

L'Institut Français de Bioinformatique : Centre de Référence Thématique en biologie et santé

Jean-François Dufayard



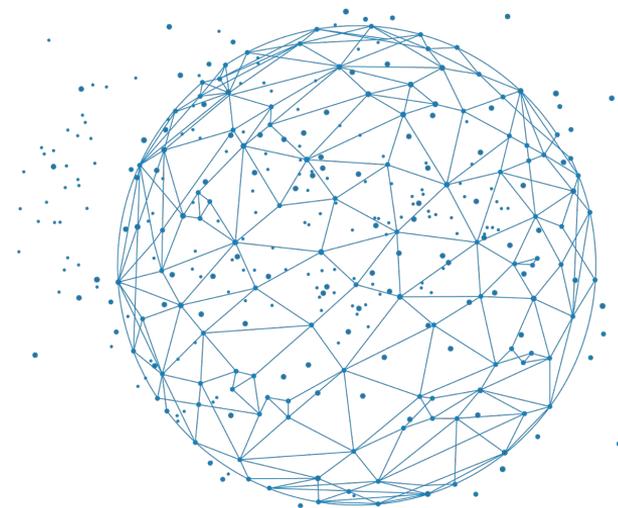
Action "réseaux des Infrastructures Nationales en
Biologie et Santé" de l'IFB



