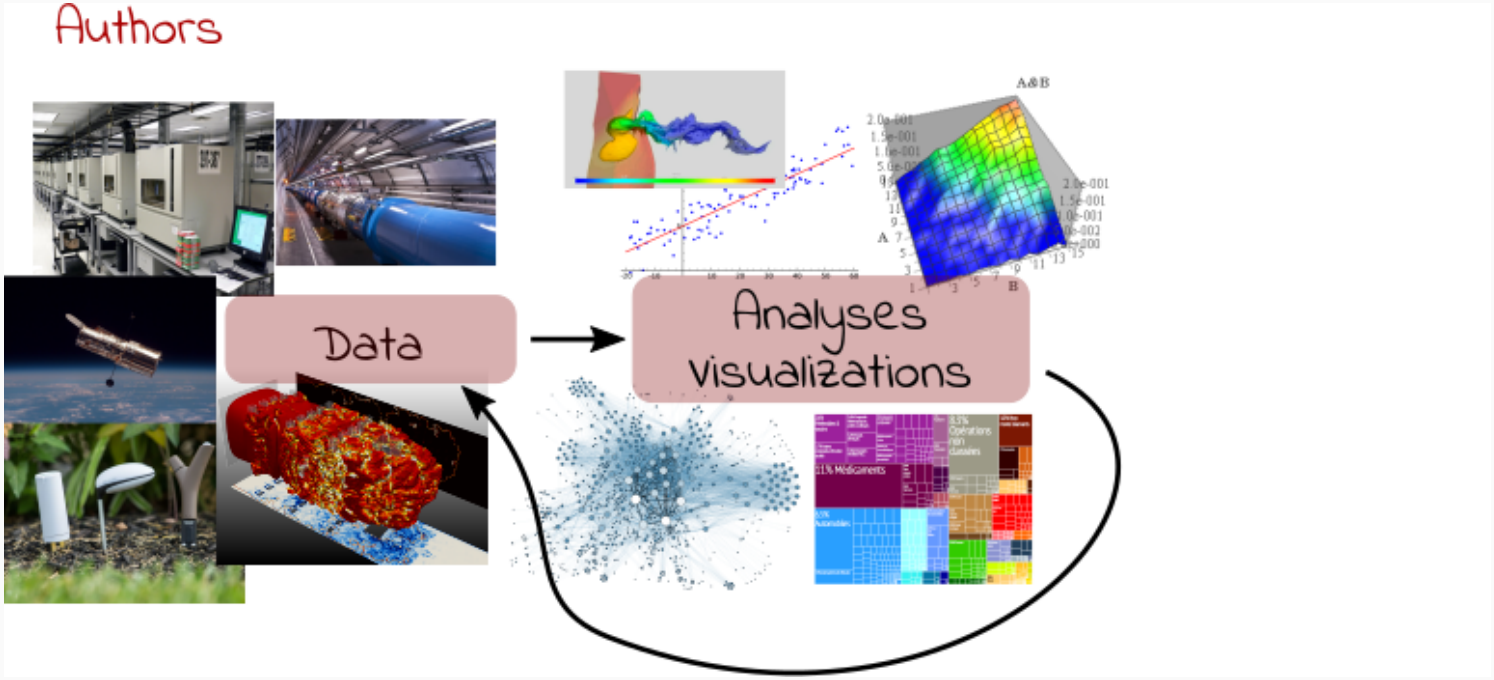
The processing steps between raw observations and findings have gotten increasingly numerous and complex



- A reproducible document links raw data with final figures
- What's at stake: verification and reuse by others

Why and how to use a computational notebook?

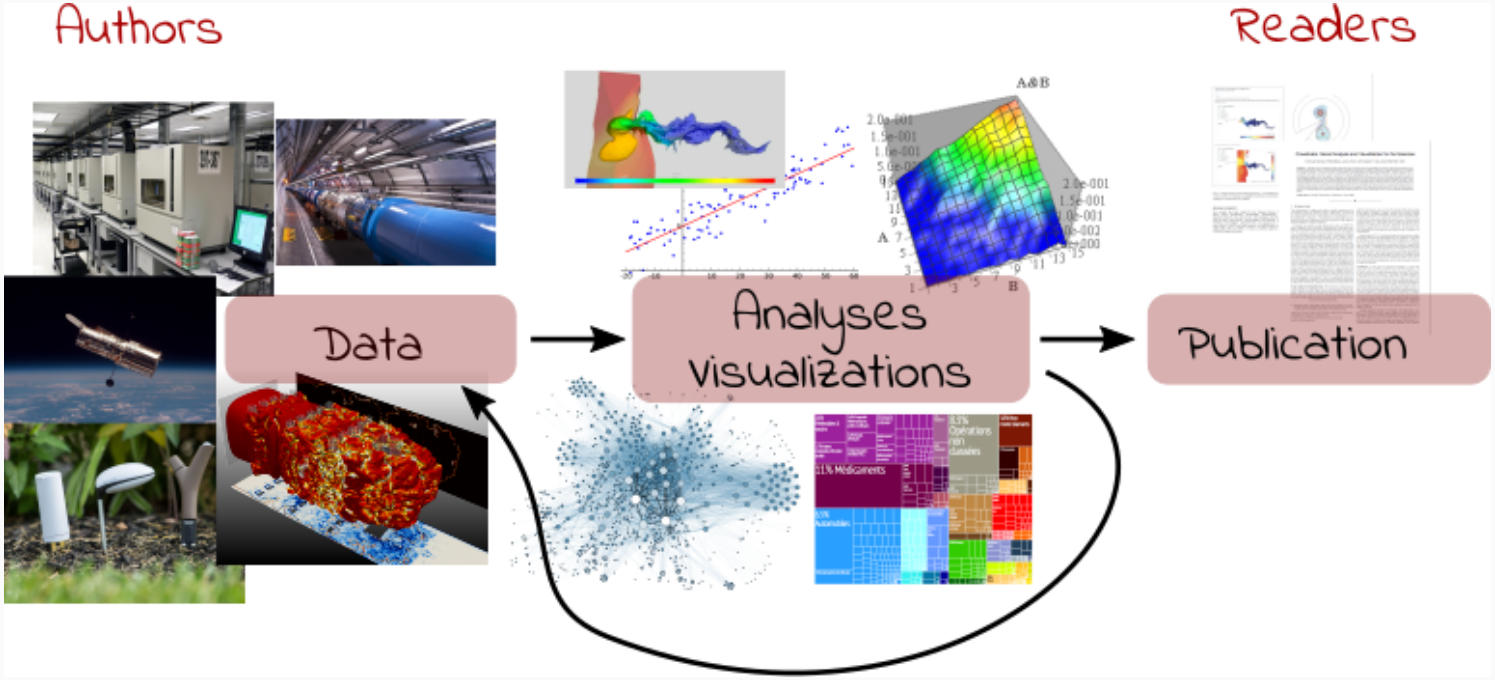
The processing steps between raw observations and findings have gotten increasingly numerous and complex



- A reproducible document links raw data with final figures
- What's at stake: verification and reuse by others

Why and how to use a computational notebook?

