INTRODUCTION

COURSE: GOOD PRACTICES FOR REPRODUCIBLE BIOINFORMATICS DATA ANALYSIS

Frédéric Lemoine
Institut Pasteur

2021/05/23

PART 1

Module organization

Progress

- Monday 23th May
  - AM: Introduction to reproducibility. Managing code with git
  - PM: Practical about git
- Tuesday 24th May
  - AM: 1) Controling software environment with Conda
  - AM: 2) Controling software environment with containers (Singularity/Docker)
  - PM: Practical about Conda and containers
- Monday 30th May
  - AM: 1) Running data analyses with Notebooks
  - AM: 2) Running data analyses with workflows
  - PM: Practical about Workflows
- Tuesday 31th May
  - AM: Good practices for software, tools and script development
  - PM: Finalizing group projects + presentations of results

Team

- Amandine Perrin, IR Pasteur (Hub Bioinfo, GEM unit)
- Etienne Kornobis, IR Pasteur (Hub Bioinfo, Biomics platform)
- Bertrand Néron, IR Pasteur (Hub Bioinfo, ALPS group)
- Frédéric Lemoine, IR Pasteur (Hub Bioinfo, GEVA unit)

Project: Analyzing SARS-CoV-2 data

- Goal: Writing a reproducible workflow to analyse SARS-CoV-2 sequencing data :
  1. Mapping
  2. SNP Calling
  3. Consensus
  4. Clade detection
- Using:
  1. git
  2. containers
  3. 1 workflow system
- By groups of 4 people
- A final presentation the last day

“Hardware”: Virtual machines from IT Dept

We will see that this afternoon but:

1. Connect to desktop.pasteur.fr
2. Go to the HTML access
3. Login with your Pasteur IDs
4. You are running a Linux Ubuntu VM with :
   - Singularity
   - Docker
   - Java
   - Conda
   - etc.

PART 2

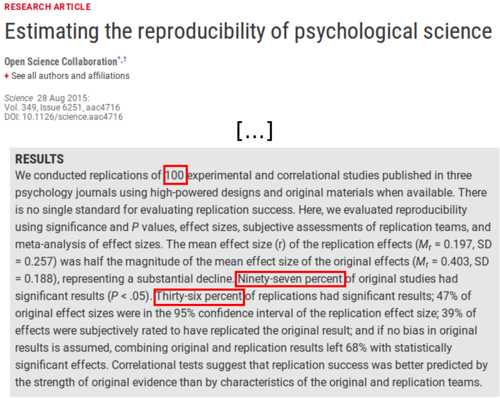
Introduction about reproducibility

2.1 Reproducibility crisis

History

In the past decade: Lots of debates around reproducibility.

Some fields that were particularly affected: Social sciences, Psychology, Clinical research;



RESEARCH ARTICLE

Estimating the reproducibility of psychological science

Open Science Collaboration[*,†]

+ See all authors and affiliations

Science 28 Aug 2015:
Vol. 349, Issue 6251, aac4716
DOI: 10.1126/science.aac4716

[…]

RESULTS

We conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original materials when available. There is no single standard for evaluating replication success. Here, we evaluated reproducibility using significance and *P* values, effect sizes, subjective assessments of replication teams, and meta-analysis of effect sizes. The mean effect size (r) of the replication effects (M_r = 0.197, SD = 0.257) was half the magnitude of the mean effect size of the original effects (M_r = 0.403, SD = 0.188), representing a substantial decline. Ninety-seven percent of original studies had significant results ($P < .05$). Thirty-six percent of replications had significant results; 47% of original effect sizes were in the 95% confidence interval of the replication effect size; 39% of effects were subjectively rated to have replicated the original result; and if no bias in original results is assumed, combining original and replication results left 68% with statistically significant effects. Correlational tests suggest that replication success was better predicted by the strength of original evidence than by characteristics of the original and replication teams.

2.1 Reproducibility crisis

Psychological research: Priming

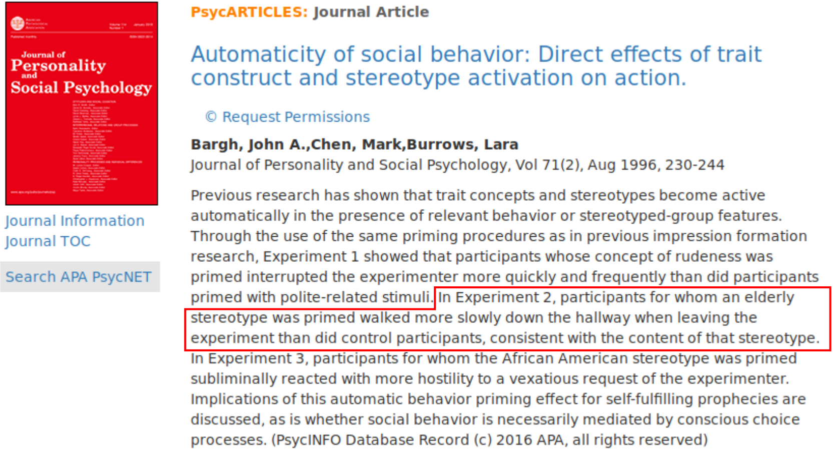
Power of Suggestion (`https://www.chronicle.com/article/Power-of-Suggestion/136907/`):

> *The studies that raise eyebrows are mostly in an area known as behavioral or goal priming, research that demonstrates how subliminal prompts can make you do all manner of crazy things. A warm mug makes you friendlier. The American flag makes you vote Republican. Fast-food logos make you impatient.*

2.1 Reproducibility crisis

Priming: the original study

Cited > 5000 times (google scholar, 06/11/2019)!



PsycARTICLES: Journal Article

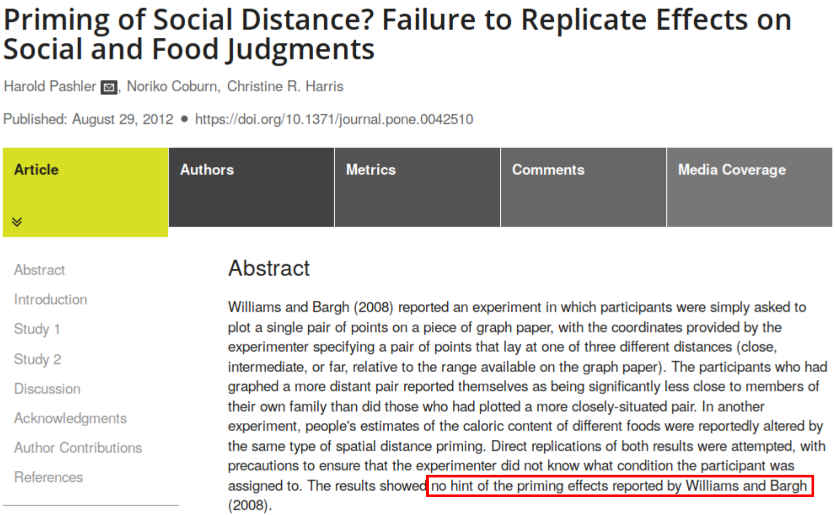
Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action.