UMR AGAP - département BIOS - référent données



I. Qu'est-ce-que l'IFB ?

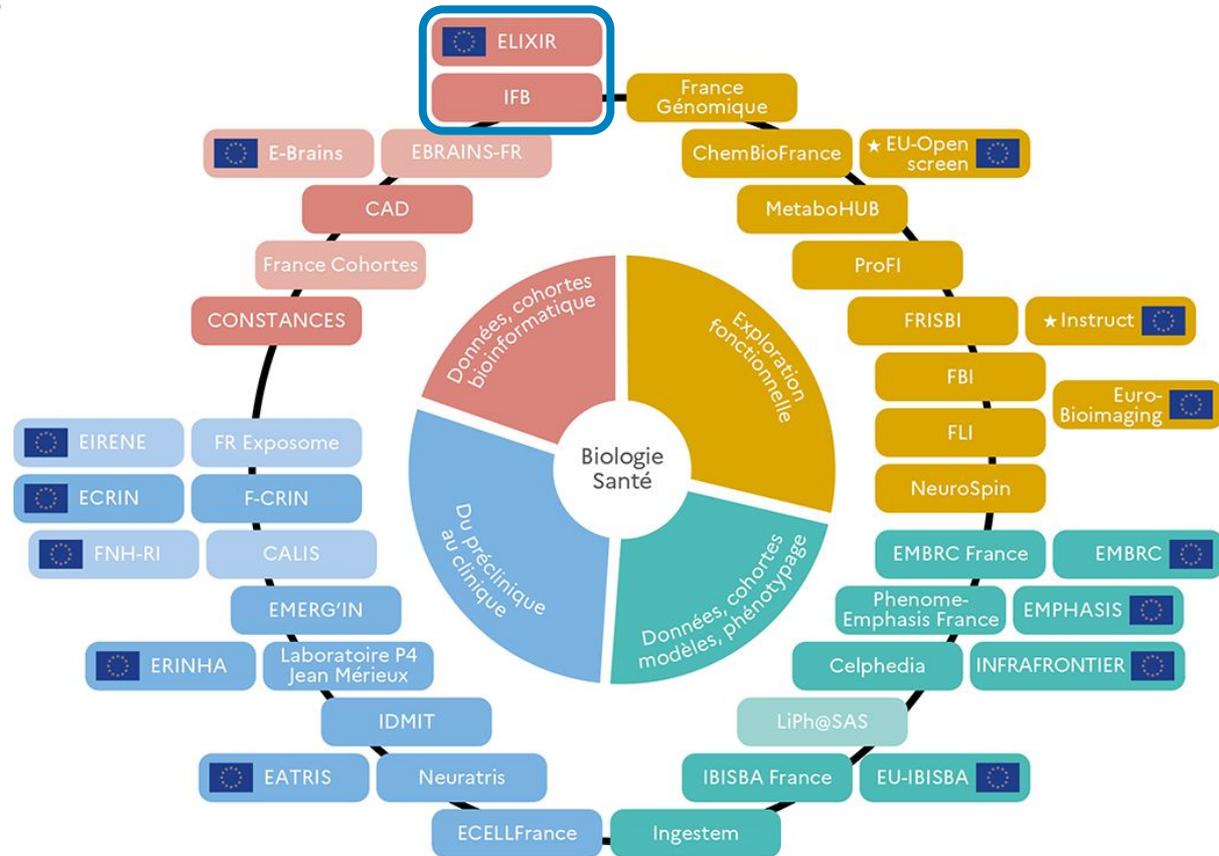


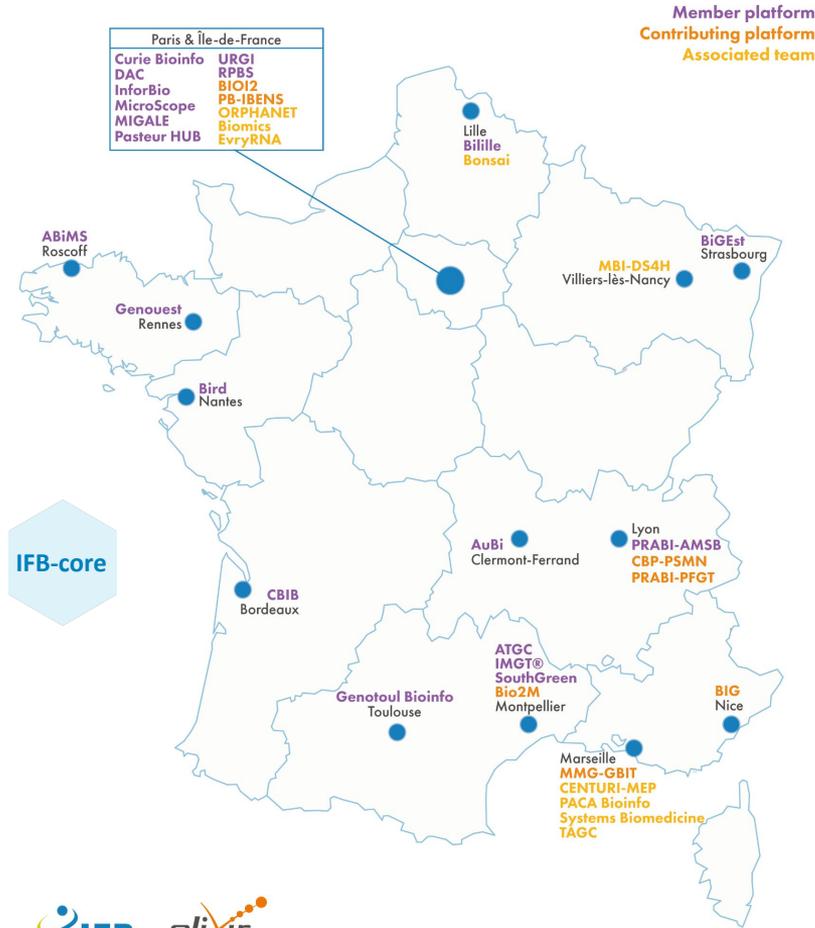
L'IFB/ELIXIR-FR, une Infrastructure Nationale en Biologie Santé (INBS)



Depuis 2012, le gouvernement français a investi dans le déploiement d'infrastructures nationales de recherche

- l'IFB/ELIXIR-FR est une **Infrastructure Nationale de Biologie-Santé (INBS)**
- 2022 - Attribution du rôle de **Centre de référence thématique biologie-santé (CRT)** dans l'écosystème Recherche Data Gouv



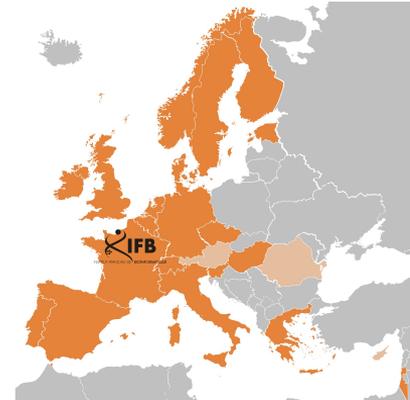


Une infrastructure décentralisée sur toute la France

- 20 plateformes membres
- 7 plateformes contributrices
- 8 équipes associées
- 1 unité de coordination (IFB-core)