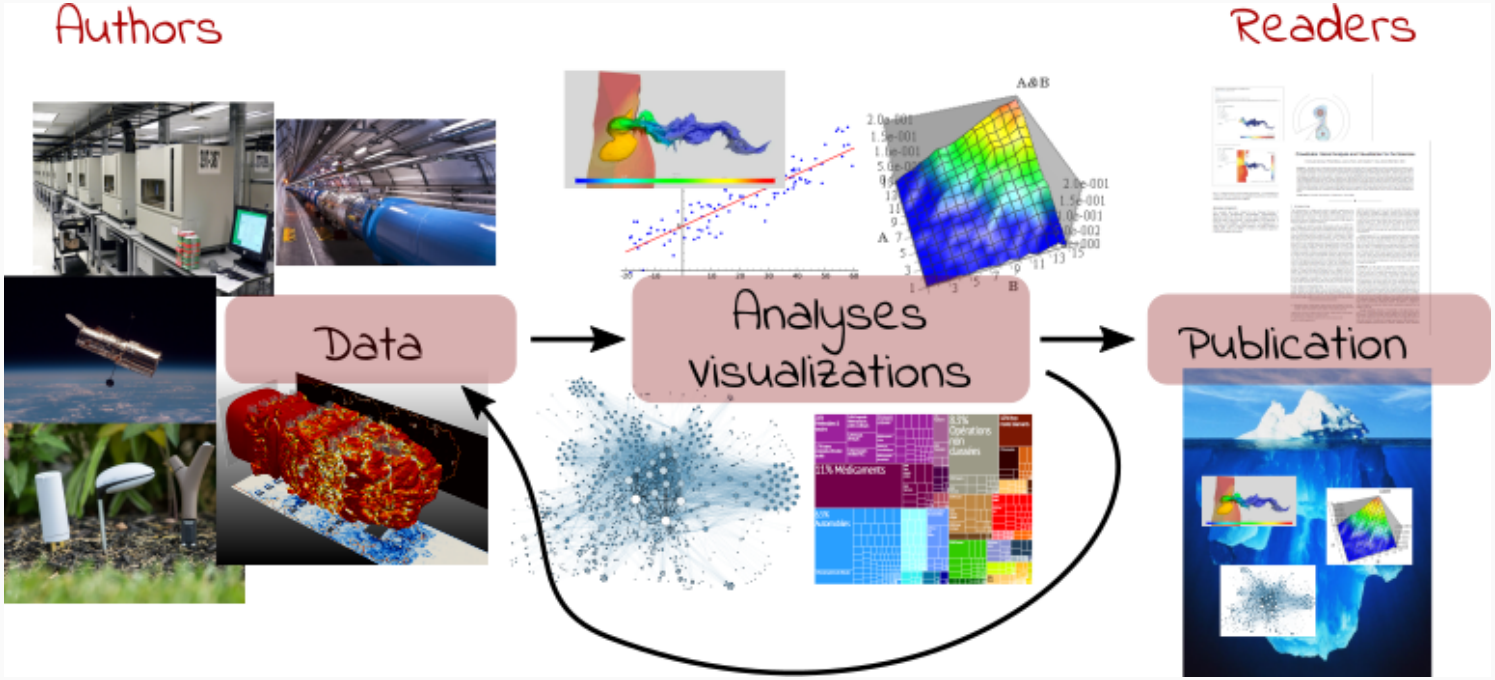
The processing steps between raw observations and findings have gotten increasingly numerous and complex



- A reproducible document links raw data with final figures
- What's at stake: verification and reuse by others

Why and how to use a computational notebook?

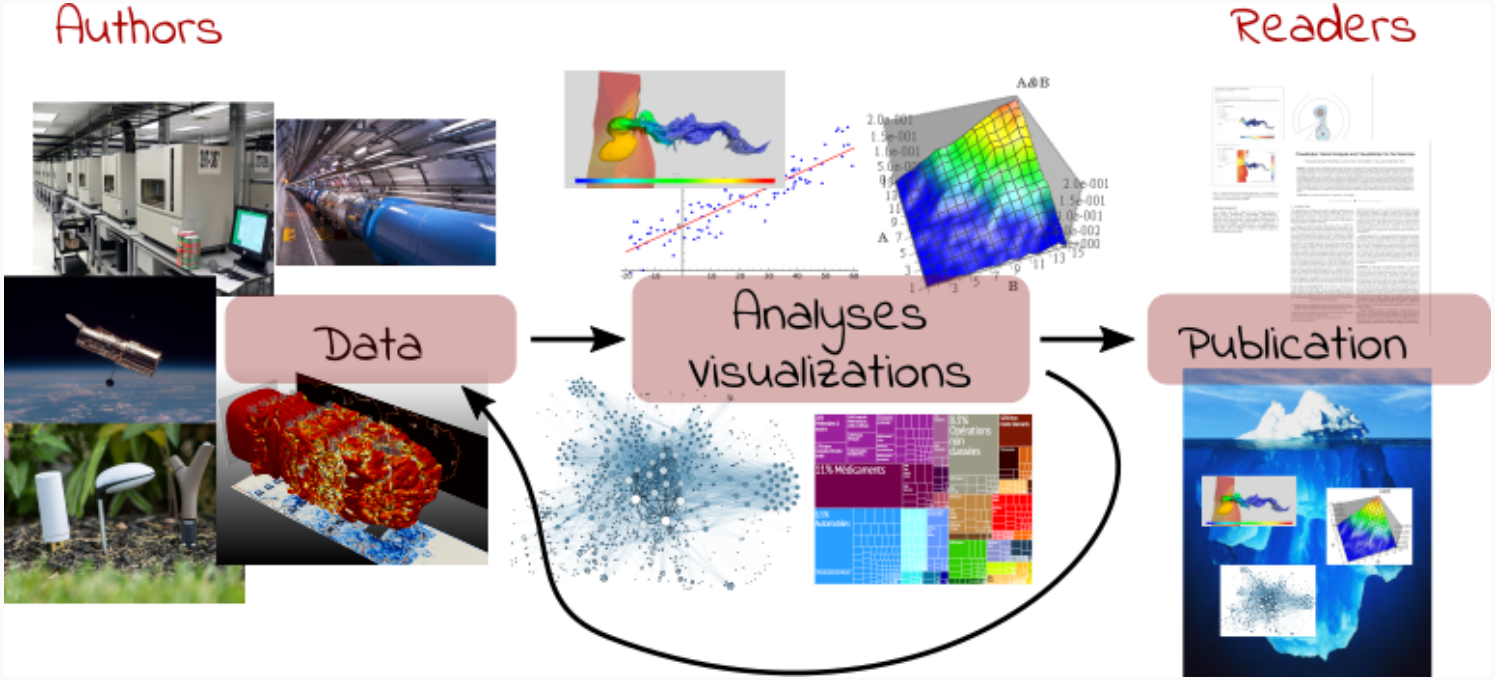
The processing steps between raw observations and findings have gotten increasingly numerous and complex



- A reproducible document links raw data with final figures
- What's at stake: verification and reuse by others

Why and how to use a computational notebook?

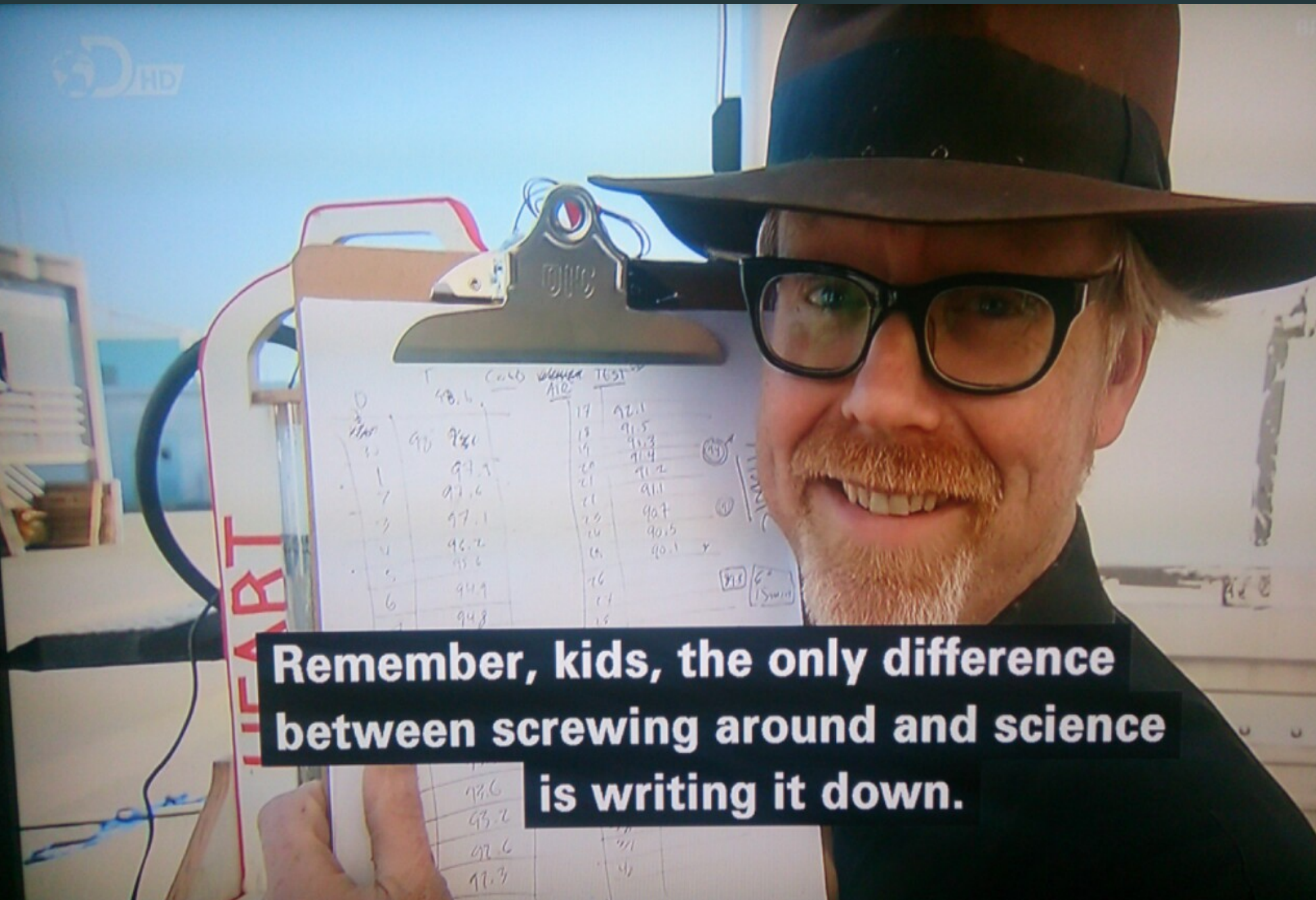
The processing steps between raw observations and findings have gotten increasingly numerous and complex