Bargh, John A.,Chen, Mark,Burrows, Lara

Journal of Personality and Social Psychology, Vol 71(2), Aug 1996, 230-244

Previous research has shown that trait concepts and stereotypes become active automatically in the presence of relevant behavior or stereotyped-group features. Through the use of the same priming procedures as in previous impression formation research, Experiment 1 showed that participants whose concept of rudeness was primed interrupted the experimenter more quickly and frequently than did participants primed with polite-related stimuli. In Experiment 2, participants for whom an elderly stereotype was primed walked more slowly down the hallway when leaving the experiment than did control participants, consistent with the content of that stereotype. In Experiment 3, participants for whom the African American stereotype was primed subliminally reacted with more hostility to a vexatious request of the experimenter. Implications of this automatic behavior priming effect for self-fulfilling prophecies are discussed, as is whether social behavior is necessarily mediated by conscious choice processes. (PsycINFO Database Record (c) 2016 APA, all rights reserved)

Priming: Replication studies



Priming of Social Distance? Failure to Replicate Effects on Social and Food Judgments

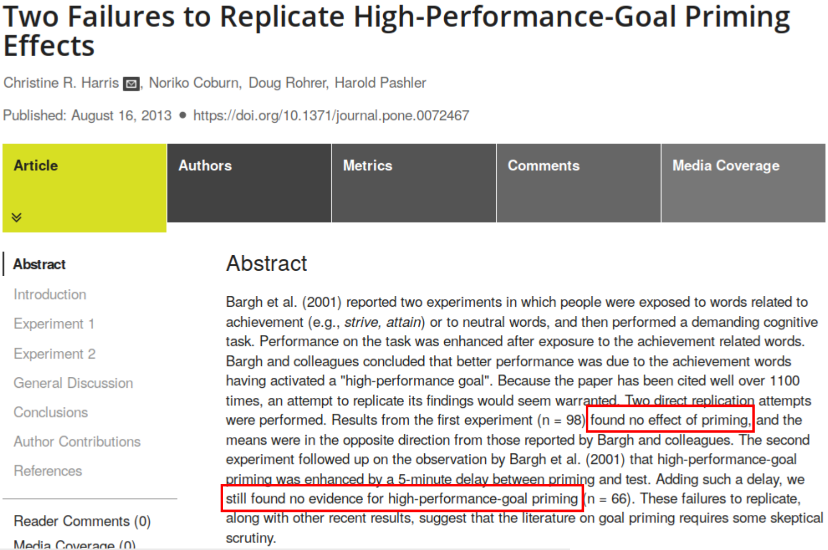
Harold Pashler ✉, Noriko Coburn, Christine R. Harris

Published: August 29, 2012 • https://doi.org/10.1371/journal.pone.0042510

Abstract

Williams and Bargh (2008) reported an experiment in which participants were simply asked to plot a single pair of points on a piece of graph paper, with the coordinates provided by the experimenter specifying a pair of points that lay at one of three different distances (close, intermediate, or far, relative to the range available on the graph paper). The participants who had graphed a more distant pair reported themselves as being significantly less close to members of their own family than did those who had plotted a more closely-situated pair. In another experiment, people's estimates of the caloric content of different foods were reportedly altered by the same type of spatial distance priming. Direct replications of both results were attempted, with precautions to ensure that the experimenter did not know what condition the participant was assigned to. The results showed no hint of the priming effects reported by Williams and Bargh (2008).

Priming: Replication studies



Two Failures to Replicate High-Performance-Goal Priming Effects

Christine R. Harris, Noriko Coburn, Doug Rohrer, Harold Pashler

Published: August 16, 2013 • https://doi.org/10.1371/journal.pone.0072467

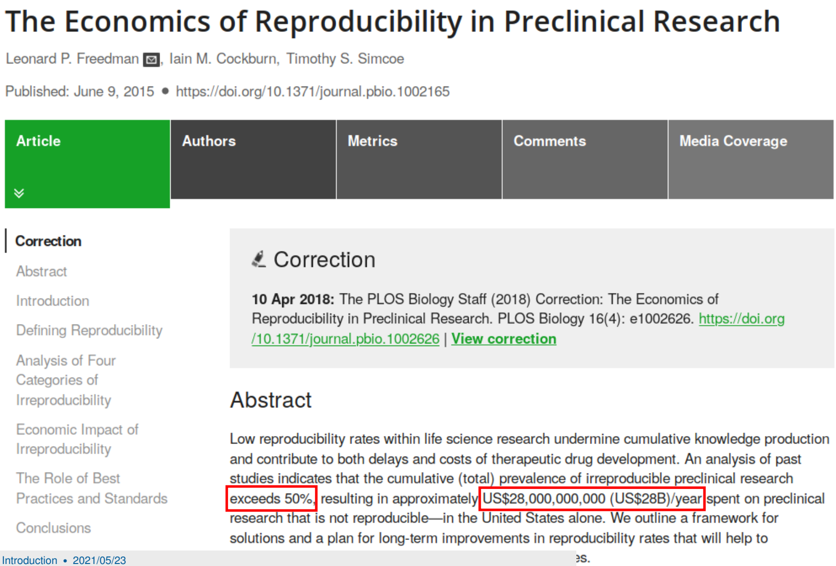
Abstract

Bargh et al. (2001) reported two experiments in which people were exposed to words related to achievement (e.g., *strive, attain*) or to neutral words, and then performed a demanding cognitive task. Performance on the task was enhanced after exposure to the achievement related words. Bargh and colleagues concluded that better performance was due to the achievement words having activated a "high-performance goal". Because the paper has been cited well over 1100 times, an attempt to replicate its findings would seem warranted. Two direct replication attempts were performed. Results from the first experiment (n = 98) found no effect of priming, and the means were in the opposite direction from those reported by Bargh and colleagues. The second experiment followed up on the observation by Bargh et al. (2001) that high-performance-goal priming was enhanced by a 5-minute delay between priming and test. Adding such a delay, we still found no evidence for high-performance-goal priming (n = 66). These failures to replicate, along with other recent results, suggest that the literature on goal priming requires some skeptical scrutiny.

INTRODUCTION ABOUT REPRODUCIBILITY

2.1 Reproducibility crisis

Preclinical research (Freeedman et al. 2015)



The Economics of Reproducibility in Preclinical Research

Leonard P. Freedman, Iain M. Cockburn, Timothy S. Simcoe

Published: June 9, 2015 • https://doi.org/10.1371/journal.pbio.1002165

- Correction
- Abstract
- Introduction
- Defining Reproducibility
- Analysis of Four Categories of Irreproducibility
- Economic Impact of Irreproducibility
- The Role of Best Practices and Standards
- Conclusions

Correction

10 Apr 2018: The PLOS Biology Staff (2018) Correction: The Economics of Reproducibility in Preclinical Research. PLOS Biology 16(4): e1002626. https://doi.org/10.1371/journal.pbio.1002626 | **View correction**

Abstract

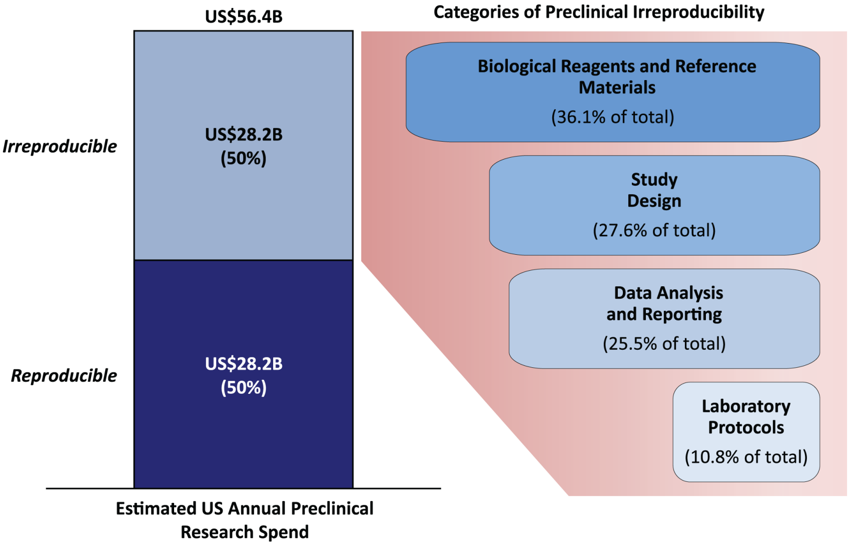
Low reproducibility rates within life science research undermine cumulative knowledge production and contribute to both delays and costs of therapeutic drug development. An analysis of past studies indicates that the cumulative (total) prevalence of irreproducible preclinical research exceeds 50%, resulting in approximately US$28,000,000,000 (US$28B)/year spent on preclinical research that is not reproducible—in the United States alone. We outline a framework for solutions and a plan for long-term improvements in reproducibility rates that will help to

es.

