

L'institut français bioinformatique (IFB), centre de référence thématique pour la biologie-santé

Jacques van Helden
*pour le groupe de travail "Science ouverte et
interopérabilité" de l'IFB*



Enjeux scientifiques et technologiques

- Changement de paradigme : biologie → sciences de données
 - technologies productrices de données massives
 - approches holistiques “omiques”
 - ambition d’approche systémique du vivant
 - → **augmentation exponentielle** de la production
- Défis scientifiques et technologiques
 - Calcul intensif
 - Intégration des données hétérogènes (multi-modales)
 - Apprentissage automatique
 - Gestion des données, interopérabilité, FAIRisation
- Très forte demande des communautés biologie-santé pour un appui bioinfo
 - Infrastructure numérique (calcul, stockage, logiciel)
 - formation, montée en compétence
 - accompagnement de projets (expertise, collaborations)
- Spécificités de la biologie-santé
 - **Complexité des objets** représentés dans les données
 - **Biologie intégrative** (multi-omique + imagerie + phénotype)
 - **Protection des données de santé**

Enjeux sociétaux de la bioinformatique

- Ouverture des données
- Rationalisation des moyens numériques
- Enjeu écologique : sobriété numérique
- Pérennité de l’infrastructure nationale
- Mutualisation des compétences
- Montée en compétence sur les domaines d’innovation
- Engagement des acteurs

Enjeux sociétaux de la biologie

- Recherche fondamentale
- Agriculture
- Environnement, biodiversité
- Santé - One Health
- Biotechnologies



L'Institut Français de Bioinformatique (IFB)

- créé en 2013 dans le cadre de l'appel à propositions “**Infrastructures Nationales en Biologie et Santé (INBS)**” du Plan d'Investissement d'Avenir ;
- Inscrit sur la [feuille de route 2021 des infrastructures de recherche](#)
- assure un **support**, déploie des **services**, organise des **formations** et réalise des **développements innovants** pour les communautés des sciences du vivant ;
- fédère 21 plateformes-membres, 7 plateformes contributrices et 8 équipes de recherche associées.

<https://www.france-bioinformatique.fr>

<https://www.france-bioinformatique.fr/>

IFB
Institut Français de Bioinformatique

Contact Login

About IFB Services Training Innovation ELIXIR-FR Resource catalogue Jobs at IFB

Search here.

Welcome to the French Institute of Bioinformatics

The French Institute of Bioinformatics is setting up a collection of computer and data analysis needs to deal with the COVID-19 health crisis and is offering the support of its human and IT resources to back up the teams directly involved.

COVID-19

The IFB is the National Bioinformatics Infrastructure that provides support, deploys services, organizes training and carries out innovative developments for the life sciences communities.

More info

Compute & Storage
Clusters • Cloud
National Network of Computing Resources (NNCR)
MORE INFO

Support
Research projects • Communities • Orientation desk
MORE INFO

Training
Professional & initial training
MORE INFO

Innovation
Integrative bioinformatics
MORE INFO

ELIXIR-FR
French node of the European bioinformatics infrastructure network
MORE INFO

Resource catalogue
Tools, data, expertise, services in bioinformatics
MORE INFO

News
Newsletter • Social Media
MORE INFO

Roadmap
KPI
MORE INFO

CNRS Inserm INRAE CEA INVESTISSEMENT D'AVENIR

Missions et actions de la feuille de route 2022-2025 de l'IFB

M1 Infrastructure numérique

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

11 Po stockage
21K cœurs

M2 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
déployés

M7 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



M3 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

M4 Formations

- Écoles 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

M6 Innovation & prospective

- Bioinformatique intégrative (multi-omique, imagerie, phénomique, structure...)
- Intégration des données de santé
- Intelligence artificielle pour la biologie-santé

M5 Science ouverte & Interopérabilité

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

Une infra numérique distribuée

- 36 plateformes et équipes de recherche
- large couverture géographique
- couverture de tous les domaines d'expertise bioinformatique

22 plateformes-membres
7 PF contributrices
8 équipes associées
>400 experts (~200 FTE)

Des ressources mutualisées

- Financements d'infrastructure:
PIA1 RENABI-IFB (22.8M€), PIA3 MUDIS4LS (16.5M€)
- missions articulées autour de "task forces" mutualisées entre plateformes

Coordination: IFB-core (UAR 3601)