Reproducible Research = Bridging the Gap by working Transparently

- A reproducible document links raw data with final figures
- What's at stake: verification and reuse by others

Tool 2: Electronic Notebooks



Taking notes

Org-mode and Markdown two simple text formats

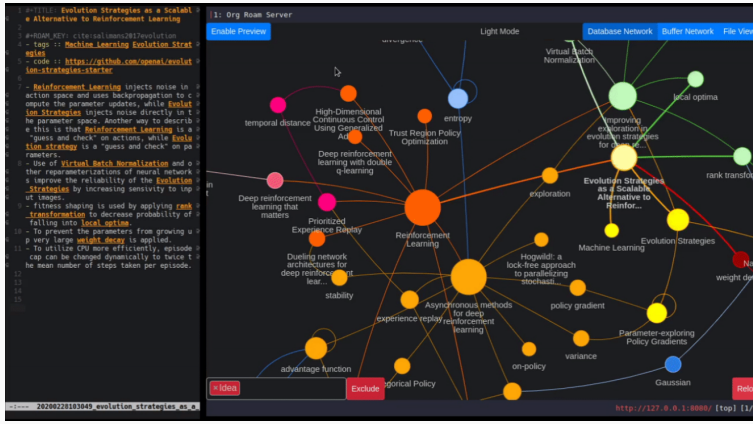
- simple formatting and export to more elaborate formats
- Hyperlinks, images, code, etc.

Journal structure

- My journal/notebook (single org file)
- Tom Cornebize's journal (single org file + Jupyter notebooks)

Zettelkasten possible structure

- Org-roam, Zettler, Roam, Obsidian, ...
- Architects, librarians, gardeners



Taking notes

Org-mode and Markdown two simple text formats

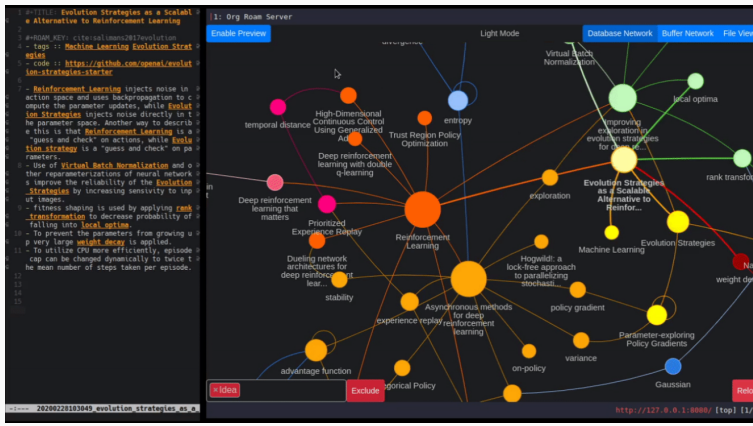
- simple formatting and export to more elaborate formats
- Hyperlinks, images, code, etc.

Journal structure

- My journal/notebook (single org file)
- Tom Cornebize's journal (single org file + Jupyter notebooks)

Zettelkasten possible structure

- Org-roam, Zettler, Roam, Obsidian, ...
- Architects, librarians, gardeners



My recommendations: Do not use a fancy cloud-based proprietary tool

- Simple open source text-based format
- Control versions and backup yourself (e.g., using gitlab, github)
- Single location if possible
- Annotate (tags in a journal, links in a Zettelkasten)

Conclusion

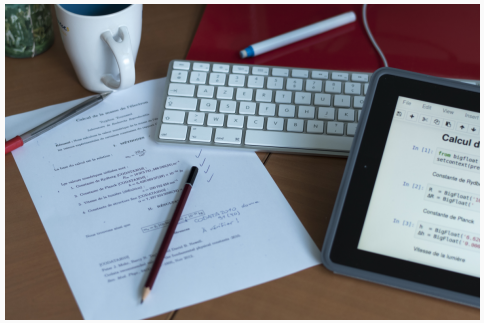
Reproducible Research = Rigor and Transparency

Good research requires time and resources

1. **Train yourself and your students**: RR, statistics, experiments
   - Beware of checklists and norms Understand what's at stake

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

- Konrad Hinsen, Christophe Pouzat
- `Markdown`, `GitLab`, `Jupyter` / `Rstudio` / `OrgMode`
- 3rd Edition: March 2020 – March 2024 (15,000+)

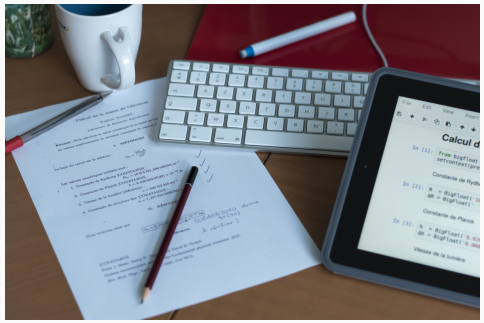


Reproducible Research = Rigor and Transparency

Good research requires time and resources

1. **Train yourself and your students**: RR, statistics, experiments
   - Beware of checklists and norms Understand what's at stake

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



- Konrad Hinsen, Christophe Pouzat
- `Markdown`, `GitLab`, `Jupyter` / `Rstudio` / `OrgMode`
- 3rd Edition: March 2020 – March 2024 (15,000+)

MOOC "Advanced RR" planned for Jan. 2024

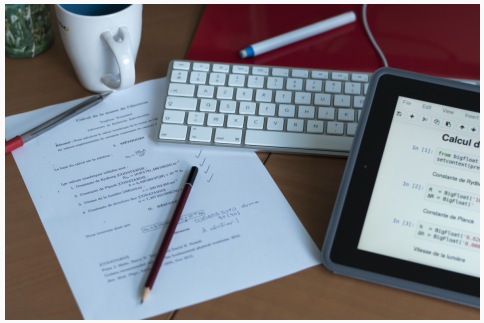
- Managing data (`FITS/HDF5`, `git annex`)
- Software environment control (`docker`, `singularity`, `guix`)
- Scientific workflow (`make`, `snakemake`)