Institut Pasteur

2.1 Reproducibility crisis

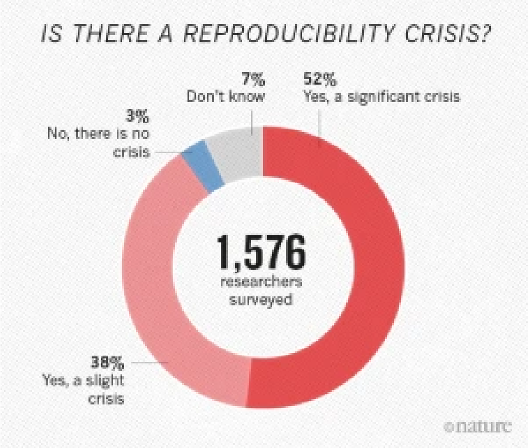
Preclinical research (Freeedman et al. 2015)



2.1 Reproducibility crisis

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

Survey of 1,576 researchers who took a brief online questionnaire on reproducibility in research:



2.1 Reproducibility crisis

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

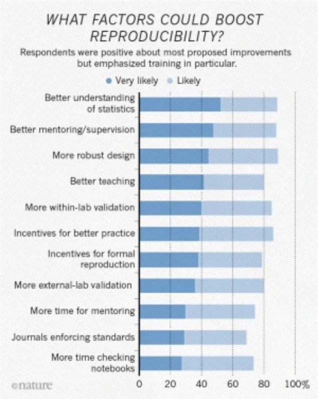
Survey of 1,576 researchers who took a brief online questionnaire on reproducibility in research:

- More than 70% of researchers have tried and failed to reproduce another scientist’s experiments,
- More than half have failed to reproduce their own experiments
- - pressure to publish and selective reporting - always or often contributed [to reproducibilty issues]. $>$ 50% pointed insufficient replication in the lab, poor oversight or low statistical power.

2.1 Reproducibility crisis

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

Survey of 1,576 researchers who took a brief online questionnaire on reproducibility in research:

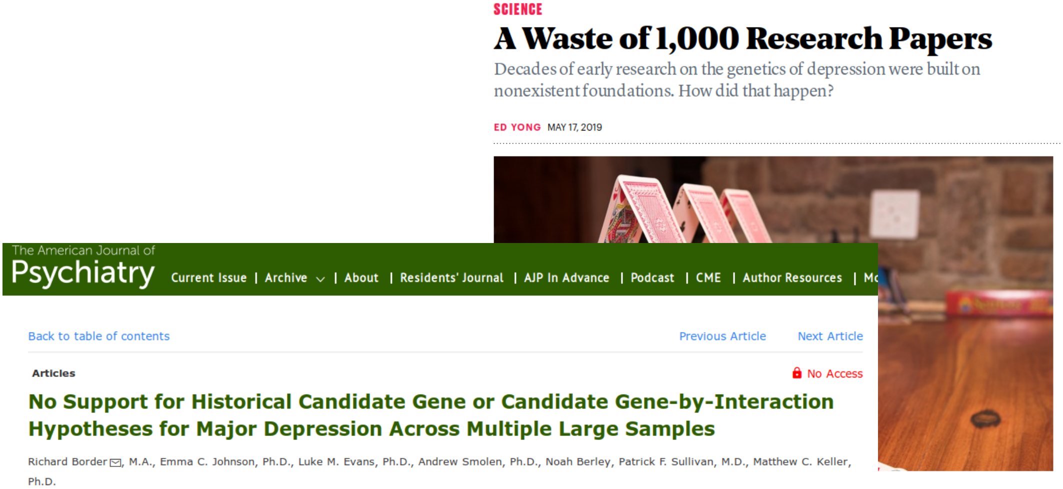


INTRODUCTION ABOUT REPRODUCIBILITY

2.1 Reproducibility crisis

Recent example

⇒ *Lack of statistical power*



SCIENCE

A Waste of 1,000 Research Papers

Decades of early research on the genetics of depression were built on nonexistent foundations. How did that happen?

ED YONG MAY 17, 2019

The American Journal of
Psychiatry
Current Issue | Archive ⌵ | About | Residents' Journal | AJP In Advance | Podcast | CME | Author Resources | Mo
Back to table of contents
Previous Article
Next Article
Articles
No Access

No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples

Richard Border, M.A., Emma C. Johnson, Ph.D., Luke M. Evans, Ph.D., Andrew Smolen, Ph.D., Noah Berley, Patrick F. Sullivan, M.D., Matthew C. Keller, Ph.D.

2.1 Reproducibility crisis

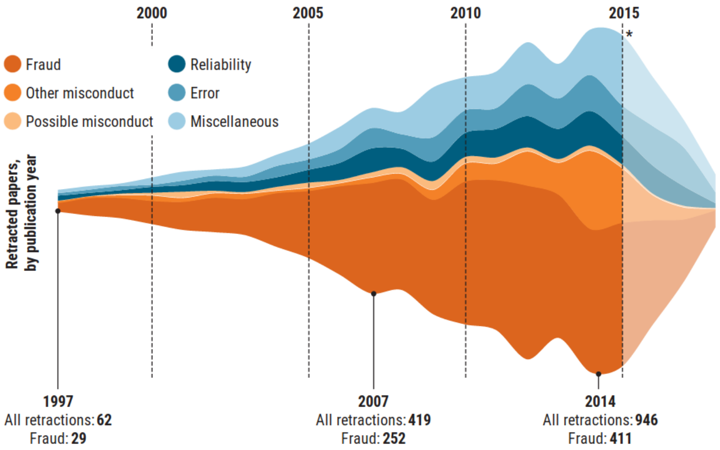
Recent example

“A waste of 1,000 research papers”, the Atlantic, 2019:

> *”What bothers me isn’t just that people said [the gene] mattered and it didn’t”, wrote the pseudonymous blogger Scott Alexander in a widely shared post. ”It’s that we built whole imaginary edifices on top of this idea of [it] mattering”. Researchers studied how SLC6A4 affects emotion centers in the brain, how its influence varies in different countries and demographics, and how it interacts with other genes. It’s as if they’d been ”describing the life cycle of unicorns, what unicorns eat, all the different subspecies of unicorn, which cuts of unicorn meat are tastiest, and a blow-by-blow account of a wrestling match between unicorns and Bigfoot”, Alexander wrote.*

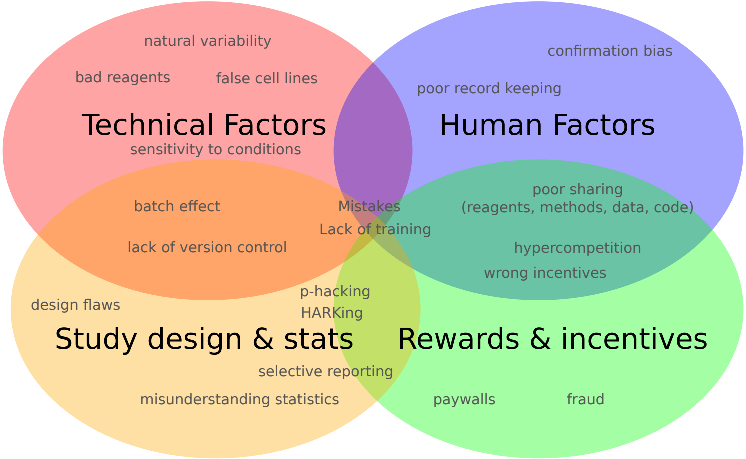
2.1 Reproducibility crisis

What about retracted papers?



