

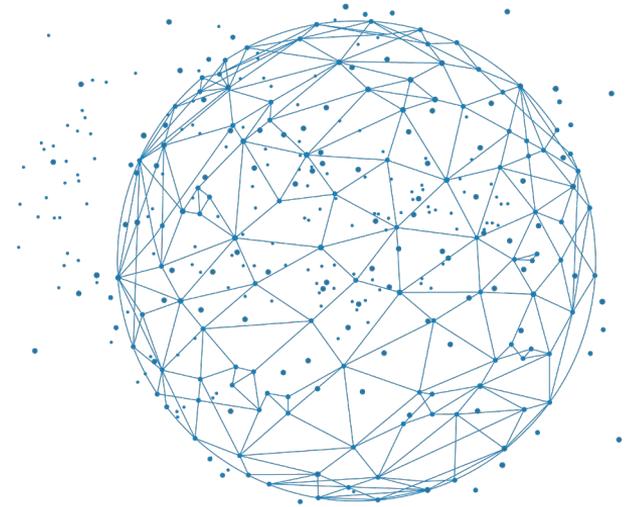
Les ateliers du GIS IBISA

Atelier PGD : session 2

Paris, les 2 et 3 février 2023

Frédéric de Lamotte
Hélène Chiapello

Introduction

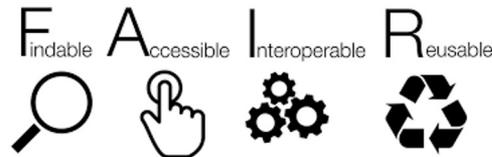




- **2020** : montage de la formation par un GT IFB pendant le confinement
- **2021** : 2 éditions “nationales” en distanciel synchrone en mars et octobre
- **2022** : 2 éditions “régionales” IFB + 1 édition IBISA en présentiel
 - édition “francilienne” - Sorbonne Univ - 22 et 23 mars
 - édition “Grand Est” - Strasbourg les 23 et 24 mars
 - session 1 IBISA en octobre
- **2023** : 2-3 février - édition avancée ciblée pour les responsables de plateformes à la demande d'IBISA



H. Chiapello T. Denecker J.-F. Dufayard P. Lieby L. Maurel G. Sarah J. Seiler F. de Lamotte



À la fin de cette formation, les participants auront

- acquis **des connaissances théoriques et pratiques** sur la gestion selon les principes FAIR de leurs données dans le contexte de la **Science Ouverte**
- initié un **PGD de structure** sur DMP Opidor
- identifié **des pistes d'amélioration** pour la gestion des données de leur plateforme.





■ Horaires :

jeudi 2 février : 10h00 - 18h30 *repas assis 12h30-14h*

vendredi 3 février : 9h00-16h30 *repas buffet 12h30-13h30*

■ En présentiel... donc **interactif** ! Nous avons planifié

Beaucoup de **séances pratiques** en petits groupes

Des **séquences d'échanges** : n'hésitez pas à partager vos expériences sur le sujet

A vos claviers

■ Un outil pour poser des questions : **slido** Join at [slido.com](https://www.slido.com) code : #3914

143

■ Un espace **moodle** pour récupérer les supports de cours :

<https://moodle.france-bioinformatique.fr/course/view.php?id=16>

slido

moodle



Organisateurs & Intervenants



Hélène Chiapello



Frédéric de Lamotte



Thomas Denecker



Célia
Michotey



Jordan Troesch

Témoignages de responsables de plateformes



Bertrand Vernay



Charles Poncet



Céline Hernandez



Oana Vigi



Ana Paula Texeira



Julien Seiler (IFB guest star)



Pascale Mercier (IBISA
super organisatrice)



- > Prénom
- > Votre plateforme
- > Localisation géographique
- > Types de données produites
- > Vos attentes (en 2 mots)



Jeudi 2 février 2023

10h00-11h00 Introduction (F. de Lamotte & H. Chiapello)

11h00-11h30 Pause

11h30-12h30 Session - PGD de structure (C. Michotey)

Déjeuner (à table)

14h00-15h30 Témoignages de plateformes

La gestion des données au sein de plateformes :

- NGS : **C. Hernandez**
- Protéomique : **O. Vigy**
- Imagerie : **A.-P. Texeira**
- Gentyane : **C. Poncet**

15h30-16h00 Pause

16h00-18h00 Atelier Sécurité des données (J. Troesch)

et séance de mise en pratique

18h00-18h30 Conclusion par le GIS IBiSA (B. Robert)

Vendredi 3 février 2023

9h00-10h30 Atelier Métadonnées (T. Denecker et al.)