Une intégration dans l'Europe

- Nœud français d'ELIXIR Europe
- Point d'entrée pour les projets européens



INNOVATION, R&D transversale dans les domaines d'expertises de la **bioinformatique** et pour les communautés des **sciences de la vie** incluant l'environnement et la santé :

- **Régional** - plateformes et équipes associées pour le développement de services
- **National** - avec les actions et communautés scientifiques IFB

FINANCEMENTS

Projets scientifiques nationaux & européens :

- ABRomics (Antibio-résistance)
- AgroServ (agriculture et écologie)
- AtlaSea : ByteSea (Biodiversité marine)
- Cloud4SAMS (métagénomique et santé)
- Neurovasc (santé et numérique)

- Ferments du futur (alimentation et santé)
- ELIXIR - STEERS (cycle de vie des données)
- ...

Projets d'infrastructures nationales

- RENABI
- MuDiS4LS (équipement de recherche)
- ABBA

FRANCE

Développement de services en bioinformatique :

- **Infrastructure numérique** (clusters et cloud)
- Développement de **logiciels**
- **Formations**
- **Support à des projets d'analyse de données** avec des biologistes
- Accompagnement et mise en œuvre de la **science ouverte** pour les données en biologie-santé
- **Cycle de vie des données**

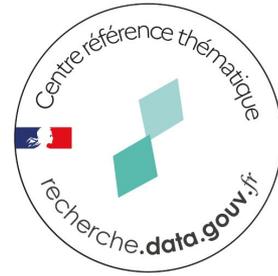
EUROPE

Mise à disposition de services nationaux :

Clusters / Cloud / formations / ...

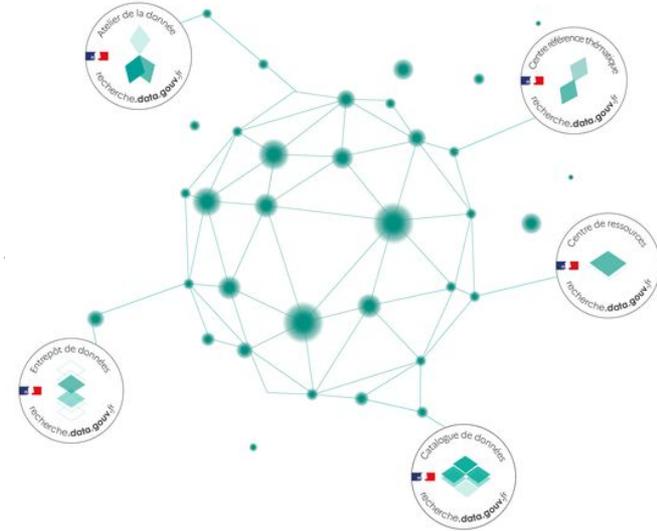
Contribution au développement de services mutualisés :