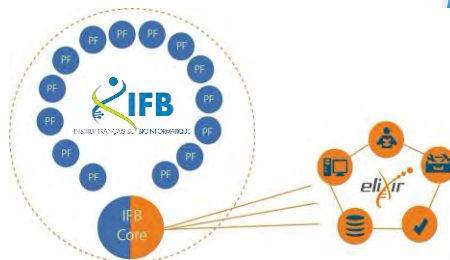
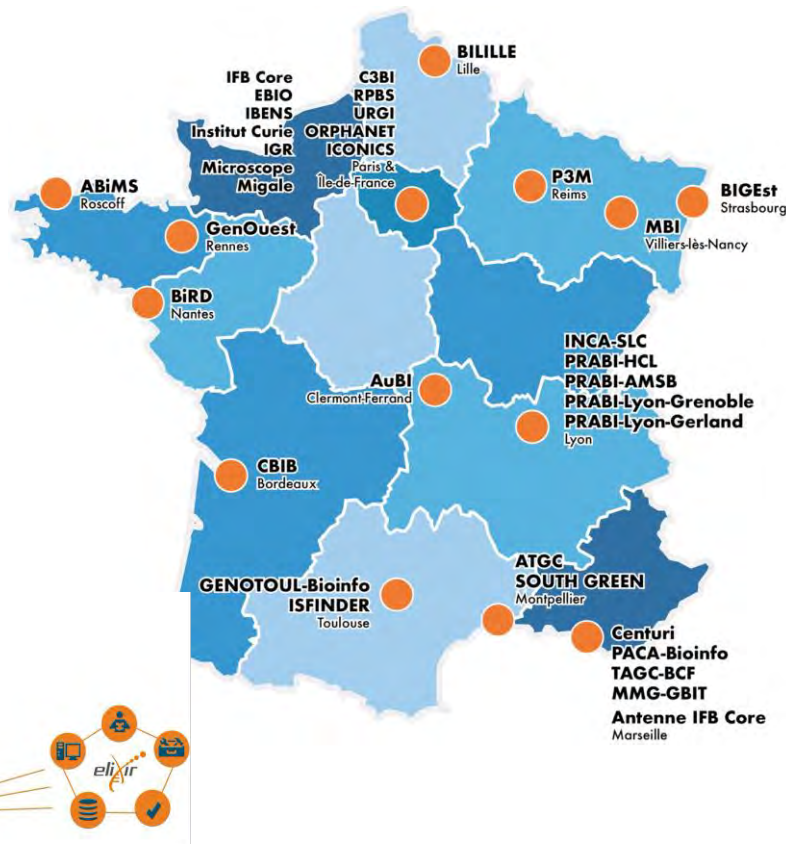
- multi-tutelles (CNRS, Inserm, INRAE, CEA)
- coordination et gestion des moyens mutualisés

Atteindre toutes les communautés

- biologie fondamentale, santé, agriculture, environnement

Une articulation forte avec l'Europe

- noeud français de l'ESFRI [ELIXIR](#) (ELIXIR-FR)
- point d'entrée pour des projets européens





Centres de Référence Thématiques (CRT)

- Définissent les normes internationales de **description des données** de leur domaine thématique
- Définissent les normes internationales de **diffusion des données** (ouverture, période d'embargo, accès restreint) de leur domaine thématique
- Définissent et diffusent les **bonnes pratiques** de collecte, documentation, traitement, et diffusion des données de leur domaine thématique
- Définissent la **liste des entrepôts de données de référence** de leur domaine thématique (nationaux et/ou internationaux) vers lesquels orienter les chercheurs pour le dépôt des données et que *Recherche Data Gouv* moissonnera
- Contribuent à la définition de l'**arborescence thématiques des données** de l'entrepôt *Recherche Data Gouv*
- Soutiennent l'**articulation entre les dispositifs thématiques spécialisés et Recherche Data Gouv**
- Soutiennent l'**inscription dans l'écosystème international**



**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**



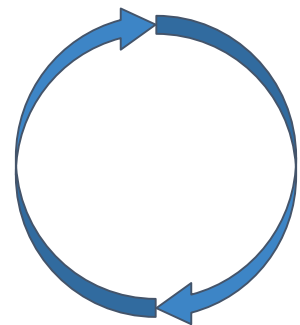
Ateliers de la donnée (politique de sites)

- Point d'entrée en proximité des chercheurs sur toute nature de besoin relatif à la donnée
- Regroupent en leur sein une palette de compétences qui constitue une alliance interprofessionnelle
- Regroupent des ressources de différents établissements
- Apportent un premier niveau d'expertise et de services
- S'appuient sur l'expertise des services juridiques, partenariales, SI de proximité
- S'appuient sur des expertises scientifiques nationales (centres de référence thématiques)
- Animés en réseau pour mutualiser leurs pratiques, formations...