2.2 Initiatives

Factors decreasing reproducibility



from https://www.repro4everyone.org

2.2 Initiatives

Quote

"“This focus on positive results is arguably one of the central drivers of the reproducibility crisis”, Russel A. Poldrack, “The Costs of Reproducibility”, Neuron, 2019”

2.2 Initiatives

A few pointers

- The Costs of Reproducibility, Neuron, 2019
- What does research reproducibility mean? Science Trans. Med., 2016
- The Economics of Reproducibility in Preclinical Research, Plos. Biol., 2015
- Community-led reproducibility workshops https://www.repro4everyone.org/
- 1,500 scientists lift the lid on reproducibility, Nature, 2016
- Leading individuals and institutions in adopting open practices to improve research rigour http://bulliedintobadscience.org/

2.3 Some definitions

Experimental variability

In experimental sciences, variablility of the results is mainly due to:

- Biological variations
  - Random nature of measured phenomena;
  - Different subjects, organisms, samples.
- Technical variations
  - Small changes in experimental conditions;
  - Noise of measurement tool;
  - Sample preparation.
- Same interpretation?

Even with all things equal otherwise

2.4 Computational reproducibility

Computational variability

In data analysis: computers and programs are supposed to be exact!
⇒ perfect reproducibility? **Actually: No**

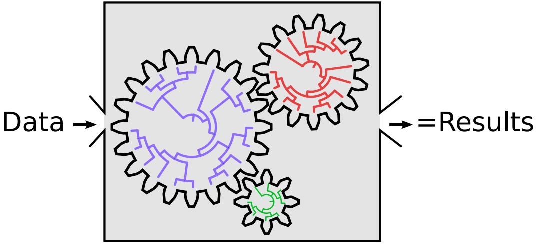
- Different versions of operating system;
- Different versions of tools used;
- Different hardware;
- Random nature of some algorithms (simulations, etc.);
- Numerical instability;
- Parallel algorithms;
- Poor method description;
- etc.

2.4 Computational reproducibility

Computational reproducibility

We can define several *levels* of reproducibility (Cohen-Boulakia et al., FGCS, 2017):

- Repeat: The data analysis experiment is performed in the exact same computational setting as the original experiment. In that case, results should be exactly the same without any variation. This necessitate to gather as many information as possible about the initial experiment, i.e. all tools versions, all operating system library versions, the state of the random number generator, etc.;

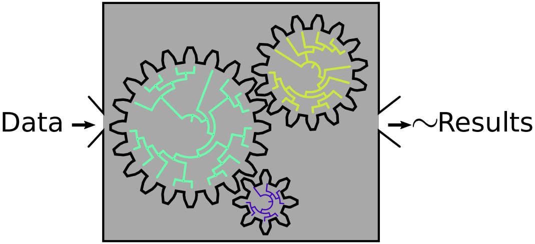


2.4 Computational reproducibility

Computational reproducibility

We can define several *levels* of reproducibility (Cohen-Boulakia et al., FGCS, 2017):

- Replicate: The data analysis experiment can be performed in a slightly different environment (different tool versions, different library versions, different random seeds, etc.), but the general protocol remains the same. In that case, results are not exactly the same, but scientific interpretation should be identical;

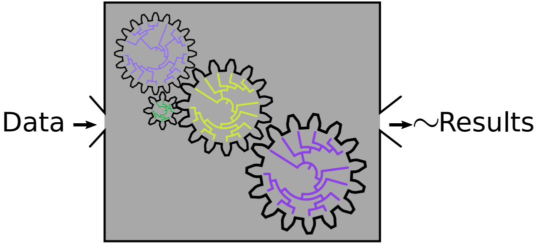


2.4 Computational reproducibility

Computational reproducibility

We can define several *levels* of reproducibility (Cohen-Boulakia et al., FGCS, 2017):

- Reproduce: The data analysis experiment aims at validating the scientific hypothesis, and can be performed in a different environment and with a different protocol (different tools, different workflow, etc.). This level of reproducibility gives us the best level of confidence about the quality of the results.



2.5 An so now...

Summary

Technical difficulties and some solutions to improve computational reproducibility:

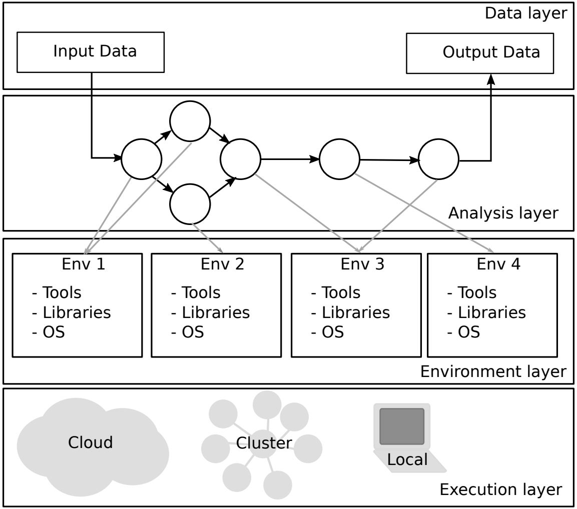
- Technical difficulties with usual practices
- Solutions:
  - Data management
  - Software/Script development
  - Environment management
  - Analysis development (workflows, notebooks, etc.)

PART 3

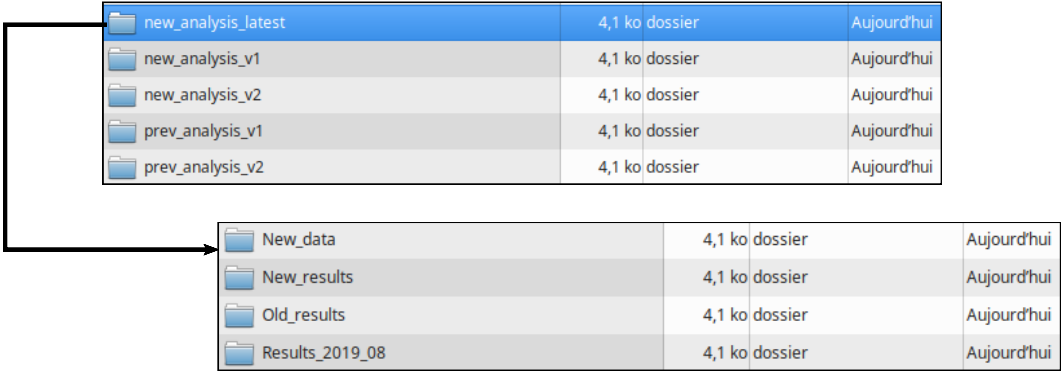
Difficulties and solutions

3.1 Whole analysis

Analysis stack



Data management: Difficulties



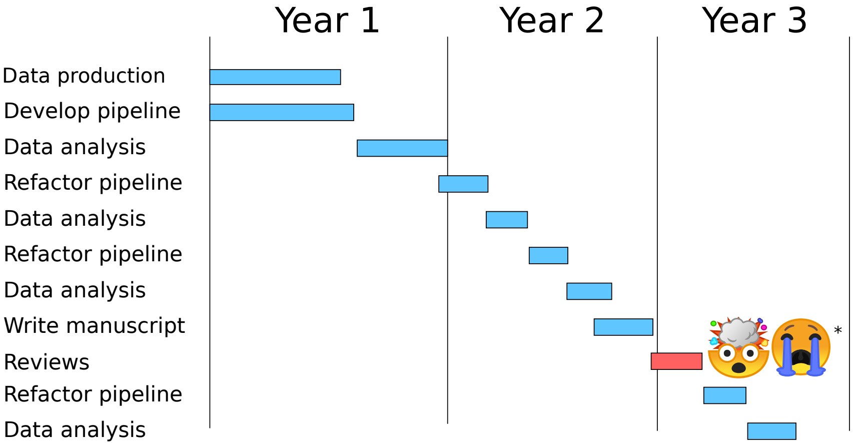
new_analysis_latest	4,1 ko	dossier	Aujourd'hui
new_analysis_v1	4,1 ko	dossier	Aujourd'hui
new_analysis_v2	4,1 ko	dossier	Aujourd'hui
prev_analysis_v1	4,1 ko	dossier	Aujourd'hui
prev_analysis_v2	4,1 ko	dossier	Aujourd'hui