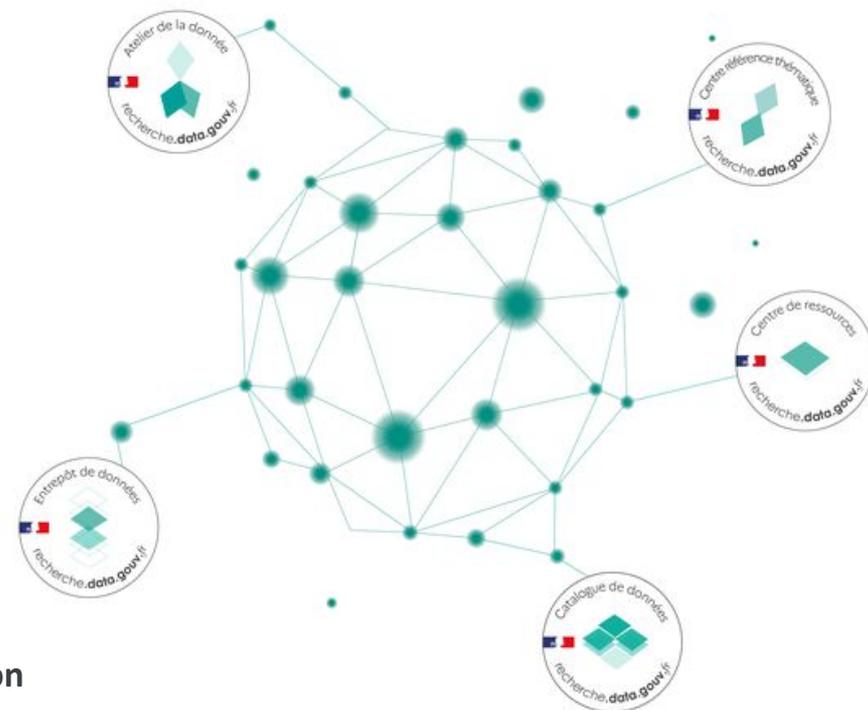
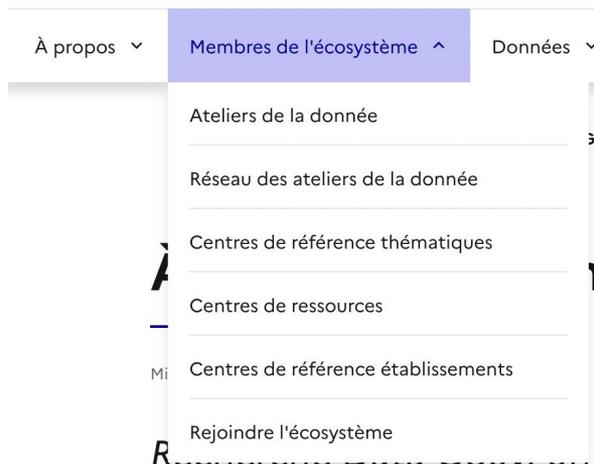
- **SDP ELIXIR** (Bio.tools / EDAM / OrphaData / ...)
- Accompagnement et mise en œuvre de la **science ouverte** pour les données en biologie-santé



II. IFB comme centre de référence thématique en biologie et santé

<https://recherche.data.gov.fr/fr>





Un écosystème d'accompagnement dans la bonne gestion des données, pour la communauté scientifique nationale

Centre de référence thématique

- Apporte une expertise par domaine
- Définie et dissémine
 - Des référentiels et bonnes pratiques
 - Des standards internationaux pour la gestion, le partage et la réutilisation de données de la recherche
 - Des entrepôts recommandés
- Collaborer avec les autres dispositifs
 - Ateliers de la donnée
 - Centres de ressources
 - ...
- Soutenir les communautés
 - e.g. proposer des outils et services spécifiques