Centres de référence thématiques (ONR et/ou infrastructures de recherche)

- Conçoivent et portent les référentiels de pratiques thématiques/disciplinaires
- Conçoivent et portent le catalogue de référence des entrepôts thématiques/disciplinaires
- Contribuent à l'arborescence thématiques/disciplinaires des données de l'entrepôt
- Soutiennent l'articulation de la plateforme avec les dispositifs thématiques spécialisés
- Soutiennent l'inscription dans l'écosystème international



Co-construction des référentiels et des catalogues

- Réponses concrètes aux problématiques de terrain, remontées des ateliers
- Les ateliers s'appuient sur les référentiels et catalogues thématiques

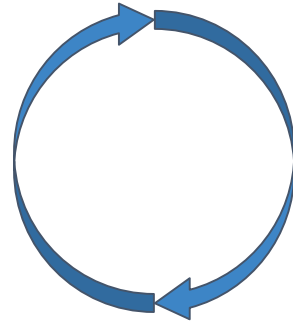
Les plateformes et équipes IFB, charnières entre CRT et ateliers de la donnée

- nombreux personnels communs entre CRT et ateliers
- /!\ Grands besoins RH

Ateliers de la donnée



Centres de recherche thématiques



- **Production de données** (workflows)
- **Structuration des données** (bases de données)
- **Analyse de données**
 - accompagnement de biologistes
- **Stockage de données**
 - **stockage à chaud** pendant la durée des projets
 - **préservation**: accompagnement des chercheurs vers les dépôts
- **Data brokering** (au sens noble du terme) :
 - Facilitation de la soumission des données
 - amélioration de la qualité des données et des métadonnées (FAIR-Checker)
 - soumission aux entrepôts internationaux
- **Formation à l'utilisation des données**
 - bioinfo "classique" (BD, séquences)
 - NGS (EBAii)
 - Analyse statistique (surtout R)
 - Bioinformatique intégrative (DUBii)
 - **FAIR-data**: FAIRisation des données
 - **FAIR-bioinfo**: FAIRisation du logiciel
- **Moissonnage des données**
 - Catalogue des ressources bioinformatiques IFB
 - Contribution au catalogue national des données de **recherche data gov**

Mutualisation des ressources entre INBS

- Coordination : Jean-François Dufayard et Christophe Bruley
- Partage des moyens numériques IFB avec infra nationales de biologie santé (INBS) productrices
- **Plans de gestion de données modulaires** pour l'intégration de données hétérogènes (Paulette Lieby)

Interopérabilité

- Coordination : Marie-Dominique Desvignes et Alban Gaignard
- **Développement de ressources d'interopérabilité**
 - FAIR-checker
 - Plateforme *Interoperability* du réseau ELIXIR
 - Bioschemas (membre du copil)

Data brokering : accompagnement des communautés

- Coordination: Hélène Chiapello & Thomas Denecker
- Objectifs
 - Faciliter le flux des données : de la création à la soumission aux dépôts de référence
 - Améliorer la qualité, l'interopérabilité et l'exploration des métadonnées
- Outil: omicsBroker

Formations FAIR-bioinfo et FAIR-data

- Coordination: Hélène Chiapello & Fred de Lamotte
- Deux formations
 - FAIR-data : gestion FAIR des données de la recherche
 - FAIR-bioinfo: principes FAIR et reproductibilité d'un workflow bioinformatique
- Approche "train-the-trainer"
 - Premières éditions orientées vers bioinformaticiens IFB
 - But : s'approprier la formation pour la redonner localement

Machine-actionable DMPs for Life Sciences (maDMP4LS)

- Coordination : Olivier Collin (PF IFB GenOuest)
- Appel ANR flash Science ouverte (2019)
- Collaboration IFB - INIST
- **Plans de Gestion des Données automatisables**
 - Interface programmatique, Modularité
 - PGD ↔ ressources numériques allouées
 - Mise à jour au fil du projet
 - Outil de gouvernance des données

OpenLink : tableau de bord de gestion des données

- Coordination : Julien Seiler
- Appel ANR flash Science ouverte (2019)
- Interfaçage avec de nombreux outils (LabGuru, OMERO, Galaxy, stockage SSH, etc.)
- Intégration de solutions de brokering