Reproducible Research = Rigor and Transparency

Good research requires time and resources

1. **Train yourself and your students**: RR, statistics, experiments
   - Beware of checklists and norms Understand what's at stake

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



- Konrad Hinsen, Christophe Pouzat
- `Markdown`, `GitLab`, `Jupyter` / `Rstudio` / `OrgMode`
- 3rd Edition: March 2020 – March 2024 (15,000+)

MOOC "Advanced RR" planned for Jan. 2024

- Managing data (`FITS/HDF5`, `git annex`)
- Software environment control (`docker`, `singularity`, `guix`)
- Scientific workflow (`make`, `snakemake`)

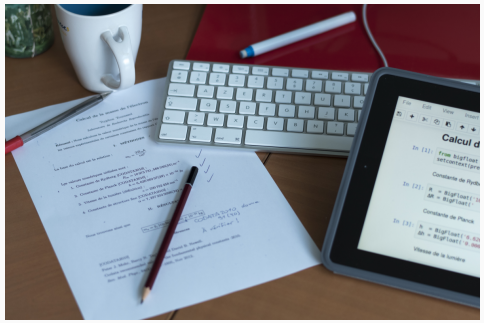
2. **Change the norm:** make publication practices evolve
3. **Incentive**: consider RR/open science when hiring/promoting

Reproducible Research = Rigor and Transparency

Good research requires time and resources

1. **Train yourself and your students**: RR, statistics, experiments
   - Beware of checklists and norms Understand what's at stake

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



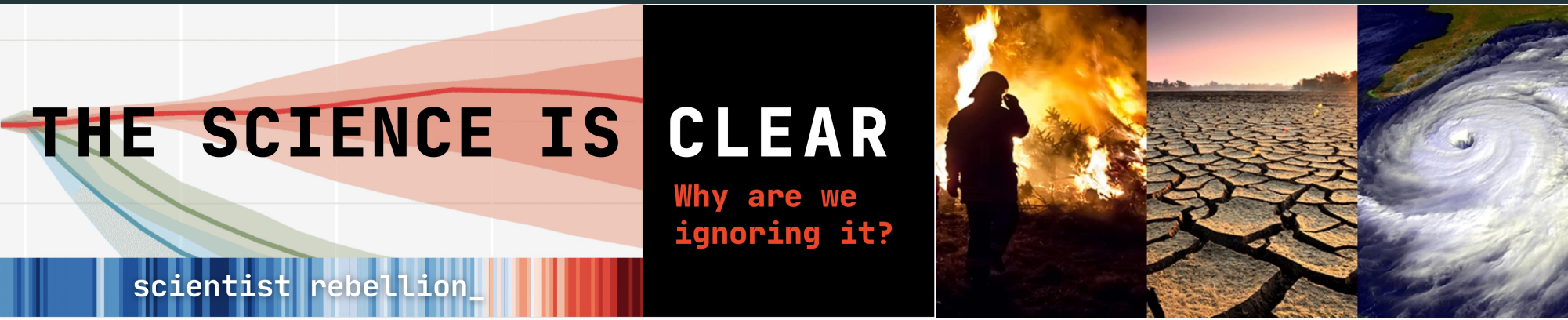
- Konrad Hinsen, Christophe Pouzat
- `Markdown`, `GitLab`, `Jupyter` / `Rstudio` / `OrgMode`
- 3rd Edition: March 2020 – March 2024 (15,000+)

MOOC "Advanced RR" planned for Jan. 2024

- Managing data (`FITS/HDF5`, `git annex`)
- Software environment control (`docker`, `singularity`, `guix`)
- Scientific workflow (`make`, `snakemake`)

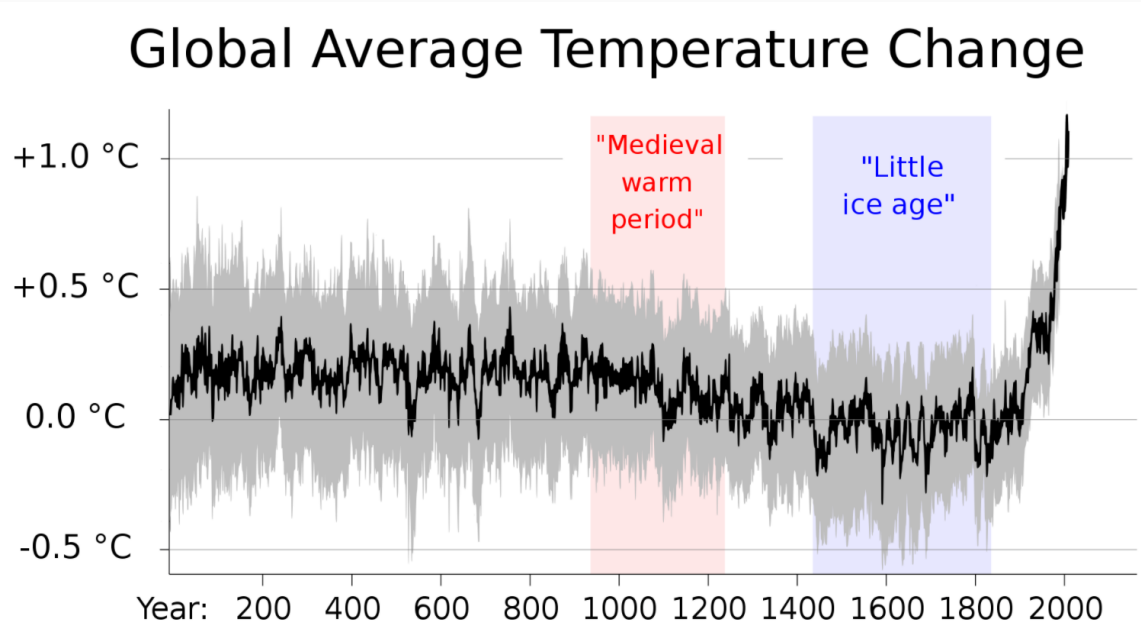
2. **Change the norm:** make publication practices evolve
3. **Incentive**: consider RR/open science when hiring/promoting