RÉPUBLIQUE FRANÇAISE recherche.data.gouv.fr
Laboratoire National de l'Humain

Les centres de référence thématiques

Centre-référence thématique
recherche.data.gouv.fr

IFB
INSTITUT FRANÇAIS DE BIOINFORMATIQUE

PRO GE DO
DATA INFRASTRUCTURE

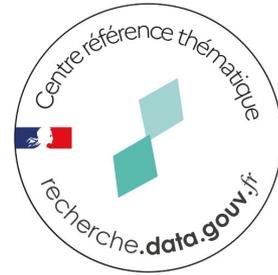
DATA TERRA

PND B Pôle National de Données de Biodiversité

HN Huma-Num IR*

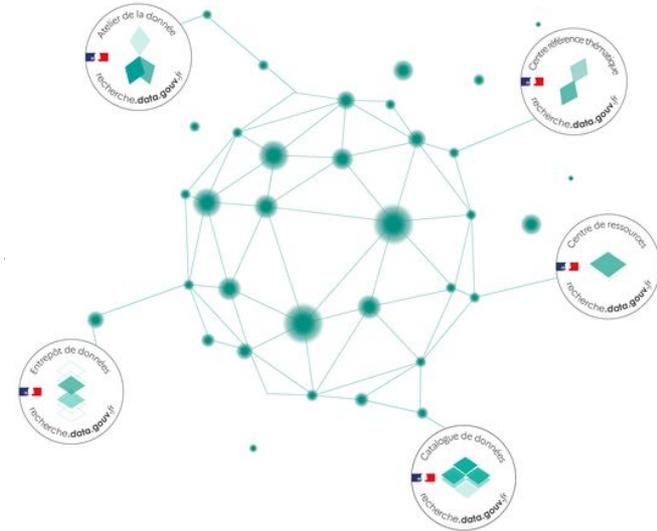
15

in @RechercheDataG



II. IFB comme centre de référence thématique en biologie et santé

A) Impliquer les communautés



Contexte - quelles communautés, quelles données ?



Beaucoup de domaines d'applications en science de la vie

- Santé
- Agronomie
- Environnement
- Microbiologie
- ...

Large panorama de données hétérogènes

- Multi-omique (méta, geno, transcripto, métabo, protéo...)
 - Séquences
 - Quantitatives
 - Structurales
 - ...
- Bioimagerie
- Spectres
- Données structurées hétérogènes

Interopérabilité avec d'autres domaines (écologie, sciences humaines...)

Les communautés sont légion et leurs pratiques et maturités en gestion des données varient grandement

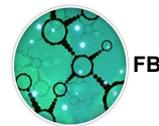
Le rôle de l'IFB comme animateur est capital



Niveau national : Les Infrastructures Nationales en Biologie et Santé (INBS), structurées au sein GIS IBISA et organisées dans le “Club des INBS”

Création d'un groupe “données de la recherche” au sein du Club des INBS:

- Se coordonner
- Inventaire des données et entrepôts
- Inventaire et état de l'art des différentes pratiques
- Préparer et co construire les référentiels de FAIRisation, implémentés avec les ressources nationales



...
25+

Niveau européen: Initiatives ELIXIR (réseau des *Research Data Managers*, RDMkit...)