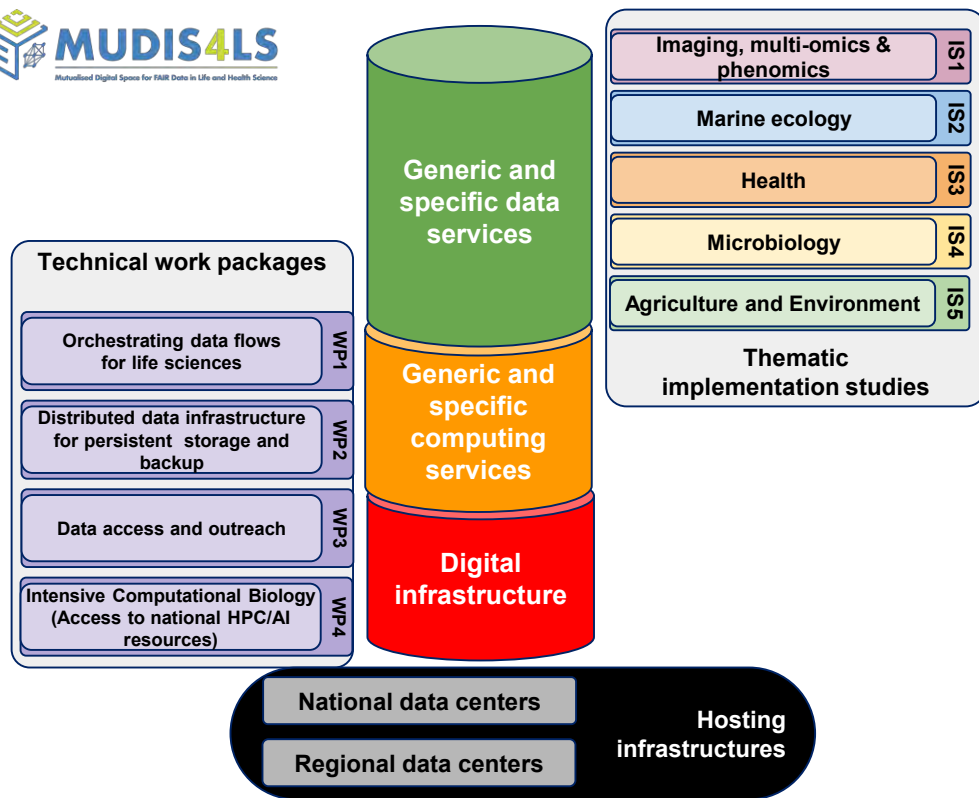


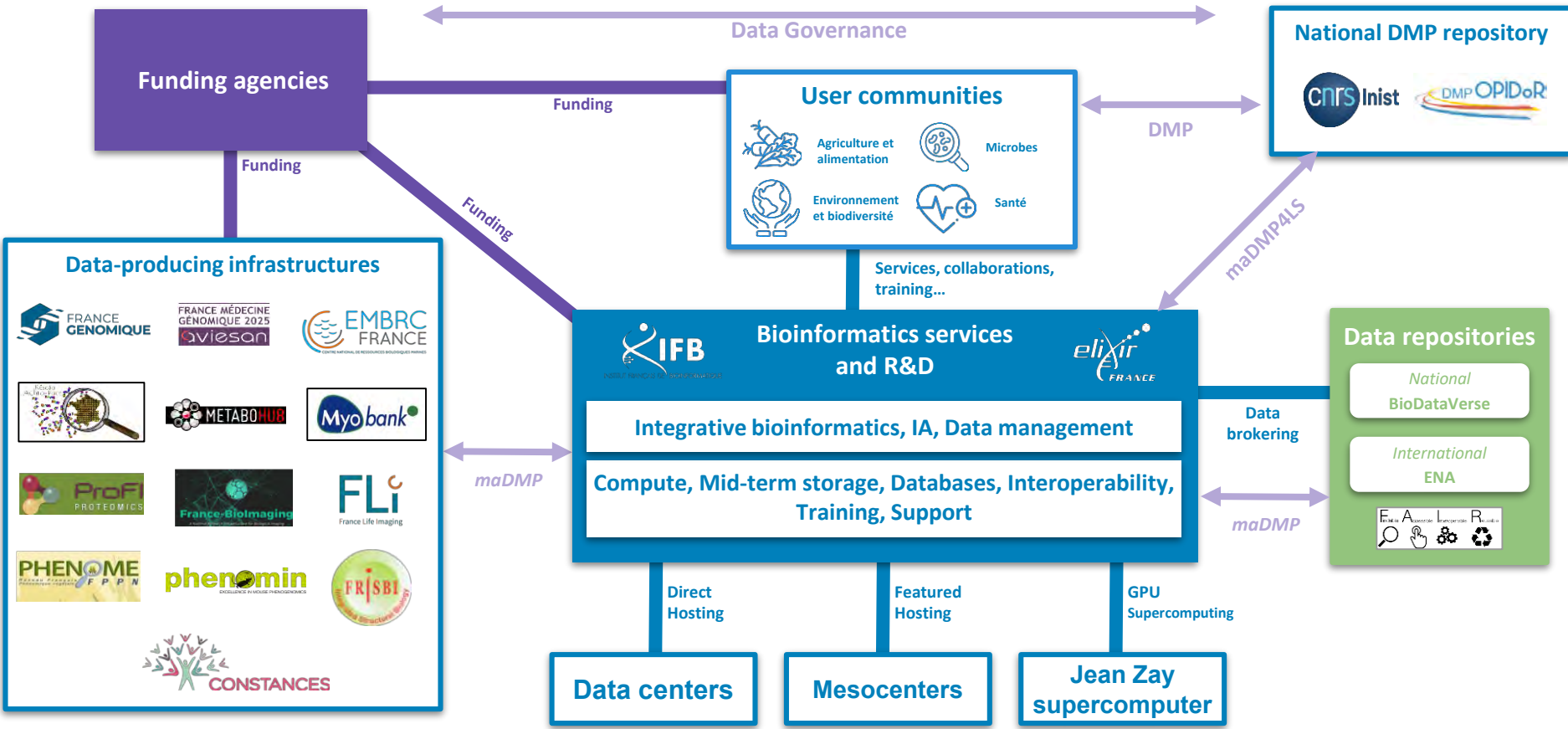
Main axes of the MUDIS4LS project

- **Enlarging the national coverage of the NNCR**
- Anchoring in national and regional **data centers**
- Subcontracting with some regional **mesocenters**
- **Data securing** during the projects (mid-term storage)
- **Orchestrating data flows** along the whole life cycle
 - maDMP-based management
 - data brokering
 - linking to electronic lab book
- Data **FAIRification**
- **5 thematic implementation studies** ensuring relevance to address user needs

Organisation

- 39 partner teams
- 14 partner organisms
- 4 national + 7 regional data centers, 4 partner mesocenters
- 170 persons, 3000 PM = 2.5 person•centuries
- Requested funding: 20M€ / **allocated : 16,5 M€**







Politique générale

- Ouvrir autant que possible, fermer autant que nécessaire
- Chaque plateforme est responsable de sa mise en conformité avec le RGPD, en lien avec la politique de ses tutelles.
- Action à mener (feuille de route IFB 2022-2025) : mutualiser les expériences et solutions.

Données personnelles

- **Métadonnées de santé** associées aux données de patients (images, séquences génomiques, ...)
- Particularité des **séquences génomiques humaines**
 - **Identifiantes**
 - **Prédictives** pour certains risques de santé
 - Personnelles mais aussi **familiales** (un individu partage 50% du génome avec ses apparentés du 1er degré)

Projets concernés

- **EMERGEN**: réseau national de surveillance génomique COVID-19 (CNRS, Santé Publique France, Inserm ANRS-MIE, DSI Inserm et CNR).