Plénière puis 3 sous-groupes : acides nucléiques/NGS (T. Denecker), protéines/structures (O. Vigy), images (B. Vernay)

10h30-11h00 Pause

11h00-12h30 Atelier les produits de recherche du PGD de structure (C. Michotey et al.)

Exemples de PGD de structure puis travail en sous-groupes sur les produits de recherche du PGD (DMP Opidor)

Déjeuner (buffet)

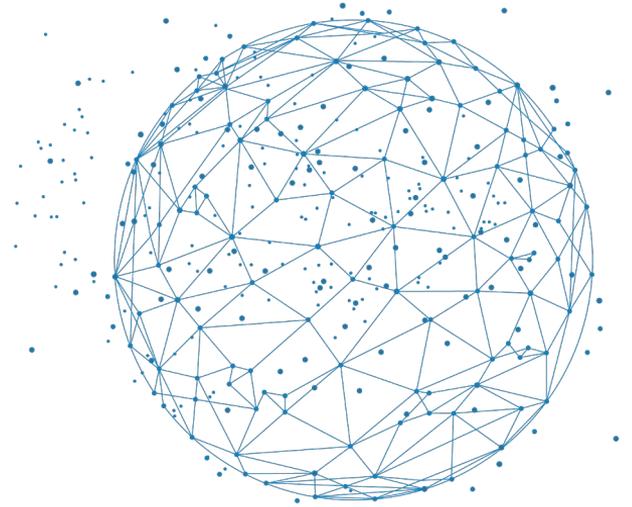
13h30-14h30 Principes FAIR & pistes d'amélioration (F. de Lamotte & T. Denecker)

14h30-15h00 Pause

15h00-16h00 Un exposé invité : OpenLink (J. Seiler)

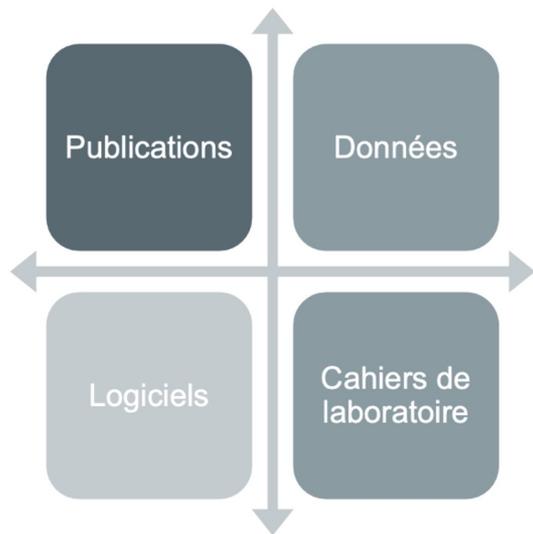
16h00-16h30 Conclusion (H. Chiapello & F. de Lamotte)

Cadre légal

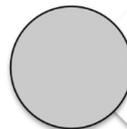




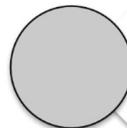
- Un des objectifs : rendre les **résultats de la recherche scientifique** accessibles à tous
- En permettant une diffusion et une réutilisation sans entrave des résultats de la recherche



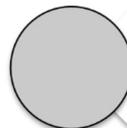
Les ambitions :



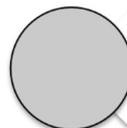
démocratiser l'accès aux savoirs



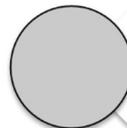
rendre la science plus cumulative, plus fortement étayée par les données, plus transparente



augmenter l'efficacité de la recherche en évitant de dupliquer les efforts, en réutilisant des données ou du matériel scientifique



favoriser les avancées scientifiques et l'innovation



favoriser la confiance des citoyens dans la science

Les données de la recherche sont des informations publiques :

- elles sont soumises à un principe d'ouverture par défaut et de libre utilisation (*Loi Lemaire - LPRN 2016*)
- elles sont soumises à un principe de gratuité (*Loi Valter 2015*)
- elles sont protégées contre les risques d'accaparement



2018



2021

Mesures



Montage du projet

Paragraphe sur la gestion des données

Budgéter la science ouverte (APC, data manager, stockage...)