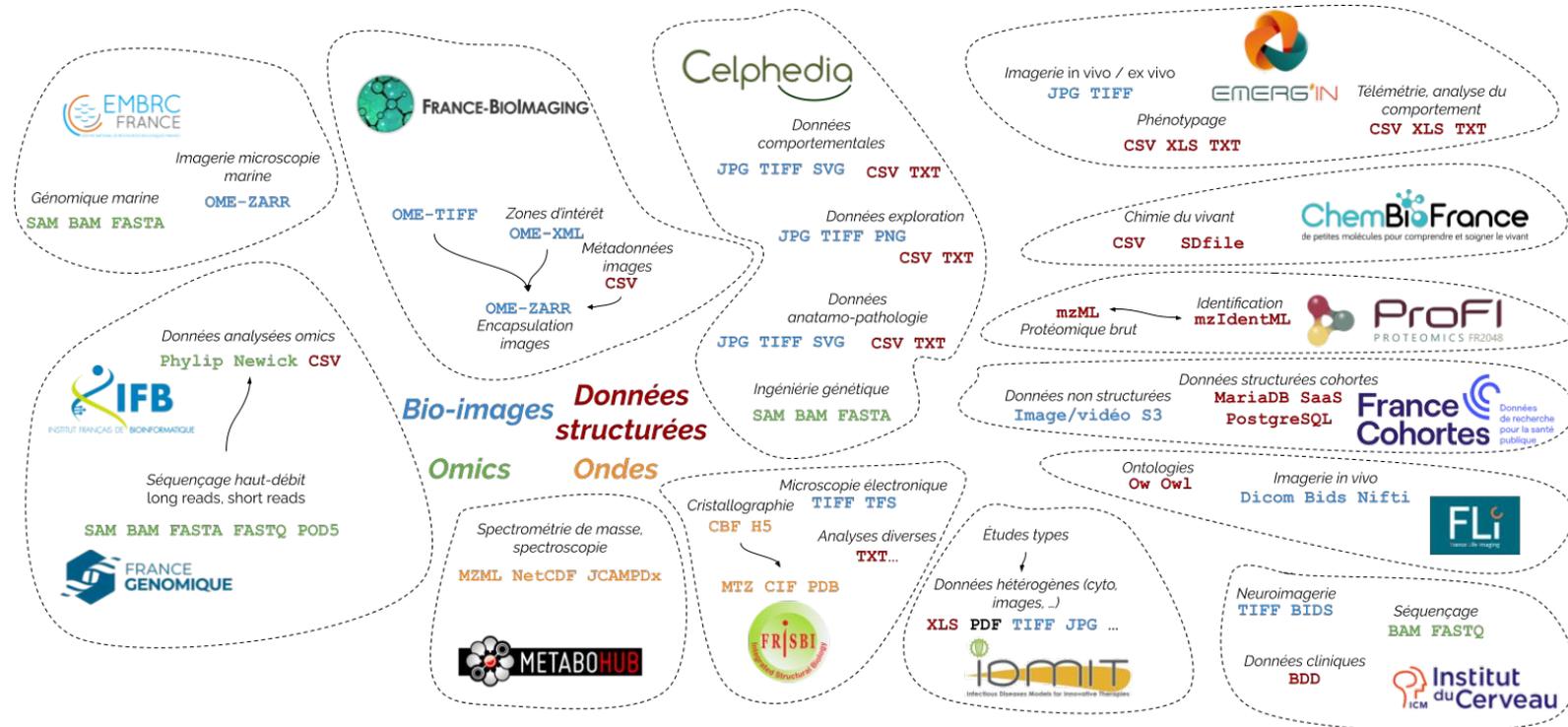




- Référentiels
 - Nécessite des expertises thématiques pour spécialiser les différents standards (e.g. protéomique, métabolomique, séquences...)
 - Toujours en construction pour certains domaines (e.g. NIRS, bioimagerie, ...)
- Infrastructures nationales
 - Produisent et analysent toutes des données
 - Partagent de nombreuses expertises
- Grande hétérogénéité des domaines d'application (les domaines et communautés des sciences de la vie)
- Beaucoup de données à forte valeur scientifique, un défi à FAIRiser

L'IFB au carrefour des données en science de la vie

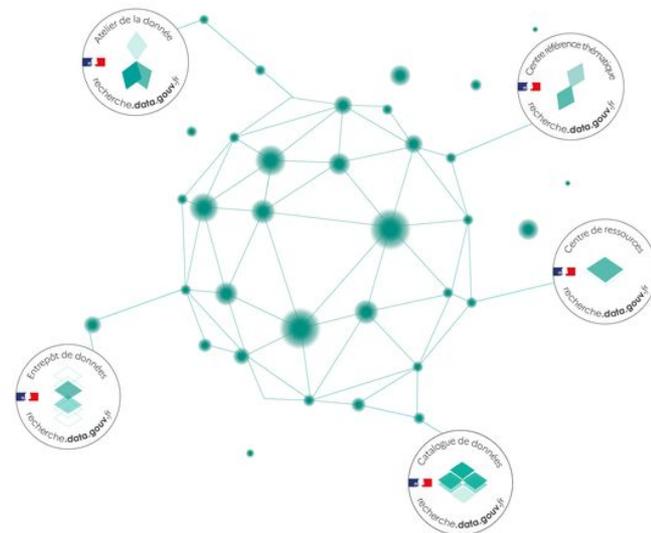
- Chaque INBS joue un rôle unique dans le paysage de la production, de l'analyse et du partage de données spécifiques.
- IFB est au carrefour:
 - Héberge tout type de données
 - Travaille sur l'intégration de données
 - Travaille sur les métadonnées et les différents standards de manière transversale (via ELIXIR)
 - Développe des outils à façon pour gérer les données et métadonnées le long de leur cycle de vie





