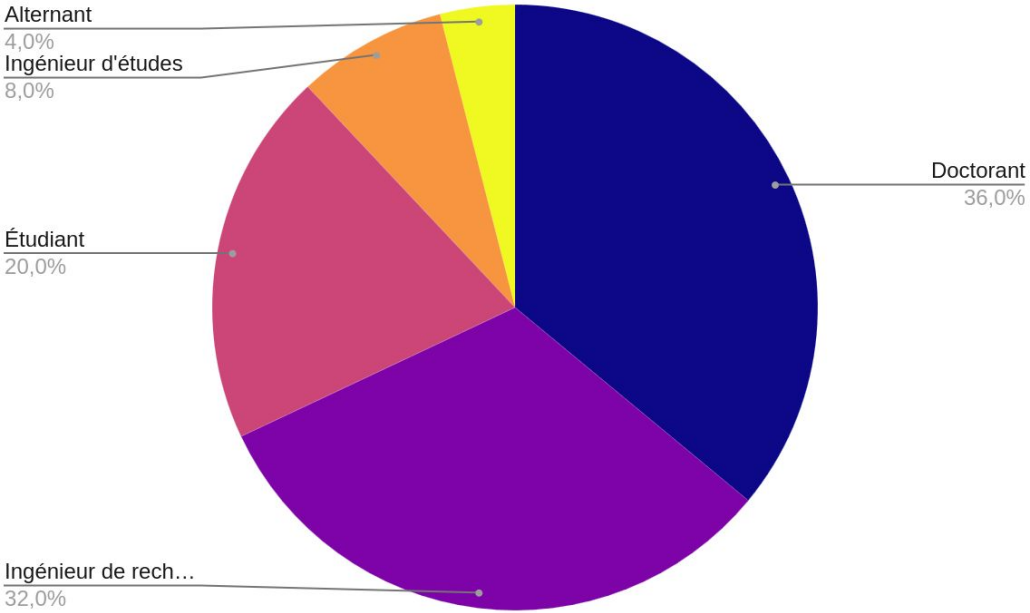
Good practices for reproducible bioinformatics data analysis

Sébastien Gradit, Elisabeth Hellec, Julien Fumey, Jérémy Rousseau, Benjamin Loire, Baptiste Imbert, Arthur Durante, Samuel Ortion, Ravy Leon Foun Lin, Maxime Ulysse Garcia, Savandara Besse

Participants



Société Française de Bioindormatique

XXXX

Program

1. Introduction: Reproducibility crisis
2. Git - https://github.com/jebif/reprohackathon-jobim2024/tree/main/resources/01_git
3. Conda - https://github.com/jebif/reprohackathon-jobim2024/tree/main/resources/02_conda
4. Nextflow & Snakemake - https://github.com/jebif/reprohackathon-jobim2024/tree/main/resources/03_workflow
5. Docker - https://github.com/jebif/reprohackathon-jobim2024/tree/main/resources/04_docker
6. It's your turn to code - https://github.com/jebif/reprohackathon-jobim2024/tree/main/workflow

Reproducibility crisis

The reproducibility crisis refers to a phenomenon in scientific research where **many published studies cannot be replicated or reproduced** by other researchers. The crisis arises due to several factors:

Reproducibility crisis

The reproducibility crisis refers to a phenomenon in scientific research where **many published studies cannot be replicated or reproduced** by other researchers. The crisis arises due to several factors:

- **Publication Bias**: Journals tend to favor publishing studies with positive results over those with negative or inconclusive findings. This can create an inflated perception of the reliability of certain findings.

Reproducibility crisis

The reproducibility crisis refers to a phenomenon in scientific research where **many published studies cannot be replicated or reproduced** by other researchers. The crisis arises due to several factors:

- **Publication Bias**
- **Methodological Issues**: Some studies may have flaws in their experimental design, data analysis, or statistical methods, making it difficult or impossible for other researchers to reproduce the results.

Reproducibility crisis

The reproducibility crisis refers to a phenomenon in scientific research where **many published studies cannot be replicated or reproduced** by other researchers. The crisis arises due to several factors:

- **Publication Bias**
- **Methodological Issues**
- **Lack of Transparency**: Insufficient reporting of methods and data in published studies can hinder replication efforts. Without access to all the necessary information, other researchers may struggle to replicate the findings.

Reproducibility crisis

The reproducibility crisis refers to a phenomenon in scientific research where **many published studies cannot be replicated or reproduced** by other researchers. The crisis arises due to several factors:

- **Publication Bias**
- **Methodological Issues**
- **Lack of Transparency**
- **Selective Reporting**: Researchers may selectively report only the results that support their hypotheses while omitting contradictory findings, leading to an inaccurate portrayal of the research.

Reproducibility crisis

The reproducibility crisis refers to a phenomenon in scientific research where **many published studies cannot be replicated or reproduced** by other researchers. The crisis arises due to several factors:

- **Publication Bias**
- **Methodological Issues**
- **Lack of Transparency**
- **Selective Reporting**
- **Pressure to Publish**: There is often pressure on researchers to publish frequently and in prestigious journals, which may incentivize rushing through experiments or selectively reporting results to meet publication deadlines.

