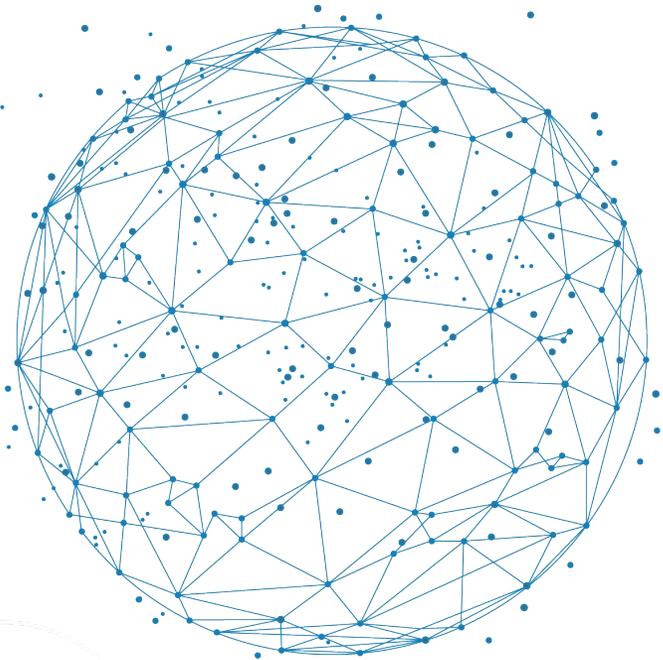


9 au 11 octobre 2023, Institut des Systèmes Complexes, Paris

# FAIR Bioinfo 2023

## Les principes FAIR dans un projet de bioinformatique

**Equipe pédagogique** : H. Chiapello, T. Denecker, C. Hernandez, G. Le Corguillé, J. Seiler, C. Toffano-Nioche  
**Helpers** : L. Khamvongsa-Charbonnier, E. Drouineau, M. Cariou



INSTITUT FRANÇAIS DE BIOINFORMATIQUE





## Objectifs

- Rendre une analyse bioinformatique **reproductible**
- Introduire et pratiquer les principes “**FAIR**” et outils associés
- **Mise en pratique sur l’infrastructure IFB**

Facilement  
découvrable



Acessible



Interopérable



Réutilisable



## Bref historique

- **Création en 2019** à l’I2BC (C. Toffano-Nioche et T. Denecker)
  - Publication : <https://doi.org/10.21105/jose.00068>
- **Partenariat avec l’IFB depuis 2020** et mise en pratique sur l’infrastructure IFB
  - [Edition 2020](#) du 31 août au 2 sept à l’ISC Paris
  - [Edition 2021](#) du 28 au 30 juin à l’ISC Paris
  - [Edition 2022](#) du 13 au 15 juin à l’ISC Paris
  - Edition 2023 du 9 au 11 octobre à l’ISC Paris



## LE NNCR : National Network of Computing Ressources

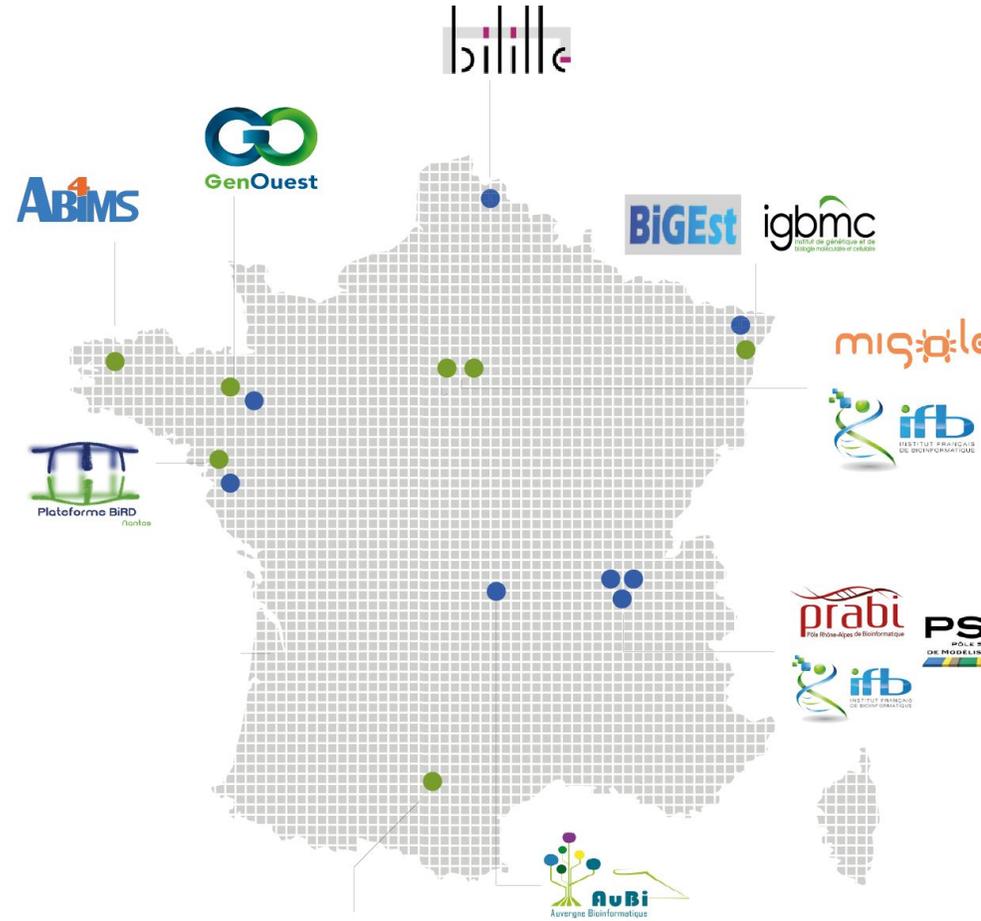
2 infrastructures complémentaires : **Cloud** et **Cluster**