Début du projet

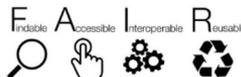
Plan de gestion des données (PGD)



Pendant le projet

Gestion des données suivant les principes FAIR

Mise à jour du PGD



Fin du projet

Partage des données « aussi ouvert que possible, aussi fermé que nécessaire »

Entrepôts de données
Open Access des publications





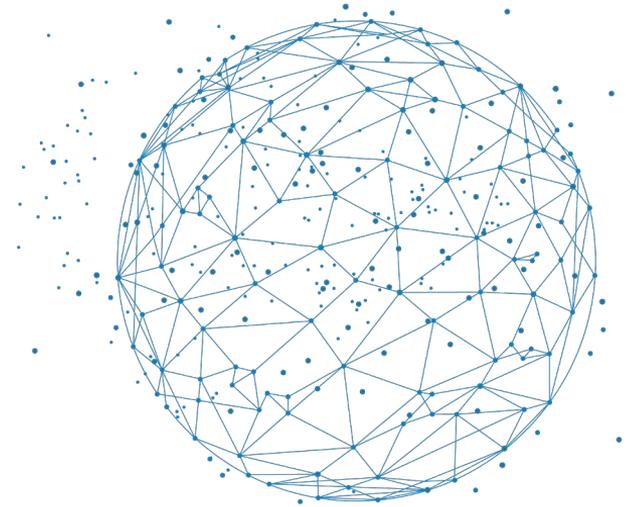
Les données produites et gérées selon les principes FAIR sont considérées comme un produit majeur des recherches menées à INRAE. **Partagés largement**, les données et leurs usages sont un **gage de l'intégrité** requise pour des résultats reconnus par la communauté scientifique et la société ;

La pratique de la réutilisation des données produites par des tiers doivent conduire à ne produire de nouvelles données **que lorsque cela est nécessaire**, gage d'efficacité et source de collaborations ;

Les données et les activités associées constituent une source de connaissance, de reconnaissance et d'échange pour la recherche, l'innovation, et l'expertise et l'appui aux politiques publiques.

Cette transformation repose sur une implication de l'ensemble des agents et des organisations. Pour cela, il est nécessaire d'accompagner ces changements de culture, d'accroître les compétences en science et gestion des données, et d'adapter notre organisation et notre stratégie partenariale pour accompagner les différentes communautés.

Éléments de contexte Science Ouverte : ressources & acteurs



Infrastructure numérique



- National Network of Computing Resources (NNCR)
- Calcul intensif, deux modalités: **cluster + cloud**
- Protection des données (GDPR)

11 Po stockage
21K cœurs

Accompagnement des usagers



- Un point d'entrée unique pour tous les services et toutes les plateformes IFB
- Soutien aux projets dès la conception
- Plans de gestion des données (PGD)
- Conception et implémentation de **workflows**
- Science des données

400 experts
>200 FTE

9115 utilisateurs
310 projets
collaboratifs

Outils logiciels



- Développement de logiciels spécialisés
- Packaging (conda)
- Virtualisation (appliances, containers)
- Bonnes pratiques d'ingénierie logicielle
- Développement collaboratif, ouverture du code

Plus de 900 outils
disponibles

Bases de connaissances



- Développement de bases de données et de connaissances
- Standards de qualité
- Support technique aux développeurs de BD
- Déploiement de bases de données développées en France

Formations



- Ecoles thématiques (NGS, multi-omiques, phylogénie, biostat, programmation, workflows...)
- Formations **Bring Your Own Data (BYOD)**
- Webinaires, MOOC
- **Nouveau: formations FAIR-data et FAIR-bioinfo**
- Adaptation permanente à l'évolution de la demande