II. IFB Centre de Référence Thématique en Biologie et Santé

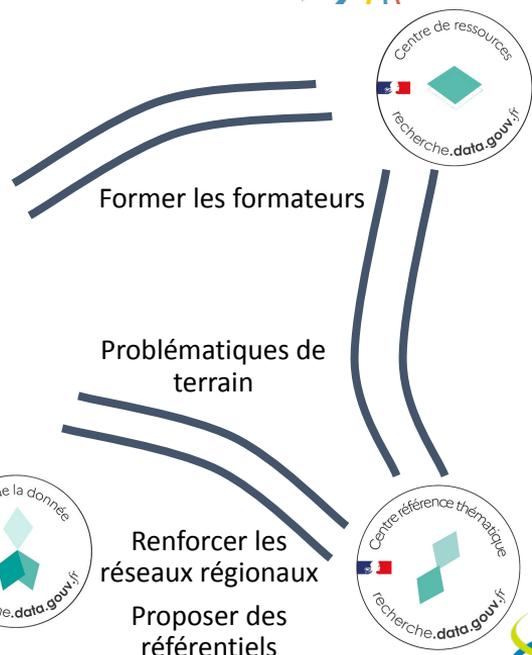
B) Être impliqué dans *Recherche Data Gouv*



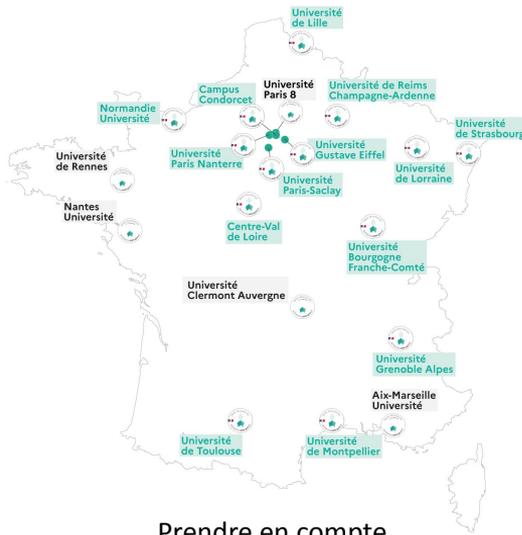


Des centres de ressources mutualisés

Faire vivre l'écosystème



Infrastructures nationales et européennes



Prendre en compte la disparité entre régions



❖ Travail sur l'entrepôt et le catalogue nationaux

- Groupe de travail régulier sur l'ergonomie et la fonctionnalité du catalogue de moissonnage
- Consultation sur les modèles de métadonnées et les interfaces thématiques de dépôt sur l'entrepôt

❖ Définition des référentiels

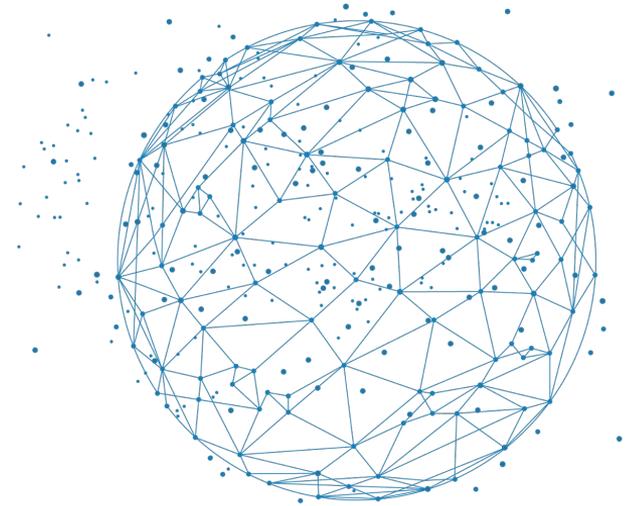
- Conception et rédaction de profils guides utilisables lors de la rédaction de PGDs (DMP OPIDoR)
- Rédaction de référentiels de FAIRisation de données
- Travail de fond sur les entrepôts thématiques