New_data	4,1 ko	dossier	Aujourd'hui
New_results	4,1 ko	dossier	Aujourd'hui
Old_results	4,1 ko	dossier	Aujourd'hui
Results_2019_08	4,1 ko	dossier	Aujourd'hui

3.2 Data Management

Data management: Difficulties

Typical process:



* licence Apache, version 2.0
https://github.com/googlefonts/noto-emoji

3.2 Data Management

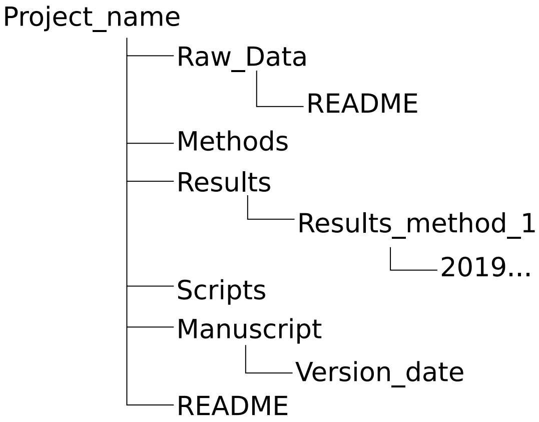
Data management: Difficulties

After Review:

- How was this figure generated?
- Where is the right data version???
- Where is the right script version???
- How was this file called???

Data management: Some solutions

- Directory structure



Project_name
Raw_Data
README
Methods
Results
Results_method_1
2019...
Scripts
Manuscript
Version_date
README

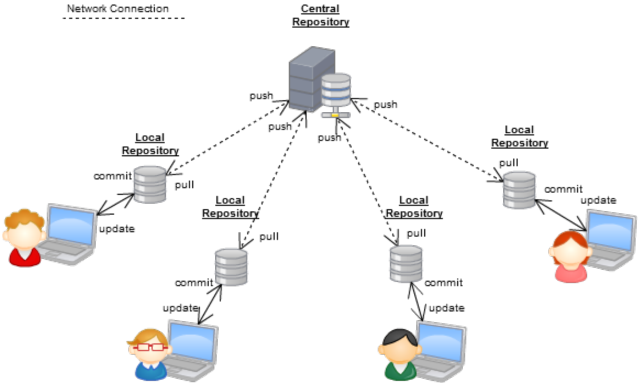
Inspired from https://www.repro4everyone.org

Data management: Some solutions

- File naming convention (from https://www.repro4everyone.org, CC BY)
  - Example: `Date_Project_Experiment_Type_ID_Version.xlsx`
- Backup!
- Versioning of scripts, manuscripts, etc.: Why not Git?

3.2 Data Management

What is a Distributed Version Control System (VCS)?



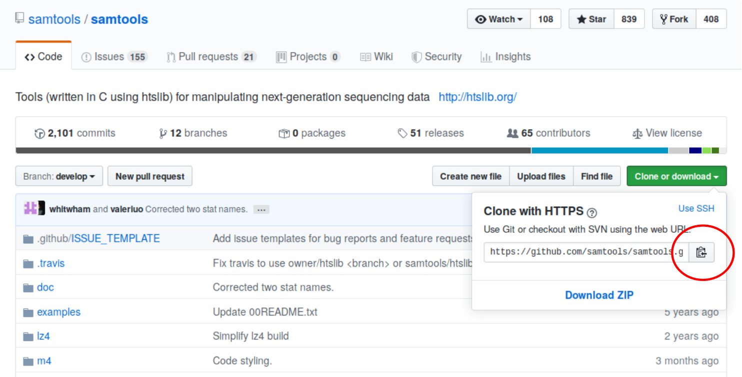
from http://www.tranthanhtu.vn/post/2017/01/11/git-branching-model

Slide from C3BI Tutorials https://github.com/C3BI-pasteur-fr/tutorials

Git Basics: Clone a github repository

GitHub

https://github.com/samtools/samtools



samtools / **samtools**

Watch 108 · Star 839 · Fork 408

<> Code · Issues 155 · Pull requests 21 · Projects 0 · Wiki · Security · Insights

Tools (written in C using htslib) for manipulating next-generation sequencing data http://htslib.org/

2,101 commits · **12** branches · **0** packages · **51** releases · **65** contributors · View license

Branch: **develop** · New pull request · Create new file · Upload files · Find file · Clone or download

whitwham and **valeriuo** Corrected two stat names. …

.github/ISSUE_TEMPLATE	Add issue templates for bug reports and feature requests	
.travis	Fix travis to use owner/htslib <branch> or samtools/htslib	
doc	Corrected two stat names.	
examples	Update 00README.txt	5 years ago
lz4	Simplify lz4 build	2 years ago
m4	Code styling.	3 months ago

Clone with HTTPS Use SSH

Use Git or checkout with SVN using the web URL.

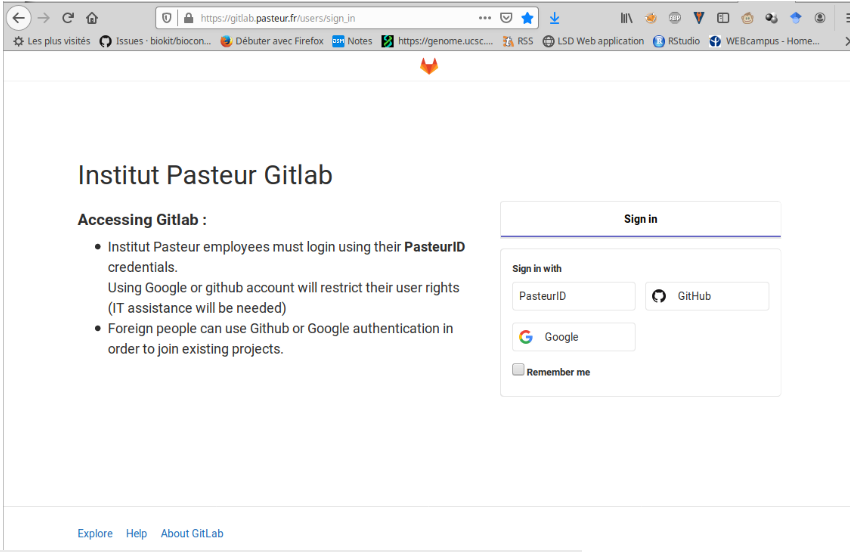
`https://github.com/samtools/samtools.g`

Download ZIP

```
$ git clone https://github.com/samtools/samtools.git
```

3.2 Data Management

Git at Pasteur: GitLab (`gitlab.pasteur.fr`: ∼ local GitHub)



Institut Pasteur Gitlab

Accessing Gitlab :

- Institut Pasteur employees must login using their **PasteurID** credentials.
  Using Google or github account will restrict their user rights (IT assistance will be needed)
- Foreign people can use Github or Google authentication in order to join existing projects.