136 formations
2326 personnes formées
341 jours de formation



Mutualised Task Forces



Prospective et innovation



- Identification des besoins émergents en bioinfo
- Projets-pilotes pour répondre aux besoins :
 - Bioinformatique intégrative
 - Intégration des données de santé
 - Intelligence artificielle pour la biologie-santé

Science ouverte & Interopérabilité



- Gestion des données
- Interopérabilité, standards, ontologies...
- Courtage des données
- PGD automatisables
- Coopération avec les infra productrices de données



Le MESRI lance le plan [Recherche.data.gouv.fr](https://www.recherche.data.gouv.fr) en 2022

Un dispositif complet d'accompagnement des équipes de recherche

- [Ateliers de la donnée](#) : des services généralistes sur tout le territoire
- [Centres de référence thématiques](#) : expertises par domaine scientifique
- [Centres de ressources](#) : un appui aux ateliers de la donnée et aux centres de référence thématique

L'IFB est désigné Centre de Référence des Données de biologie-santé.



Cet écosystème consiste à fédérer les actions des dispositifs existants et à amplifier leur présence et leurs actions sur tout le territoire et dans tous les domaines thématiques

L'IFB est désigné Centre de Référence des Données de biologie-santé.

Attentes du MESRI

- Établir développer des **standards**
- **Gestion** des données
- **Soin** à la donnée
- **Moissonnage** des données pour le catalogue
- **Courtage** des données

→ **Animation** des communautés biologie santé, des autres INBS, **co-construction des standards**



Juillet 2022 : un premier CRT en biologie et santé ...

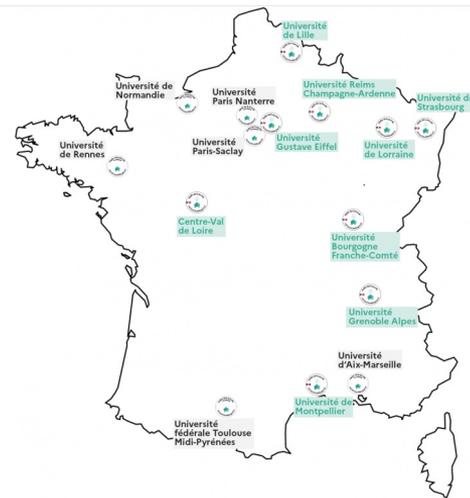
... d'autres suivront, notamment pour

- les données de santé,
- les données biologiques qui ne relèvent pas de la bioinformatique

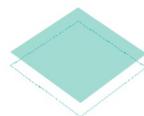


- Les ateliers de la donnée : des services généralistes sur tout le territoire en proximité géographique des équipes de recherche pour leur apporter une première expertise dans la gestion raisonnée des données de recherche
- Les centres de ressources : un appui aux ateliers de la donnée et aux centres de référence thématique

Ateliers de la donnée



Centres de ressources



Centre de ressources
entrepôt-catalogue
Recherche Data Gov



Centre de ressources
supports pédagogiques, e-formations
Recherche Data Gov



Les **ADAC** (Administrateur des Données Algorithmes et Codes de la recherche)

Le **délégué à la protection des données** (DIL) ou **Data Protection Officer (DPO)**

Les **Référents Données** dans les Instituts (RDS et RDO) ou **data steward**

Les Responsables de la Sécurité des Systèmes d'Information (**RSSI**)

- + un tas d'autres : PRADA, responsables valorisation & partenariats, responsables qualité,..

Gouvernance des données

Protection des données

Bonnes pratiques de gestion des données

Sécurité des données

Valeur des données

- La CNIL <https://www.cnil.fr/>
- Le site Doranum <https://doranum.fr>
- Le site Ouvrir la Science
<https://www.ouvrirlascience.fr/accueil/>
- Le site RDMkit <https://rdmkit.elixir-europe.org/>





Merci & bonne formation !



INSTITUT FRANÇAIS DE BIOINFORMATIQUE