- **ABRomics**: Plateforme multi-omique antibiorésistance, (Inserm, CNRS, Institut Pasteur, CNR et des labos de microbiologie).
- **MUDIS4LS**: implementation study (IS3) consacrée aux données de santé et de génomique personnelle (CNRS, Inserm, CINES et CAD du PFMG2025).
- **Genomic Data Infrastructure (GDI)**. Projet européen piloté par ELIXIR visant à partager les informations de génomique pour la santé en préservant la sécurité de la donnée génomique. Partenaires français: CAD du PFMG2025 + IFB.

Défi technologique : déployer des **espaces certifiés pour l'hébergement de données de santé (HDS)** accessibles aux chercheurs.



ELIXIR: a distributed infrastructure for Life Science Information

<https://www.elixir-europe.org/>

23 noeuds nationaux
1 hub coordinateur
200 instituts



Année d'entrée sur la roadmap ESFRI : 2006 - reached its implementation phase by 2015.

Statut juridique actuel de l'infrastructure européenne/internationale : ELIXIR Consortium Agreement (ECA)

23 countries in ELIXIR :

 EMBL (June 2013)	 UK (Sept 2013)	 Sweden (Sept 2013)	 Switzerland (Oct 2013)
 Czech Republic (Nov 2013)	 Estonia (Dec 2013)	 Norway (Jan 2014)	 Netherlands (Feb 2014)
 Denmark (Mar 2014)	 Israel (May 2014)	 Portugal (July 2014)	 Finland (Sept 2014)
 France (Oct 2015 as Provisional Member)	 Belgium (Nov 2015)	 Italy (Jan 2016)	 Slovenia (Feb 2016)
 Luxembourg (July 2016)	 Ireland (July 2016)	 Germany (Aug 2016)	 Hungary (Jan 2017)
 Spain (Oct 2017)	 Greece (Feb 2019)	 Cyprus	