Data management: Take home message

- Project Structure
- Versionning
- Storage/Backup

3.3 Software/Script development

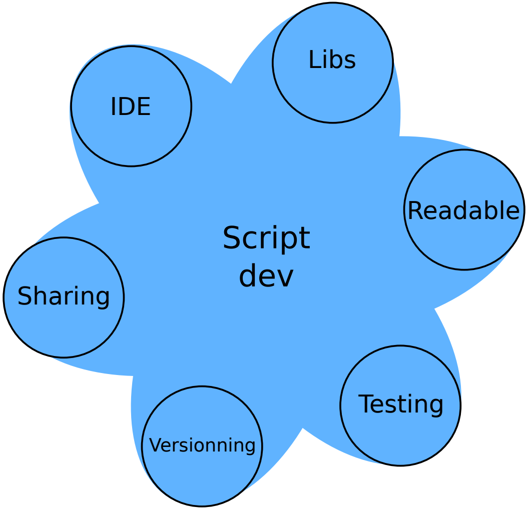
Difficulties

- Developing good code
- Do not reinventing the wheel
- Writing understandable code
- Do not introducing (too much) bugs
- Maintaining/Sharing the code
- Versioning the code (again)
- Executing, Deploying (dependencies, etc.)

Your software/script has at least ONE user: the future you, think about them!

3.3 Software/Script development

Coding: Good practices



3.4 Environment management

What is the computational environment?

The computational environment is made of:

- The operating system:
  - Type: MacOS, Linux: CentOS, RedHat, Ubuntu, Debian?
  - Version!
- The core libraries: Versions!
- The tools used: Versions and OS!
- Their dependencies: Versions and OS!

Difficulties

- LOTS of Informations needed to **capture** and **describe** the whole environment!
- Different tools may have conflicting dependencies...
- Keep old tools to rerun old analyses...
- Manage a huge tool repository...

3.4 Environment management

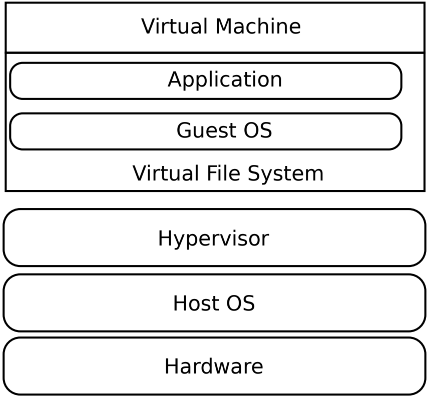
Computational environment: solutions

Several solutions:

- Virtual machines
- Conda / Bioconda
- Docker
- Singularity
- etc.

3.4 Environment management

Computational environment: Virtual Machines (VMs)



GPL v2
https://fr.wikipedia.org/wiki/Oracle_VM_VirtualBox

- Pros
  - Everything is included
  - No dependency problem
- cons
  - VMs include their own OS
  - VMs consume lots of resources
  - VMs are heavy: large storage
- Not ideal in a context of data analysis

3.4 Environment management

Computational environment: Conda/Bioconda

Conda: package and environment management system (`https://docs.conda.io/en/latest/miniconda.html`).

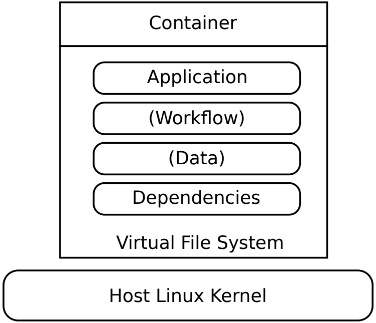
CONDA

- pros
  - Create several environments with their own tools and dependencies (VERY useful)
  - Keep as many environments as the number of old analyses
- cons
  - Still OS specific: tools may behave differently in MacOS and Linux...
  - Difficult to freeze all dependency versions

3.4 Environment management

Computational environment: Containers

Docker and Singularity allow to package applications/tools and their dependencies in an isolated “Container” that can be executed on any server.



- Lightweight! (dozens of MB for main single application containers)
- Fast and efficient: No need to boot a full OS
- Can keep all containers (very old anlyses, etc.)

3.5 Analysis development

Next: Analysis development

We know how to:

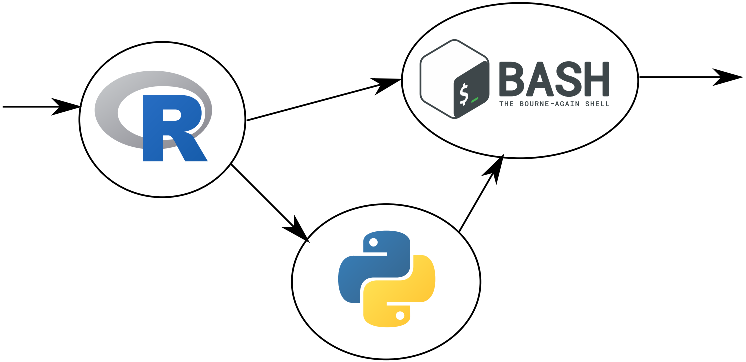
- Manage data
- Develop independent analysis scripts
- Manage environment (tool versions, etc.)

But...

How to link all of them together in a full analysis pipeline?

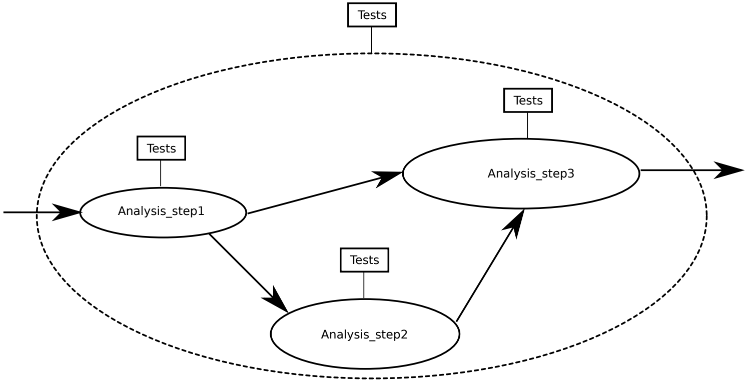
Analysis development: Difficulties

- Multi-language: R, python, bash, awk, etc.



Analysis development: Difficulties

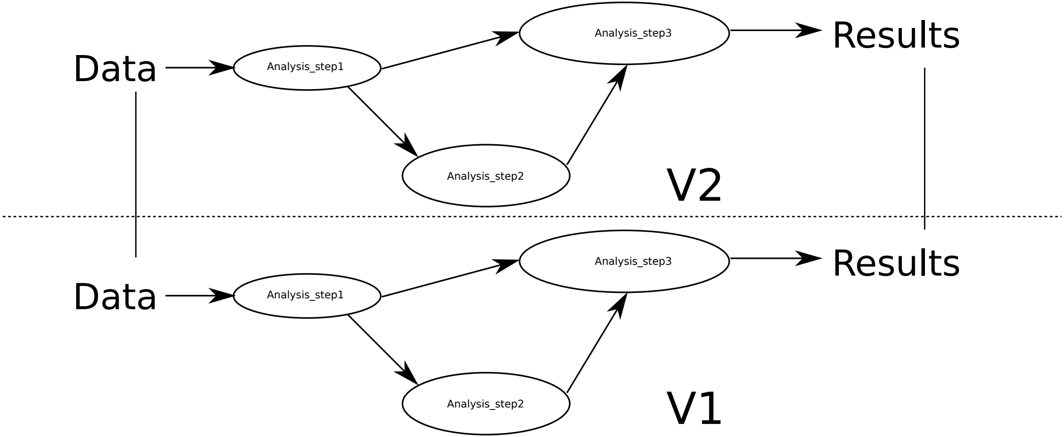
- Testing: How to test the whole pipeline?



