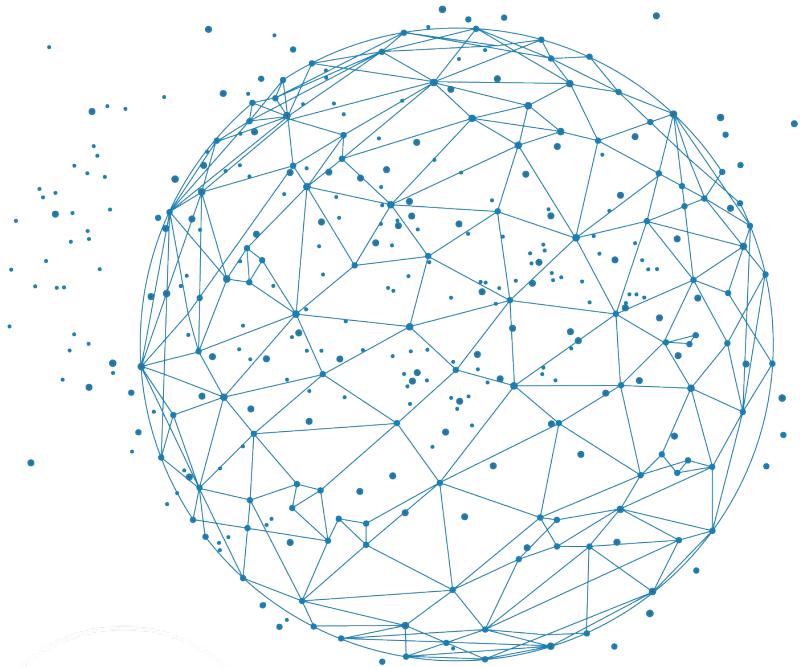


9 au 11 octobre 2023, Institut des Systèmes Complexes, Paris

FAIR Bioinfo 2023

Les principes FAIR dans un
projet de bioinformatique

Conclusion



INSTITUT FRANÇAIS DE BIOINFORMATIQUE



INRAE



Inserm





- Introduction to FAIR & Open Science (webinaire, questions)
- Project management (data, repository)
- Traceability with notebooks (jupyterlab)
- History management (Git)
- Introduction to encapsulation
- Environment management (Conda)
- Containerisation (docker)
- Share & disseminate, code & project (GitHub)
- HPC : cluster (Slurm)
- Analysis workflow (snakemake)
- HPC : containerisation (singularity)
- Conclusion



- Introduction to FAIR & Open Science (webinaire, questions)

- Project management (data, repository)
- Traceability with notebooks (jupyterlab)
- History management (Git)

e-labbook

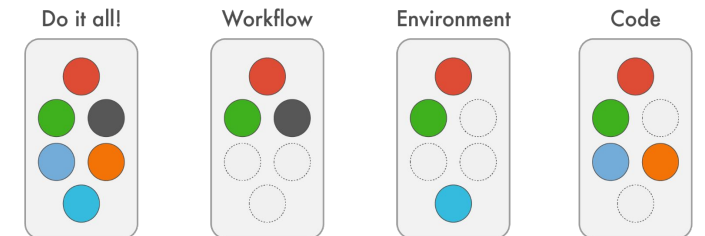
- Introduction to encapsulation
- Environment management (Conda)
- Containerisation (docker)
- Share & disseminate, code & project (GitHub)

code development

- HPC : cluster (Slurm)
- Analysis workflow (snakemake)
- HPC : containerisation (singularity)
- Conclusion