Activités d'ELIXIR pertinentes pour la science ouverte

- **Service Delivery Plan** : labellisation européenne des services (outils logiciels, bases de données, formations) développés par les noeuds nationaux
- **Core Data Resources** : bases de données indispensables à la recherche en biologie.
- **Data Brokering** (au sens noble) : rôle des noeuds nationaux pour faciliter la soumission des données nationales.
- **EDAM**: ontologie des concepts bioinformatiques (SDP co-porté par la France et la Norvège)
- **Bio.tools** : catalogue des outils et bases de données bioinformatiques
- **Bioschemas** : métadonnées minimales pour la découvrabilité des ressources biologiques
- **TeSS** : réutilisation des supports de formation.
- **ELIXIR-CONVERGE** : standardisation de la gestion des données (voir diapo suivante).

ELIXIR unites Europe's leading life science organisations in managing and safeguarding the increasing volume of data being generated by publicly funded research. It coordinates, integrates and sustains bioinformatics resources across its member states and enables users in academia and industry to access services that are vital for their research. See [About us](#).



Services

ELIXIR services make it easier to discover, store, and analyse life science data.



Platforms

ELIXIR's activities are divided into five areas called 'Platforms'.



EU Projects

ELIXIR both collaborates in and coordinates EU projects.



Communities

Communities develop services that are specialised for particular life science domains.



ELIXIR-CONVERGE

- Objectif: **standardiser la gestion des données** des sciences du vivant au niveau de l'Europe
- Communauté internationale d'experts en gestion de la donnée
- Développement du [Research Data Management Toolkit for Life Sciences](#) ([RDMkit](#))
- Financement européen
 - 2019: 5M€ (WP1 - 6)
 - 2020: complément de 5M€ pour mettre le réseau au service du partage des données COVID-19

Work packages

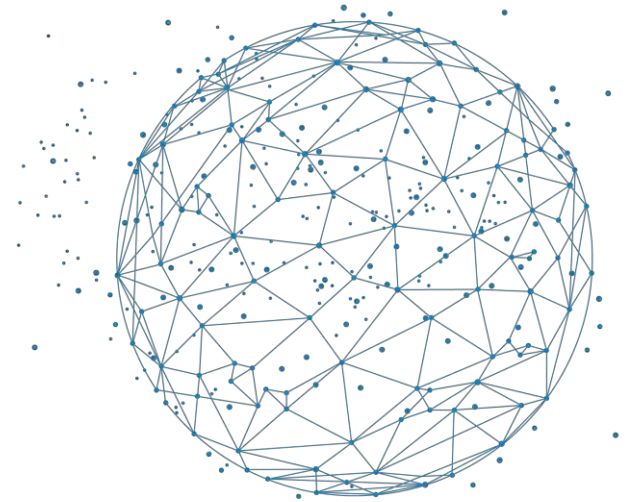
- **WP1: Expert network**
- **WP2: Training and Capacity Building**
- **WP3: Common Data Management Toolkit**
- **WP4: Communications, Industry, International, Impact and Sustainability**
- **WP5: Demonstrator Projects**
- **WP6: Project Management and Scientific Coordination**
- **WP7: Federated European Genome-phenome Archives for transnational access of COVID-19 host data**
- **WP8: ELIXIR-CONVERGE European COVID-19 Data Platform**
- **WP9: Mobilisation of SARS-CoV-2 variant surveillance data tracking services and tools**





- Défis
 - Mise en oeuvre des standards
 - Concilier ouverture et protection des données sensibles (santé, valeur économique)
 - Toucher l'ensemble des communautés biologie-santé
 - Intégration de données hétérogènes de la biologie
 - Mutualiser les réalisations entre infrastructures numériques
 - Elargir aux données transversales : biologie + physiques + humaines
- Points d'appui
 - Les initiatives financées par les PIA (RENABI, MUDIS4LS)
 - L'engagement fort de la communauté bioinformatique (plateformes, équipes de recherche, enseignants)
 - La collaboration active entre Infrastructures Nationales de Biologie-Santé (club des INBS, IBISA)
 - Les réseaux nationaux mis en place par le MESR
 - Comité de soutien aux Infrastructures Numériques (CoSIN)
 - Comité de Soins à la Donnée (CSD)
 - Comité Science Ouverte (COSO)
 - La mutualisation des réalisations avec pairs européens (ESFRI ELIXIR)