3.5 Analysis development

Analysis development: Difficulties

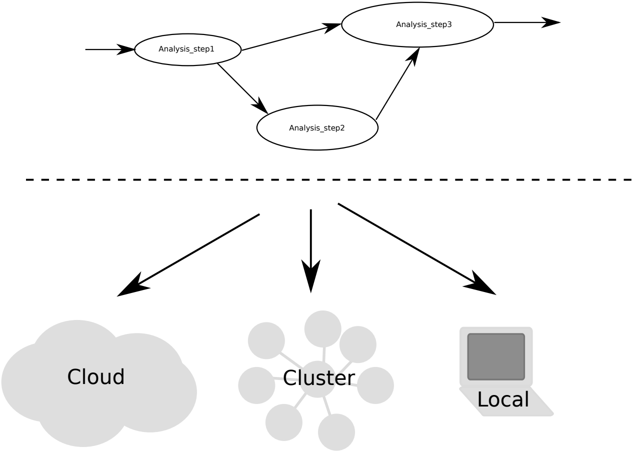
- Versionning



3.5 Analysis development

Analysis development: Difficulties

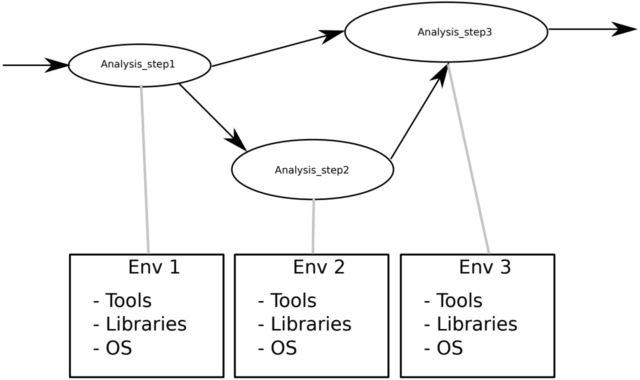
- Different execution machines: local, cluster, cloud, etc.



3.5 Analysis development

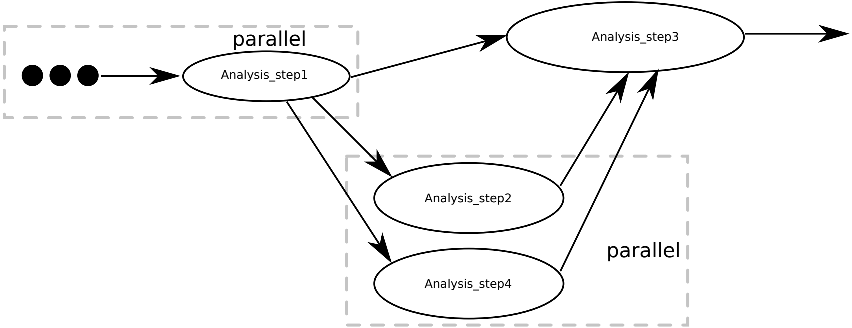
Analysis development: Difficulties

- High dependency to the environment (cluster scheduler, tool versions, etc.)



Analysis development: Difficulties

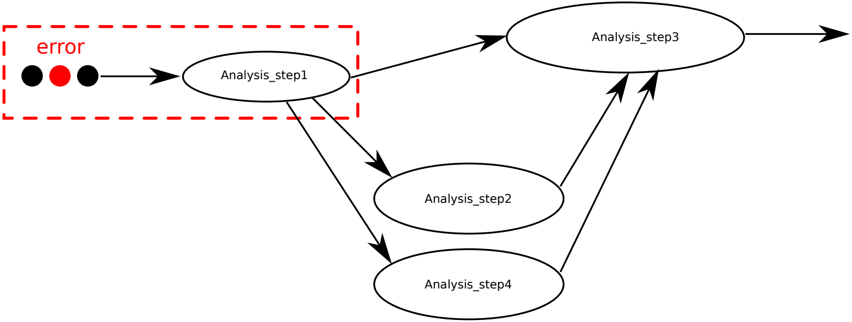
- Parallel processing



3.5 Analysis development

Analysis development: Difficulties

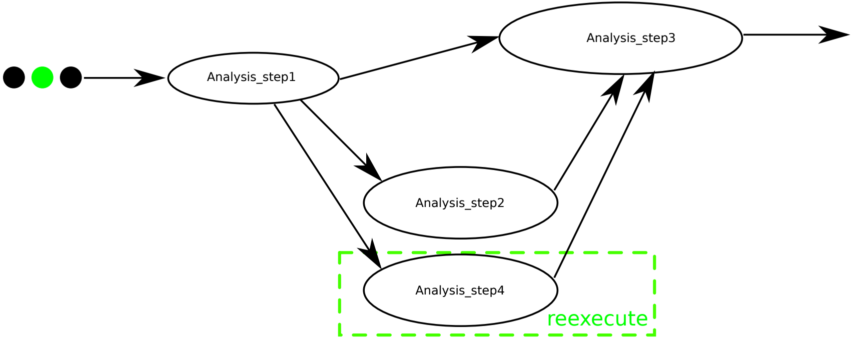
- Error handling



3.5 Analysis development

Analysis development: Difficulties

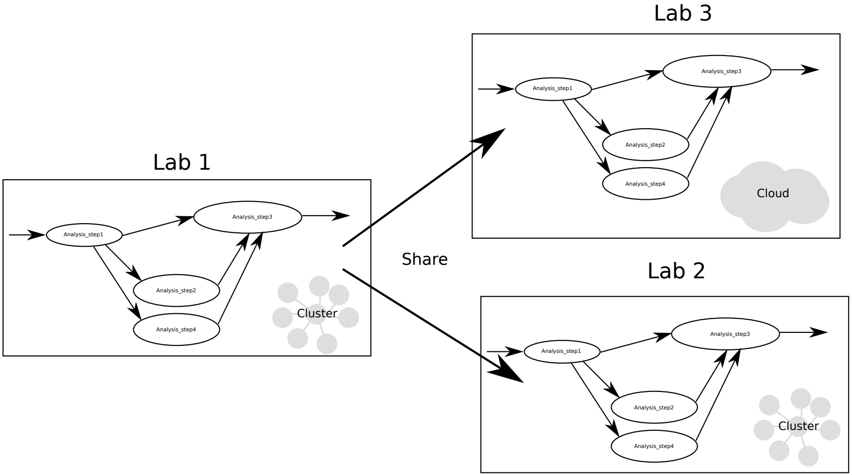
- Re-execution



3.5 Analysis development

Analysis development: Difficulties

- Maintenance, sharing, reuse, reproduce



3.5 Analysis development

Analysis development: Solutions

- Notebooks (Jupyter, RMarkdown)

- Workflow systems (Nextflow, Snakemake)

3.5 Analysis development

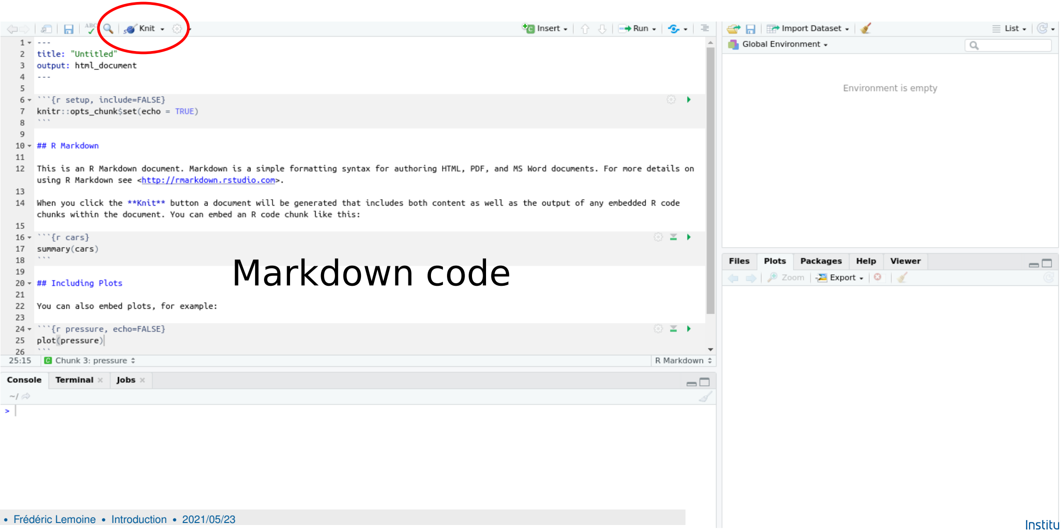
Analysis development: Notebooks

Interactive notebooks are a relatively new analysis technology which allows highly interactive data exploration, visualizations, exhaustive documentation, and sharing of an analysis process. Associated with the concept of Literate Programming, these notebooks are valuable tools towards better analyses comprehension and reproducibility.
There are many notebooks solutions to choose from:

- Jupyter: `https://jupyter.org/`
- R Markdown: `https://rmarkdown.rstudio.com/`
- Apache zeppelin: `https://zeppelin.apache.org/`
- Google Colaboratory: `https://colab.research.google.com`
- Observable (client-side): `https://observablehq.com/`
- Spark notebooks : `http://spark-notebook.io/`
- Beaker (engulfed by Jupyter)
- And more ...