The Elephant in the Room: Climate Change

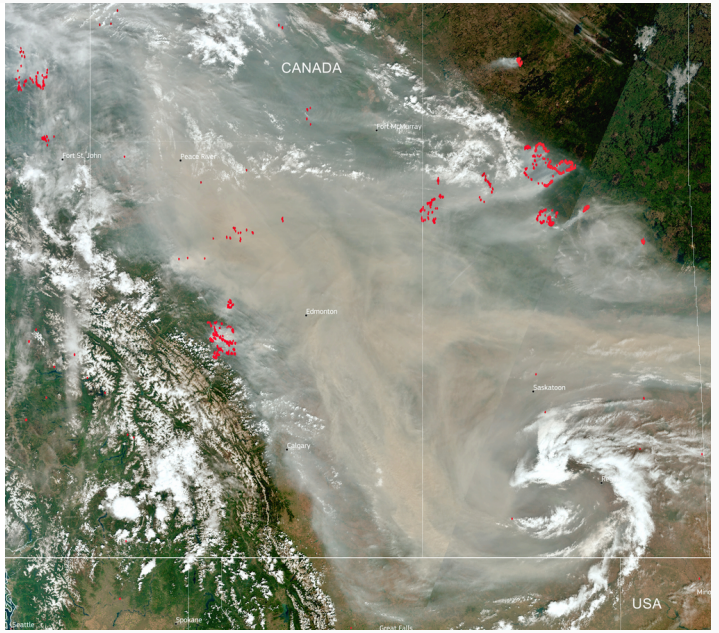


IPCC, IPBES, https://climate.nasa.gov/

1. Global climate change is **not** a future problem

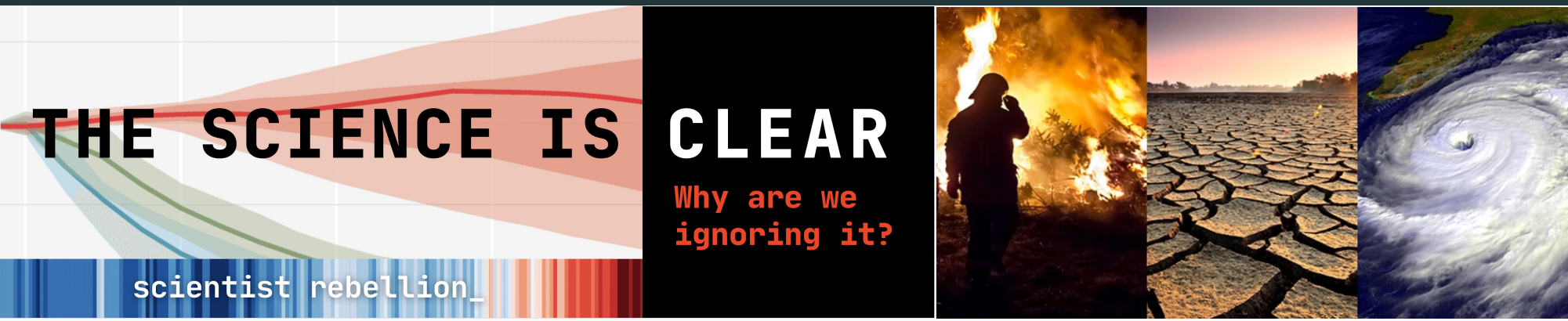


https://en.wikipedia.org/wiki/Global_temperature_record



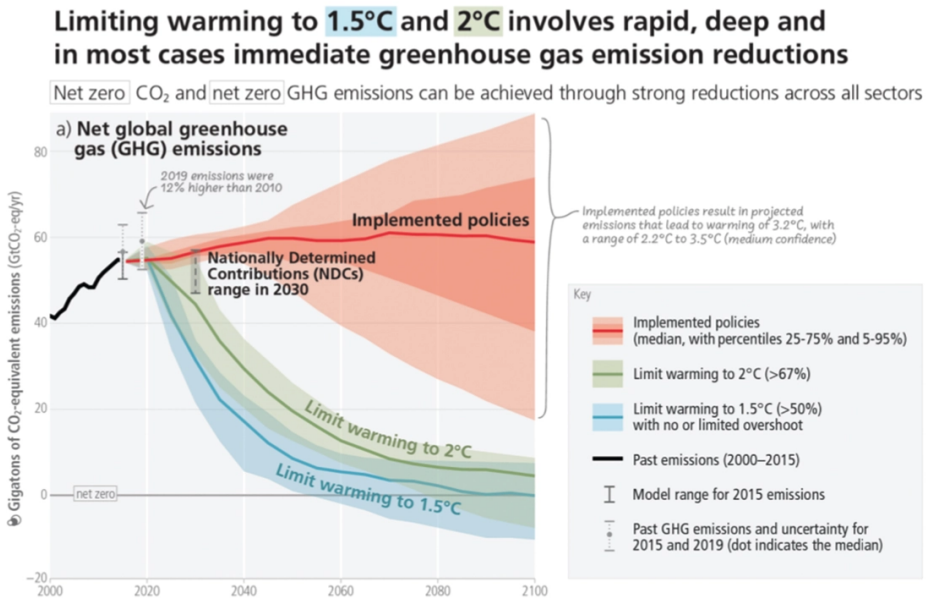
2023 Alberta wildfires (> 1 Mha)

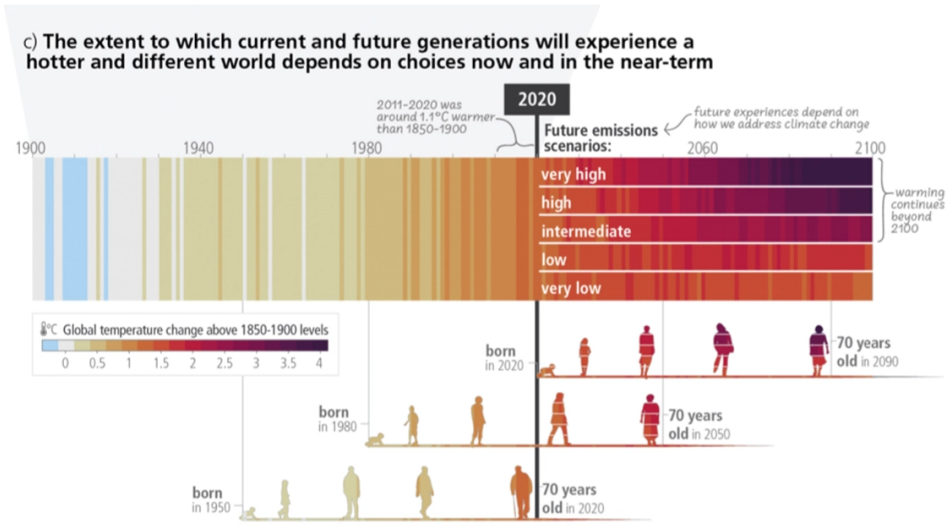
The Elephant in the Room: Climate Change