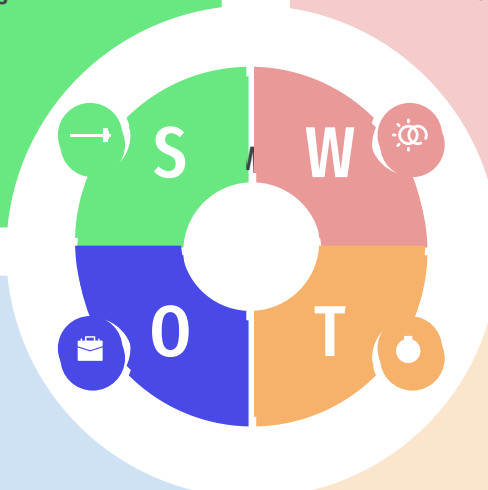
Matériel supplémentaire





Forces

- **Dynamique collective des plateformes IFB**
- Proximité des équipes de recherche
- **Large spectre de compétences techniques**
- **Point d'entrée pour appuyer les projets dans tous les domaines thématiques biologie-santé**
- Catalyse de **projets inter-INBS**
- **Connecteur à l'international**
- **Formation**
- **Capacité à appuyer des politiques publiques de surveillance sanitaire**



Faiblesses

- **Limitation des postes statutaires** pour affronter la **croissance de la demande** et des attentes
- Soutien administratif pour projets complexes
- **Lacune à combler : Absence (actuelle) de ressource numérique HDS**

Opportunités

- (France: HDS; projets EU 1+MG, B1MG, GDI...)
- **Transposer au niveau EU** la capacité structurante et l'impact
- IFB centre de référence des données biologie-santé
- Certification organisme de formation
- Interaction avec **INBS environnement**, biodiversité

Menaces

- **Limitation du vivier des ressources humaines recrutables**
- Manque d'attractivité des salaires → fuite des compétences
- Epuisement des ressources humaines du fait de la sur-sollicitation
- Manque de visibilité sur la pérennisation post-2024

- **Bases de données** : collecte, structuration, analyse, diffusion de données produites par différentes technologies, notamment à haut débit
- **Bases de connaissances**: curation par les experts
 - Couverture de **tous les domaines de biologie-santé**
- Quelques exemples
 - **OrphaData** : maladies orphelines (réseau OrphaNet)
 - **IMGT**, the international ImMunoGeneTics Information system
 - **MICROSCOPE** : génomes et voies métaboliques bactériens
 - **GenomeHubs**: crop-specific genomic datasets with focus on tropical and Mediterranean plants
 - **Ocean Gene Atlas** : sequences (ENA), environmental context (PANGAEA), gene abundances estimates
 - **GENOMICUS**: comparative genome data and ancestral genome reconstruction
 - **Norine**: nonribosomal peptides (NRPs)
- Depuis 2021: sollicitation de l'IFB pour développer des bases de données de **projets nationaux**
 - **Emergen-DB**: surveillance génomique et recherche COVID-19
 - **ABRomics-PF**: plateforme multi-omiques pour la surveillance et recherche en antibiorésistance

Labellisation internationale des bases de connaissances

- **ELIXIR Services Delivery Plan (SDP)**
 - L'IFB a déclaré 29 services au SDP ELIXIR
 - 16 de ces services reposent sur des bases de données ou de connaissances
- **ELIXIR Core Data Resources**
 - La plateforme "Data" de l'ESFRI ELIXIR labellise des "Core Data Resources", définies comme des ressources internationales indispensables au bon fonctionnement de la recherche en biologie
 - La BD **Orphadata a été déclarée core data resource.**
 - **IMGT**: démarches de labellisation en cours.



The screenshot shows the ELIXIR website's news section. The header includes the ELIXIR logo and navigation links: ABOUT US, SERVICES, HOW WE WORK, EVENTS, NEWS, and INTRANET. A search bar is located in the top right corner. The main content area features a news article titled "Orphadata selected as ELIXIR Core Data Resource". The article text states: "Orphadata, a platform to aggregate rare disease datasets from Orphanet database, has joined the list of ELIXIR Core Data Resources, a set of data resources of fundamental importance for life science research. The ELIXIR Core Data Resources serve as a mark of the highest quality in infrastructure service provision and play a critical role in the long-term preservation of life science data. Orphadata has been added to the list, following a detailed review by independent panel of reviewers and their decision to adopt open access licence." To the right of the text is the ELIXIR Core Data Resource logo. Below the main text, there is a smaller paragraph: "Orphadata gives access to aggregated scientific data related to rare diseases from the Orphanet knowledge base, as well as to ORDO - the Orphanet rare disease ontology and to HOOB - the HPO-ORDO ontological module. It is an extension of Orphanet, serving as a dedicated download platform for the datasets collected by Orphanet. It launched in 2011 to meet the demands from researchers and clinicians who want to reuse and combine the rare disease data provided by Orphanet."