Reproducibility crisis

The reproducibility crisis refers to a phenomenon in scientific research where **many published studies cannot be replicated or reproduced** by other researchers. The crisis arises due to several factors:

- **Publication Bias**
- **Methodological Issues**
- **Lack of Transparency**
- **Selective Reporting**
- **Pressure to Publish**

Addressing the reproducibility crisis requires efforts to improve research practices, promote transparency, and encourage replication studies. This includes preregistration of study protocols, sharing of data and code, conducting replication studies, and reevaluating incentive structures within academia.

Reproducibility crisis

The reproducibility crisis refers to a phenomenon in scientific research where **many published studies cannot be replicated or reproduced** by other researchers. The crisis arises due to several factors:

- **Publication Bias**
- **Methodological Issues** 
- **Lack of Transparency**
- **Selective Reporting**
- **Pressure to Publish**

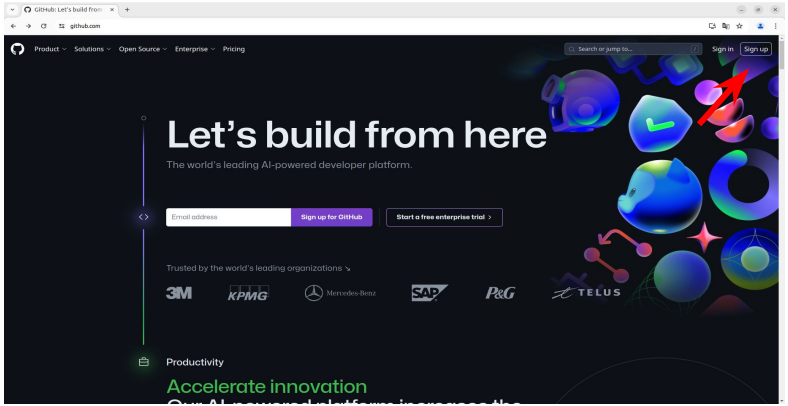
Addressing the reproducibility crisis requires efforts to improve research practices, promote transparency, and encourage replication studies. This includes preregistration of study protocols, sharing of data and code, conducting replication studies, and reevaluating incentive structures within academia.

Git

XXXX

Git

- Git :
  - Created by Linus Torvalds in 2005
  - Decentralized versioning software (free and open source)
- GitHub :
  - Web hosting and software development management service
  - Owner: Microsoft



Git

IT'S UP TO YOU

Go to GitHub

https://github.com/jebif/reprohackathon-jobim2024/tree/main/resources/01_git

Conda

XXXX

Conda

- **Conda :**
  - **Characteristics : open-source, cross-platform, package manager and environment management system**
  - **Developed to solve package management challenges (Python and R)**
- **Anaconda :**
  - **Distribution of the Python and R programming languages for scientific computing**
  - **Simplify package management and deployment**
- **Miniconda :**
  - **Small version of Anaconda**

Conda

IT'S UP TO YOU

Go to GitHub

https://github.com/jebif/reprohackathon-jobim2024/tree/main/resources/02_conda

Nextflow

Maxime Ulysse Garcia

Snakemake

Sébastien Gradit

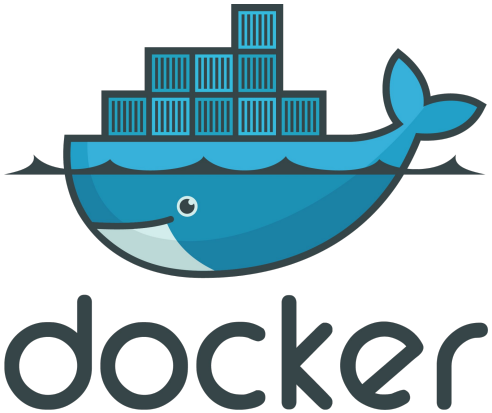
Nextflow & Snakemake

IT'S UP TO YOU

Go to GitHub

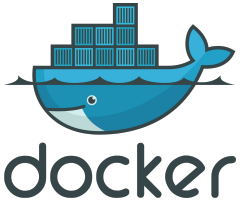
https://github.com/jebif/reprohackathon-jobim2024/tree/main/resources/03_workflow

Docker



Benjamin Loire & Jérémy Rousseau

Project

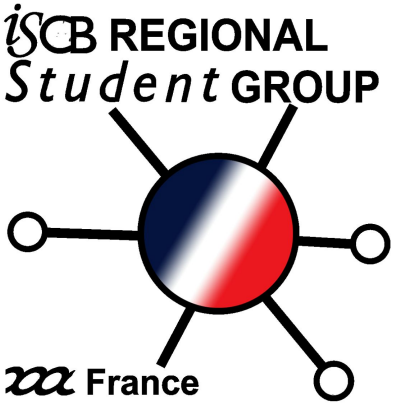


It's your turn to code

Go to GitHub

https://github.com/jebif/reprohackathon-jobim2024/tree/main/workflow

What's JeBiF ?



XXXX

THE END !