IPCC, IPBES, https://climate.nasa.gov/

1. Global climate change is **not** a future problem
2. It is **entirely** due to human activity

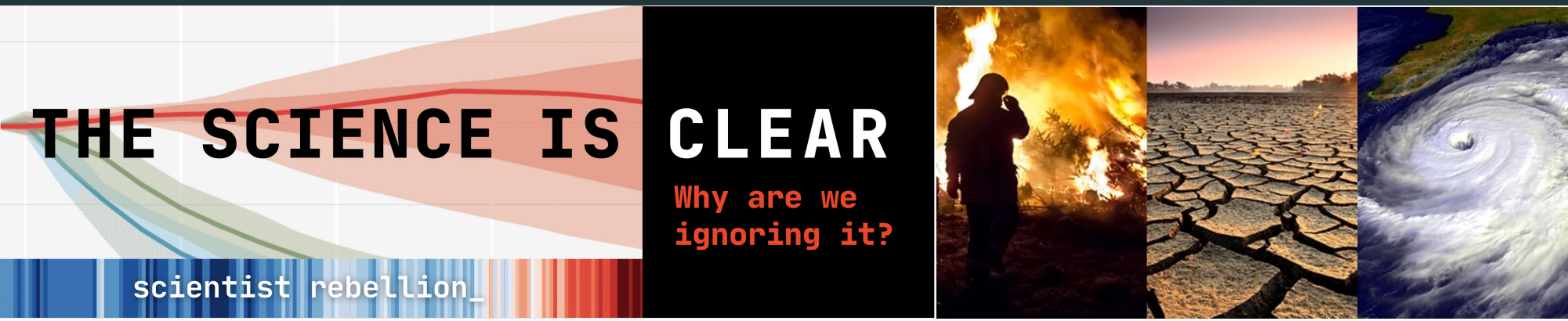




Paris Agreement'15 ∼ Net Zero by 2050

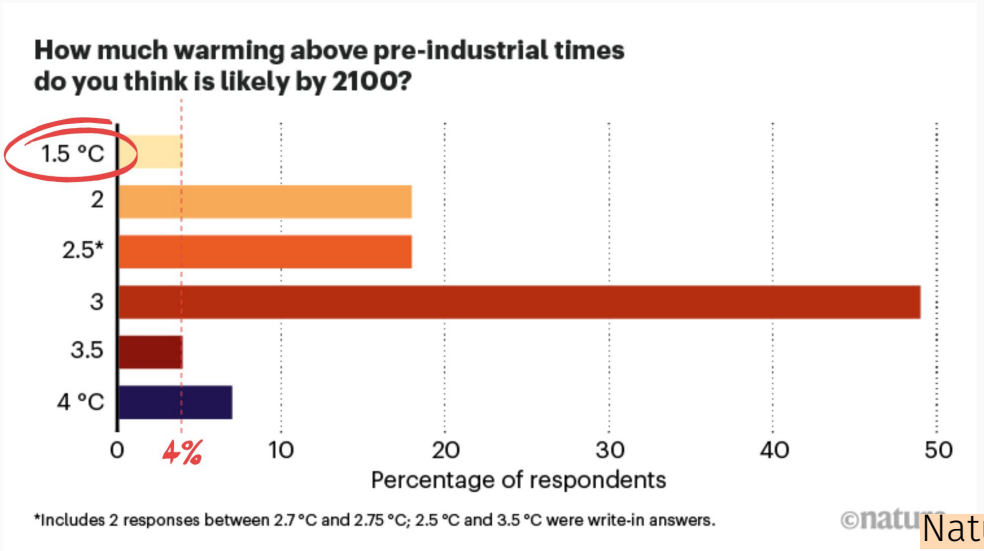
Latest IPCC report

The Elephant in the Room: Climate Change



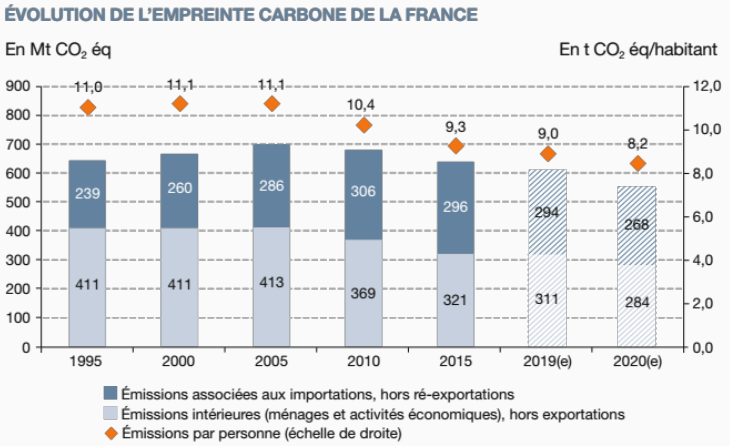
IPCC, IPBES, https://climate.nasa.gov/

1. Global climate change is **not** a future problem
2. It is **entirely** due to human activity
3. *9 out of 10 IPCC scientists believe overshoot is likely*

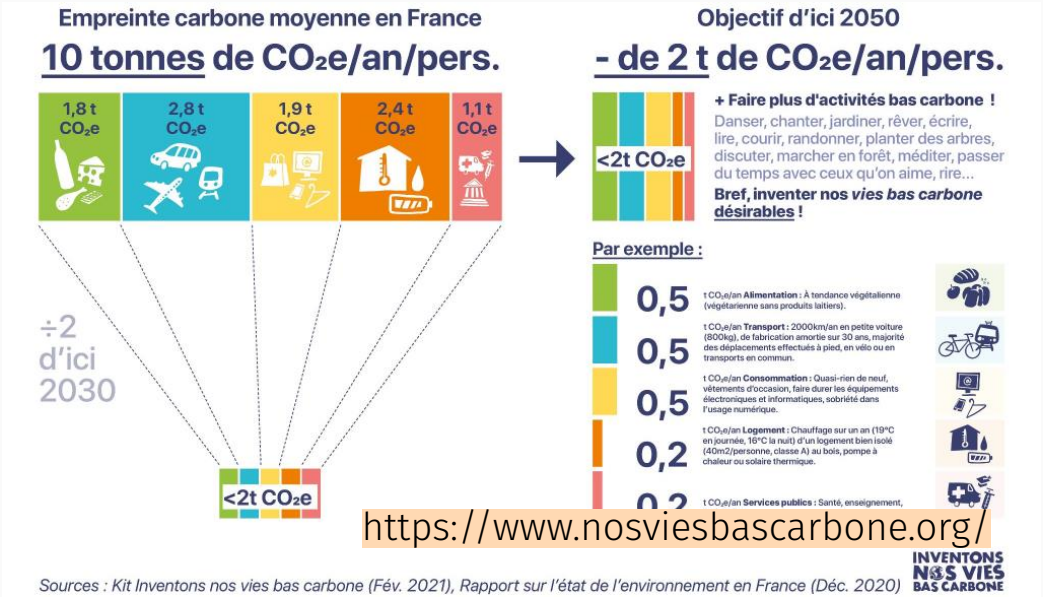


Nature survey, Nov. 2021

Put aside biodiversity loss, pollution, freshwater, land system change...

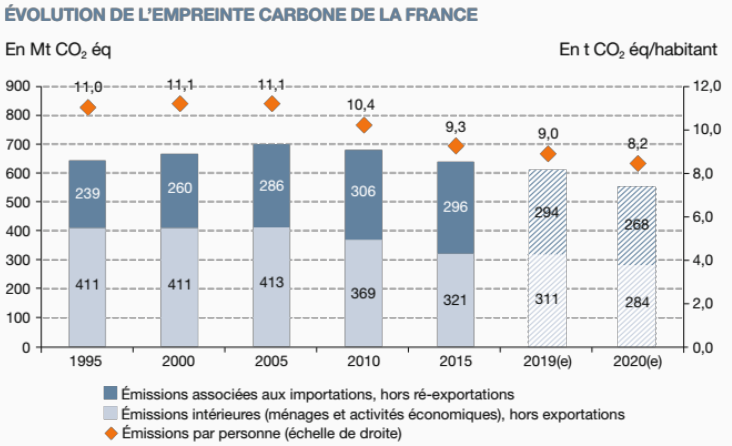


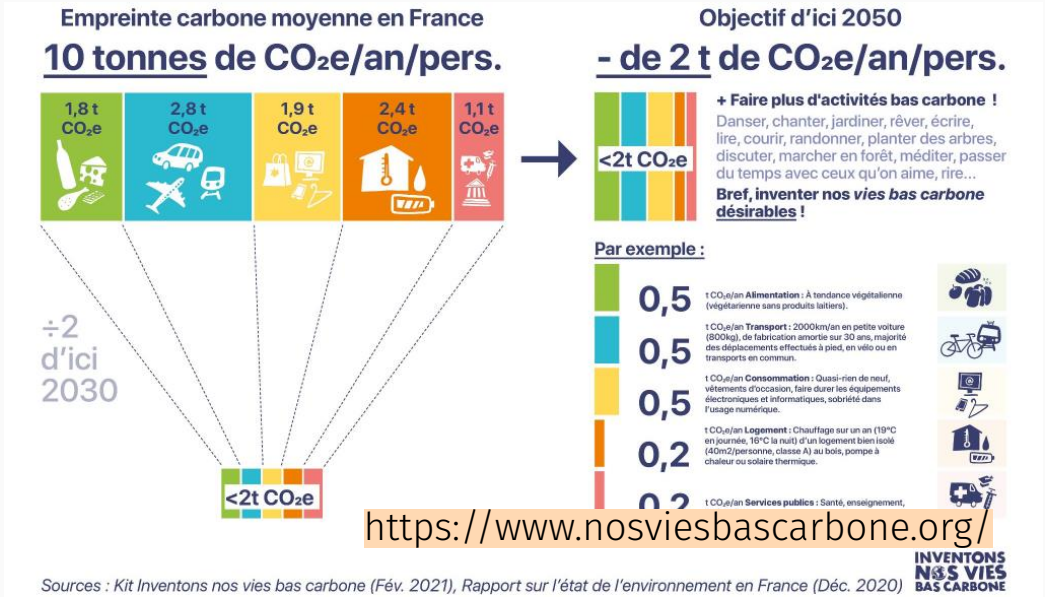
(e) = estimations.
Note : l'empreinte carbone porte sur les trois principaux gaz à effet de serre (CO_2, CH_4, N_2O). En 2021, la méthodologie a été ajustée afin de mieux tenir compte de l'évolution des coûts du pétrole brut, du gaz et du charbon ; l'ensemble de la série a ainsi été révisée, l'essentiel des ajustements portant sur les émissions importées de CH_4.
Champ : périmètre Kyoto (métropole et outre-mer appartenant à l'UE).
***Sources :** Citepa ; AIE ; FAO ; Douanes ; Eurostat ; Insee. Traitement : SDES, 2021*



Sources : Kit Inventons nos vies bas carbone (Fév. 2021), Rapport sur l'état de l'environnement en France (Déc. 2020)

Put aside biodiversity loss, pollution, freshwater, land system change...