Analysis development: RMarkdown

Mix R code, comments, and report in a text file (Versionning, etc.):



~~~
---
title: "Untitled"
output: html_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
~~~

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```{r cars}
summary(cars)
```

Including Plots

You can also embed plots, for example:

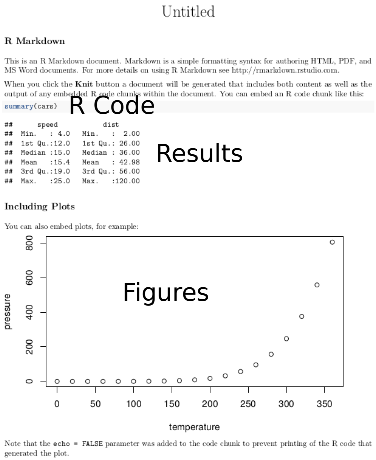
```{r pressure, echo=FALSE}
plot(pressure)
```
```

Markdown code

Institut Pasteur
```

## Analysis development: RMarkdown

Nice rendering (report or slides):



Untitled

### R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00  
```

Including Plots

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

3.5 Analysis development

Analysis development: Jupyter notebooks

Mix several languages, comments, and report in a unique notebook, with live execution:

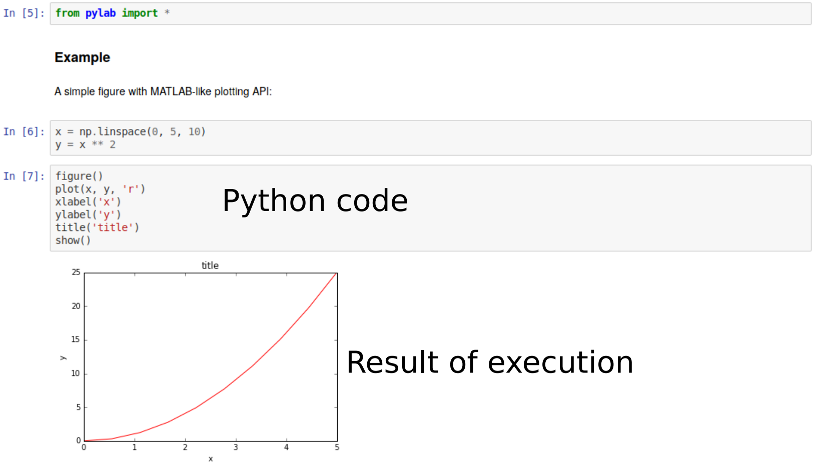
- support of 40+ languages
- extensibility (numerous plugins)
- recognition and pretty formatting by GitLab and GitHub
- helpful community

Getting started (with conda!):

```
conda create -n jupyter jupyter jupyterlab
conda activate jupyter
jupyter lab
```

Analysis development: Jupyter notebooks

Nice interactive rendering:



```
In [5]: from pylab import *
```

Example

A simple figure with MATLAB-like plotting API:

```
In [6]: x = np.linspace(0, 5, 10)
        y = x ** 2
```

```
In [7]: figure()
        plot(x, y, 'r')
        xlabel('x')
        ylabel('y')
        title('title')
        show()
```

Python code

Result of execution

3.5 Analysis development

Analysis development: Notebooks downsides

- They can be cumbersome (cell order execution...) for development and not necessarily a good first entry point in Programming
- They do not actively help in developing good coding practices (no proper module/library design)
- Source control difficulties
- Testing difficulties
- Not easy to use on different clusters...
- Not easy to use several environments (tool versions, etc.)

Analysis development: Workflow systems

In the past 20 years, bioinformatics workflow systems have been developed to solve lots of the difficulties described earlier. They allow to:

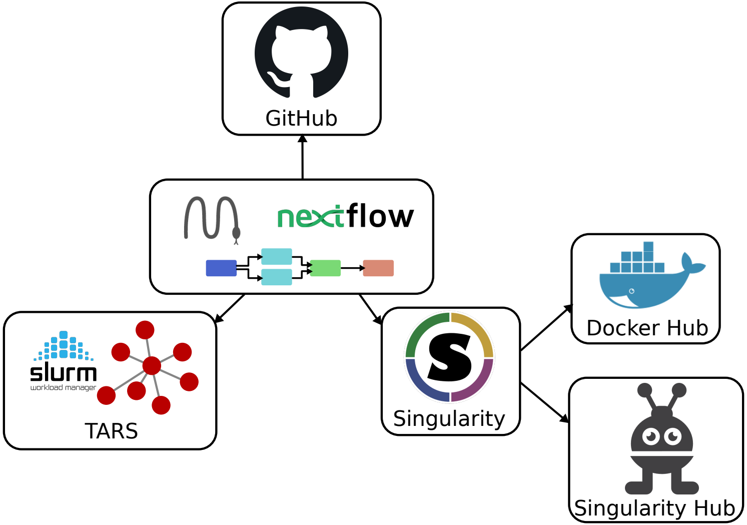
- Structure and develop their analysis pipelines
- Automate and monitor their execution
- Trace data flow and results
- Abstract from the execution machines and environment

Most popular options:

- Galaxy: `https://usegalaxy.org/`
- SnakeMake: `https://snakemake.readthedocs.io/en/stable/`
- Nextflow: `https://nextflow.io`

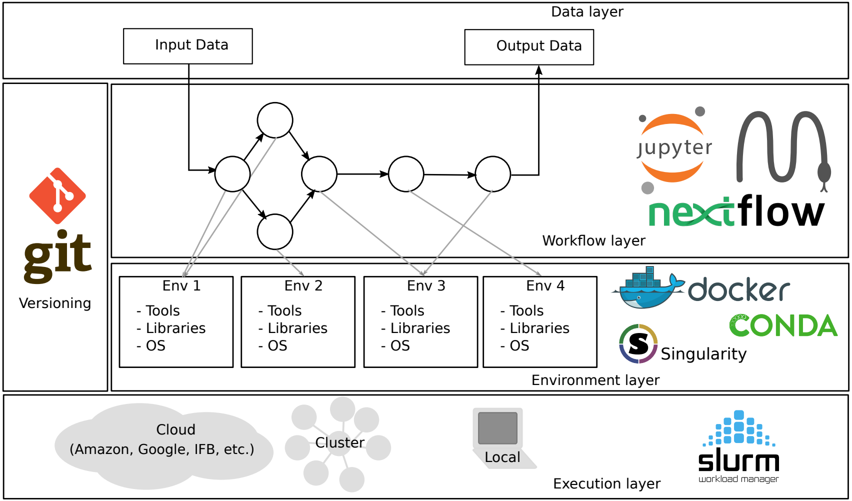
3.5 Analysis development

Analysis development: Workflow systems



3.6 Summary

Full stack



Institut Pasteur
25-28, rue du Docteur Roux
75724 Paris Cedex 15 - France
www.pasteur.fr