Réseau national composé :

- D'équipements de plateformes régionales IFB
- De 2 infrastructures Core (IFB core cluster + IFB core cloud)

Administré par une "task force" mutualisée

- Les plateformes partagent X% d'ETP d'ingénieurs
- Bénéfice mutuel des % d'ETP partagés
  - Forte émulation du personnel
  - Éviter la déperdition d'énergie
  - Partager l'expérience
  - Partager les développements
  - Traitement collaboratif des problèmes
  - Solutions robustes



### cluster

- 6 SLURM Clusters
- 16,900 vCPU
- 92 TB RAM
- 10.5 PB
- 10 GPU cards
- 4935 yearCPU in 2020

8 sites clouds avec 8,000 vCPU et 36To RAM

● Cluster  
● Cloud



Catalogue de nos formations : <https://www.france-bioinformatique.fr/formations>

## FAIR bioinfo

### Objectifs

- > Fondamentaux de la **reproductibilité** en bioinformatique
- > Introduire et pratiquer les outils qui permettent d'adopter les principes "FAIR"
- > Mise en pratique sur l'infra IFB

### Modalités

- > **Présentiel** à Paris, payant
- > 2 jours + 1 (optionnel pour futurs formateurs)
- > 4 éditions 2020, 2021, 2022

## FAIR data

### Objectifs

- > Fondamentaux de l'**Open Data** (pratiques, juridiques)
- > Le PGD dans un contexte bioinformatique
- > Les standards de métadonnées pour les « omiques »

### Modalités

- > **Distanciel** synchrone, gratuit
- > 4 ou 5 demi-journées
- > 2 éditions distancielles nationales en 2021, 2 éditions présentiels régionales en 2022, 1 édition en partenariat avec IBISA en 2023



Librement inspiré de : <https://openlifesci.org/code-of-conduct>

- **Soyons amicaux et patients.** Nous sommes ici entre collègues. Demain vous serez (peut-être) formateur
- **Soyons accueillant.** Pas de discrimination d'aucune sorte ici
- **Soyons prévenants.** Respectez le travail des autres
- **Soyons respectueux.** Restons calme en cas de désaccord. Essayons de résoudre les points de désaccord de manière constructive. Par ailleurs l'erreur est humaine... et reproductible (en général)

Toutes les questions sont bienvenues !

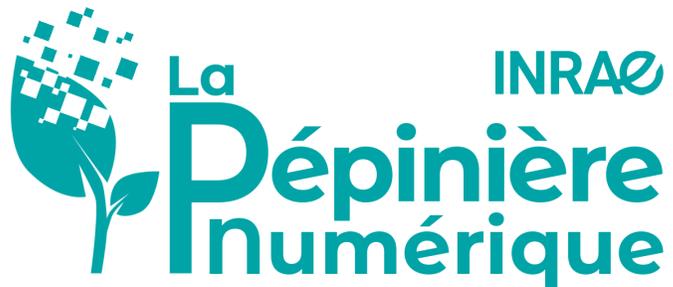
# Nouveauté 2023 : le webinaire introductif

Lien : <https://www.youtube.com/watch?v=FDOIIIUh2Ew>

Diapos : <https://zenodo.org/record/8392952>

Vos retours :

<https://postit.colibris-outilslibres.org/WebinaireFAIRbioinfo>



Rechercher

CC BY SA

Introduction à la reproductibilité en bioinformatique

H. Chiapello, IFB, <https://orcid.org/0000-0001-5102-0632>  
T. Denecker, IFB, <https://orcid.org/0000-0003-1421-7641>  
C. Hernandez, I2BC, <https://orcid.org/0000-0001-8664-1340>  
C. Toffano-Nioche, I2BC, <https://orcid.org/0000-0003-4134-6844>