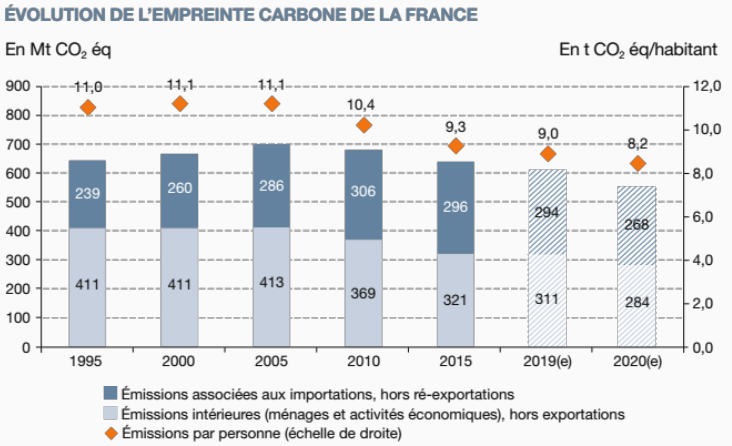


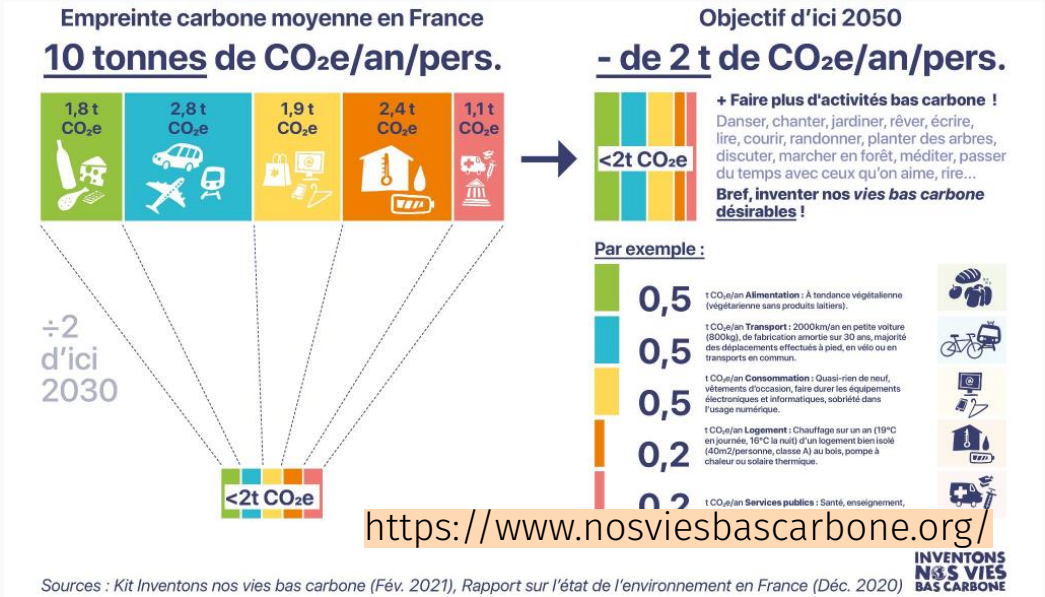


French government response

- *Verdissement de l'industrie: « pause » sur les normes environnementales*
- *Loi de programmation militaire (+41%)*
- *Nous devons préparer la France à une élévation de la température de 4 °C*
- Academia ? PEPR 5G, Cloud, NUMPEX, Quantique, IA, Agroécologie et numérique

Put aside biodiversity loss, pollution, freshwater, land system change...





French government response

- *Verdissement de l'industrie: « pause » sur les normes environnementales*
- *Loi de programmation militaire (+41%)*
- *Nous devons préparer la France à une élévation de la température de 4 °C*
- Academia ? PEPR 5G, Cloud, NUMPEX, Quantique, IA, Agroécologie et numérique

~~Several~~ scenarios on the table

- What will research/CS look like/be used for in such a world?
- Energy optimization/saving ≠ sobriety and frugality

Other nightmares and tools

Nightmare 1: Fighting Software Environments Nightmare

What is hiding behind a simple

```
1 import matplotlib
```

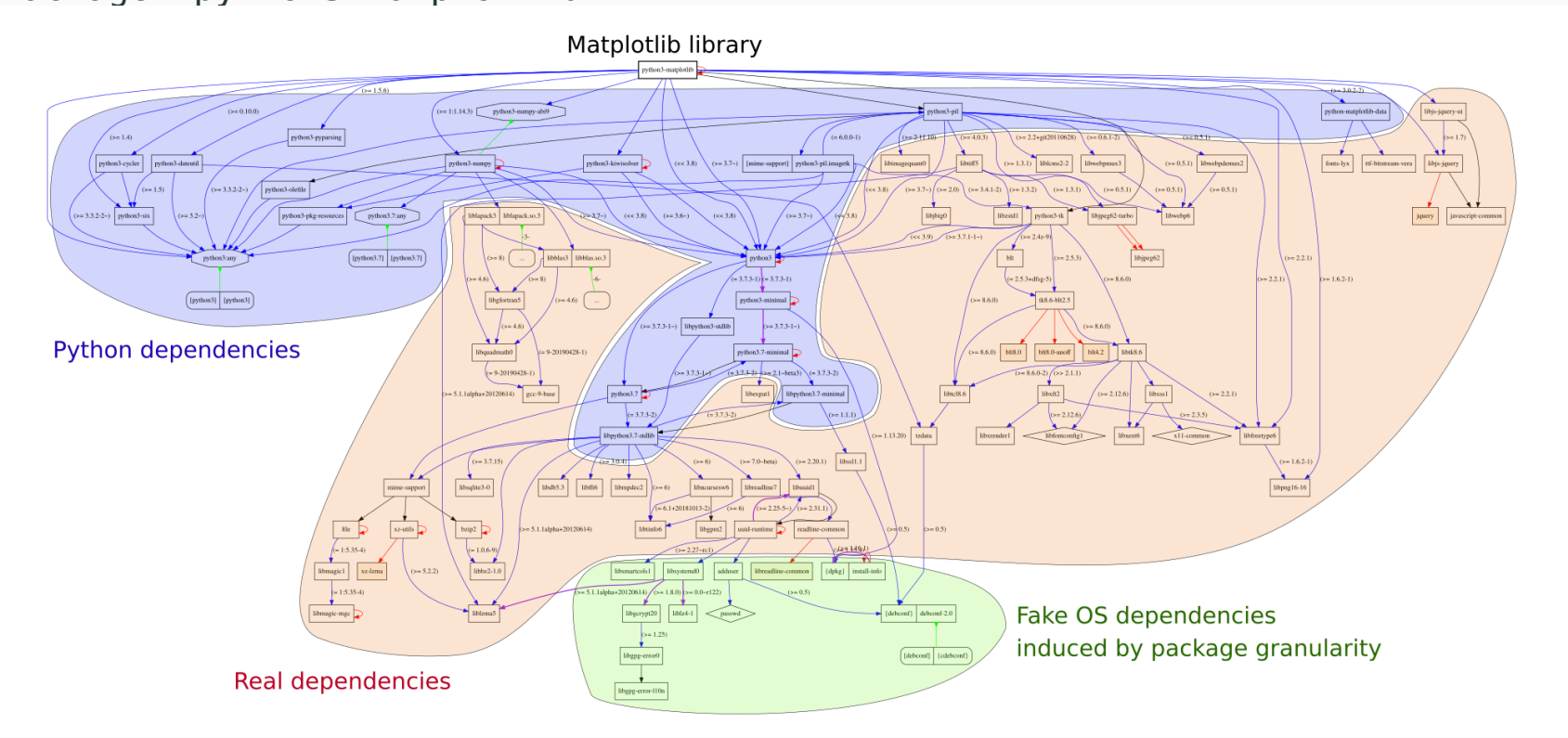
```
Package: python3-matplotlib
Version: 2.1.1-2
Depends: python3-dateutil, python-matplotlib-data (>= 2.1.1-2),
python3-pyparsing (>= 1.5.6), python3-six (>= 1.10), python3-tz,
libjs-jquery, libjs-jquery-ui, python3-numpy (>= 1:1.13.1),
python3-numpy-abi9, python3 (<< 3.7), python3 (>= 3.6~),
python3-cycler (>= 0.10.0), python3:any (>= 3.3.2-2~), libc6 (>=
2.14), libfreetype6 (>= 2.2.1), libgcc1 (>= 1:3.0), libpng16-16 (>=
1.6.2-1), libstdc++6 (>= 5.2), zlib1g (>= 1:1.1.4)
```