❖ Participation aux plénières

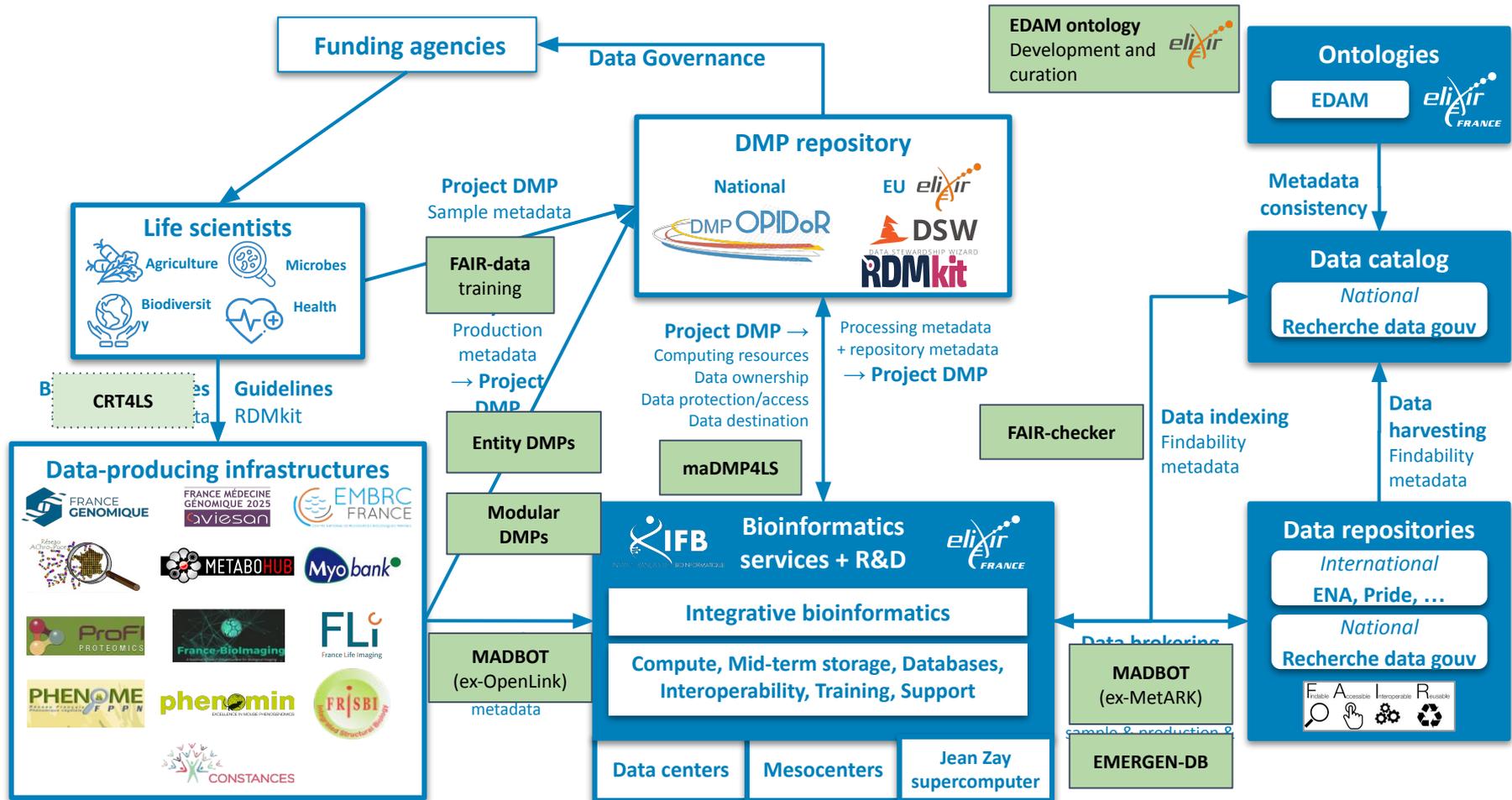
- 2 plénières Recherche Data Gouv par an
- Animation d'ateliers, participation active
- Sollicitation de partenaires des autres INBS

II. IFB Centre de Référence Thématique en Biologie et Santé

C) Des activités connexes au Centre de Référence Thématique



Orchestration des flux de données



- **DSW@IFB : PGD entité** pour des données multi-omiques
- **MADBOT** : un portail pour la science ouverte
- **FAIR-Checker** : un outil automatique d'évaluation FAIR
- De nombreuses **formations** au catalogue:
 - *Principes FAIR pour la gestion d'une plateforme de données IBISA*
 - *Principes FAIR dans un projet de bioinformatique*
 - *Principes FAIR pour la gestion des données en science de la vie*
 - *Science Ouvertes et PGDs*
- **RDMkit** : une boîte à outils pour les biosciences écrite par la communauté des biosciences