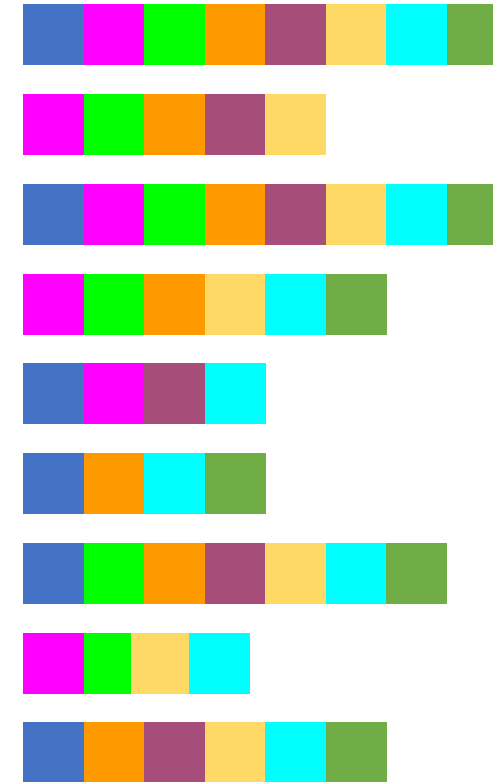
HPC

- Introduction to FAIR & Open Science (webinaire, questions)
- Project management (data, repository)
- Traceability with notebooks (jupyterlab)
- History management (Git)
- Introduction to encapsulation
- Environment management (Conda)
- Containerisation (docker)
- Share & disseminate, code & project (GitHub)
- HPC : cluster (Slurm)
- Analysis workflow (snakemake)
- HPC : containerisation (singularity)
- Conclusion





- Project management (data, repository)
- Traceability with notebooks (jupyterlab)
- History management (Git)
- Environment management (Conda)
- Containerisation (docker)
- Share & disseminate, code & project (Git***)
- HPC : cluster (Slurm)
- Analysis workflow (snakemake)
- HPC : containerisation (singularity)



Céline (NGS)

Claire (NGS analysis)

Emilie (NGS)

Gildas (admin)

Julien (Dev)

Pauline (WGA)

Hugo (Dev & NGS)

Thomas (Dev)



F 
indable

A 
ccessible

I 
nteroperable

R 
eusable

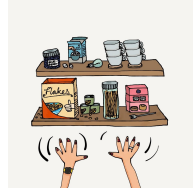




F
Findable



A
Accessible



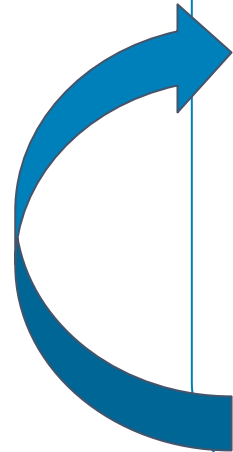
I
Interoperable



R
Reusable



A virtuous
cycle



Objective:

FAIR raw data

+

FAIR scripts

=

FAIR processed data





Automation

Manual command lines



Write a shell script



Use a workflow manager



Tests and continuous integration (*)

User analysis (trial-and-error)

Offer a GUI (eg. with R-Shiny) (*)



Save and re-import choices (*)

Softwares

Local installation



Package manager



Conda environment



Image / container



Virtual machine (*)

(*) not in the course

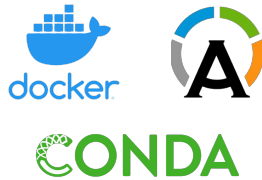


Reproductibilité

Versionner
Partager
Archiver



Créer un
environnement
virtuel



Installer des
outils par un
gestionnaire



Créer un
script
d'analyses



Déporter
l'analyse sur
un serveur



Portabilité
de
l'exploration
des résultats



Éditer un
rapport
d'analyses



Alternative tools:





Reproducibility to the exact bit?

container uses some resources of the support machine

⇒ version control of the env.
(Nix, Guix)

HPC and parallelization?

loss of computational order,
multithreading, identical hardware?

⇒ ...





Pedagogical team (our guardian angel): H el ene

IFB Core Cluster taskforce: Julien, Gildas, and all those who provide in the shadows

Helpers: Emilie, Marie, Lucie

Organisations:

CNRS, INRAE, IFB, I2BC, Paris Saclay University, CEA