Nightmare 1: Fighting Software Environments Nightmare

What is hiding behind a simple

```
1 import matplotlib
```

`Package: python3-matplotlib`



Nightmare 1: Fighting Software Environments Nightmare

Python and its rapidly evolving environment

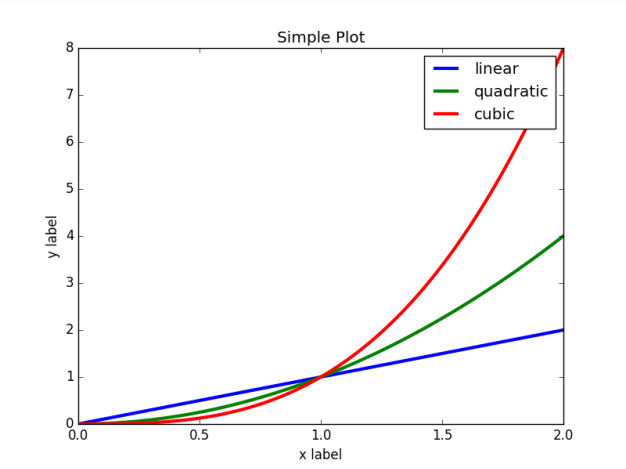
```
python2 -c "print(10/3)"
python3 -c "print(10/3)"
```

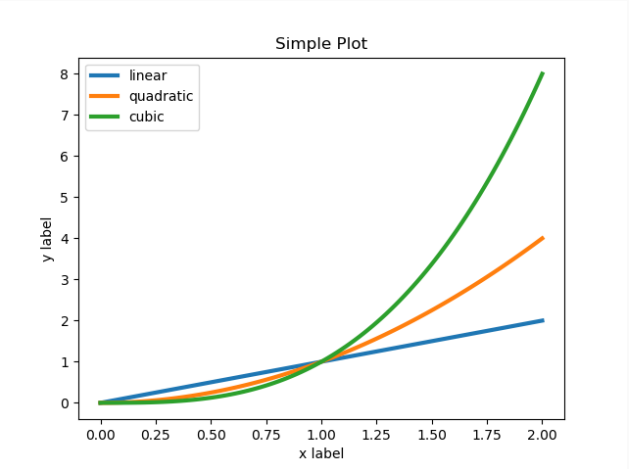
```
3
3.3333333333333335
```

Nightmare 1: Fighting Software Environments Nightmare

Python and its rapidly evolving environment

```
python2 -c "print(10/3)"
python3 -c "print(10/3)"
```

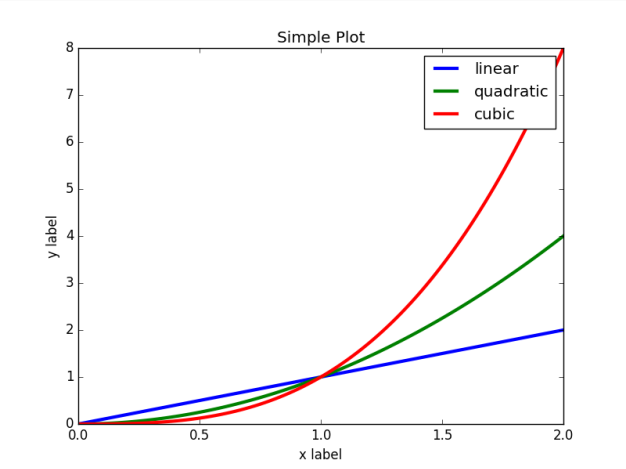


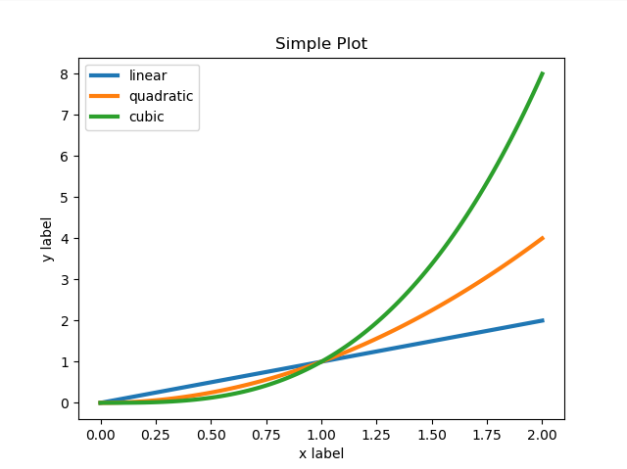


Nightmare 1: Fighting Software Environments Nightmare

Python and its rapidly evolving environment

```
python2 -c "print(10/3)"
python3 -c "print(10/3)"
```



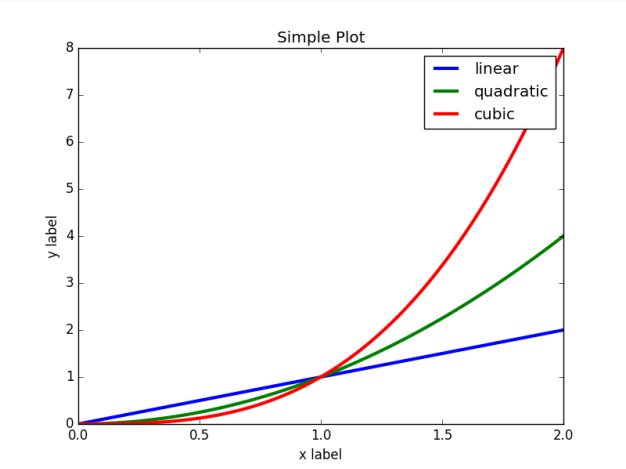


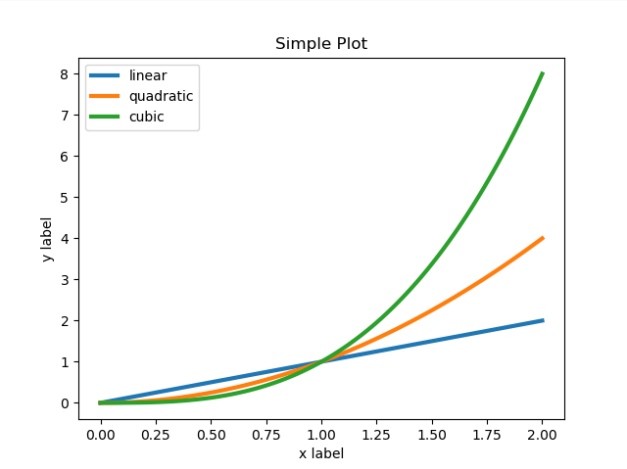
Cortical Thickness Measurements (PLOS ONE, June 2012): *FreeSurfer: differences were found between the Mac and HP workstations and between Mac OSX 10.5 and OSX 10.6.*

Nightmare 1: Fighting Software Environments Nightmare

Python and its rapidly evolving environment

```
1 python2 -c "print(10/3)"
2 python3 -c "print(10/3)"
```





Cortical Thickness Measurements (PLOS ONE, June 2012): *FreeSurfer: differences were found between the Mac and HP workstations and between Mac OSX 10.5 and OSX 10.6.*

ReproZip

Guix

Nightmare 2: Fighting Information Loss with Archives

D. Spinellis. *The Decay and Failures of URL References*. CACM, 46(1), Jan 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*

Nightmare 2: Fighting Information Loss with Archives

D. Spinellis. *The Decay and Failures of URL References*. CACM, 46(1), Jan 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*

Article archives arXiv.org HAL archives-ouvertes.fr

Data archives figshare zenodo

Software Archive Software Heritage


or  = awesome collaborations ≠ archive

That's all Folks!