0:07 / 33:28

IFB Institut Français de Bioinfo... 31 abonnés S'abonner

0 Partager

## Introduction à la reproductibilité en bioinformatique

Non répertoriée



Institut Français de Bioinfo...  
31 abonnés

S'abonner

0



Partager



14 vues il y a 2 jours

Formation IFB à propos de la reproductibilité en bioinformatique

Retrouvez les Slides de cette formation ici : <https://zenodo.org/record/8392952> ...afficher plus



Céline Hernandez, I2BC



Claire Toffano-Nioche, I2BC



Thomas Denecker, IFB



Gildas Le Corguillé, IFB



Julien Seiler, IFB



Hélène Chiapello, IFB



Emilie Drouineau, I2BC



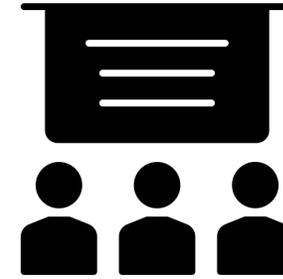
Lucie Khamvongsa-Charbonnier, IFB



Marie Cariou, MHN



## Les participants



Vous ? Votre équipe / unité ?

Votre première expérience de la reproductibilité et des principes FAIR ?

Vos attentes ?

+ Une question sur le contenu du webinaire





## Tout est sur moodle [ici](#)

### ▪ **Lundi 9 octobre - 9:30-17:30**

Accueil, présentation de la formation, tour de table

Retour sur le webinaire

#### **Usecase 1 - e-labbook**

*Organiser son projet, Notebook, concept des liens entre texte et code (JupyterLab local - Démo live)*

*Versionning de code (Git)*

#### **Usecase 2 - Développement (en local, exposition)**

*Premiers pas vers un environnement contrôlé : Conda. Exposer son travail (GitHub/GitLab). Intégration github et jupyter*

*DOI avec Zenodo. GitHub Pages*

### ▪ **Mardi 10 octobre- 9:30-17:00**

Distribuer son code et l'environnement (Docker)

#### **Usecase 3 - cluster**

*Cluster de calcul de l'IFB (Jupyterlab, slurm). Workflow (Snakemake). Singularity*

*Conclusion - Questionnaires évaluation*

### ▪ **Mercredi 11 octobre- 9:30-16:00 (journée optionnelle pour futur formateurs)**

*Bilan des évaluations des participants et discussions sur les points d'amélioration*

*Témoignages et retour d'expériences d'éditions régionales ou de réutilisation de FAIR bioinfo (Pierre Marin, Marie Cariou, Jacques Dainat, Pauline François)*

*Les principes FAIR appliqués au matériel pédagogique (L. Khamvongsa-Charbonnier)*

*Travail sur le matériel de cours (diapos, exercices)*

*Conclusion*

***Déjeuner sur place avec plateaux repas :  
12:45-13:45***



1- Contexte : Pouvez-vous citer 1 ou 2 objectifs de la science ouverte ?

2- Contexte : Pour qui fait-on de la science ouverte ?

3- Contexte : Quelle est votre expérience de la science ouverte ? Quelle (expérience) pratique en avez-vous ?

4- Contexte : Avez-vous déjà échoué à reproduire une expérience/analyse ? L'une de vos collaborateurs ? L'une des vôtres ?

5- Contexte : Quels sont les bénéficiaires directs de la reproductibilité ?

6- Disp. Nat. : Quelle est la spécificité du 2<sup>e</sup> plan national pour la science ouverte ?

7- Disp. Nat. : Quelle est la différence entre atelier de la donnée et centre de référence thématique ?

8- Disp. Nat. : Quels interlocuteurs identifiez-vous dans le domaine de la Science Ouverte et de la reproductibilité ?

9- Concepts : Reproducible research, Repeatability, Replicability, Reproducibility, Replication, ... : même définition / sens ? Pourquoi ?

10- Concepts : Pouvons-nous gérer du code comme on gère des données ?

11- Concepts : Pouvons-nous appliquer les principes FAIR des données aux codes ?



12- Proposition : Pensez-vous qu'il est obligatoire de tout faire tout ou pouvons-nous n'en faire qu'une partie ?

13- Proposition : Est-ce que les outils sont interchangeables ?

14- Proposition : Combien utilisez-vous d'outils au quotidien ? Séparément ou ensemble ?

15- Proposition : Quelle(s) différence(s) entre GitHub/GitLab et Software Heritage ?

16- Questions ouvertes : Vous sentez vous concernés par ces questions ? Les avez-vous déjà entendues/vécues ?