Activités nationales

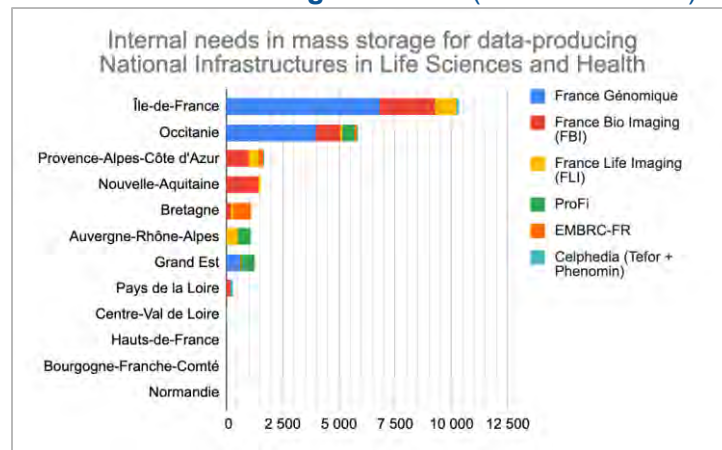
- **PGD interopérables (machine-actionable DMP)**
 - **maDMP4LS**: machine-actionable Data Management Plans for Life Sciences
 - ANR appel flash science ouverte (95k€, 2020-2021)
 - Collaboration IFB - INIST
 - plans de données automatisables \leftrightarrow allocation de ressources numériques NNCR
- **PGD modulaires** (collaboration multi-INBS + INIST)
 - Projets combinant des données hétérogènes (multi-omiques, d'imagerie, phénomiques)
 - Modularité institutionnelle
- **MUDIS4LS** : Mutualized Digital Spaces for FAIR Life Sciences
 - orchestration des flux de données tout au long du cycle de vie
 - généralisation des plans de données automatisables
 - PIA3 ESR/Equipex+ (16.5M€, 2021-2029)



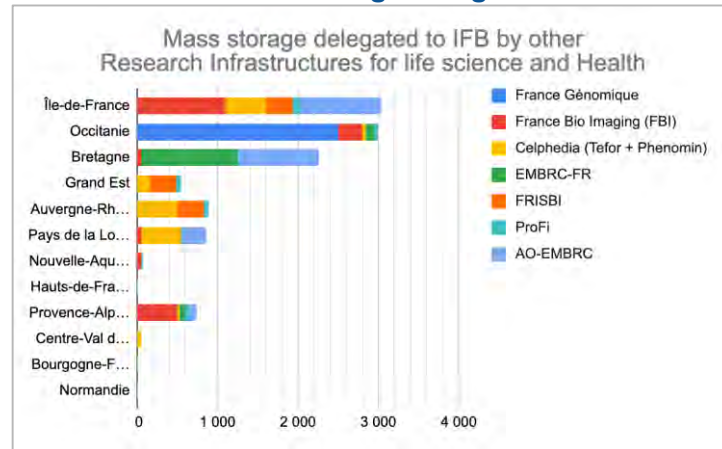
Données à haut débit

- La biologie actuelle repose crucialement sur les technologies de production de données massives
- L'IFB **ne produit pas de données primaires**, mais
 - assure le **stockage à chaud et le traitement** des données massives produites par différentes technologies (génomique, protéomique, métabolomique, structure, imagerie, ...)
 - **produit des données secondaires** à partir de ces données primaires (volume équivalent)
 - **accompagne les chercheurs** pour le dépôt de ces données (primaires et secondaires) dans les entrepôts internationaux
- Ordres de grandeur (figure) : **cartographie (2020) des besoins de stockage des INBS (en To)**
 - en interne
 - délégué à l'IFB
- L'IFB produit également du **code**, qui relève de la donnée
 - Par rapport aux données massives, négligeable en volume – mais pas en valeur !

Besoins de stockage internes (estimation 2020)



Besoins de stockage délégués à l'IFB



1. **Mutualisation des ressources des plateformes** française de bioinformatique :
 - développement des infras numériques et des services nationaux
 - articulation au niveau européen et international
2. **Fourniture et accès à des ressources bioinformatiques** pour les sciences de la vie et de la santé.
3. **Appui aux communautés** (académiques et privées) **en biologie, santé, agronomie et environnement**: expertise, services, et développements méthodologiques.
4. **Développement d'une vision stratégique commune du domaine** pour maintenir la France au plus haut niveau d'expertise dans les différents domaines de la bioinformatique.
5. **Renforcement de la représentation internationale de la communauté bioinformatique française**, en particulier au sein du réseau européen ELIXIR.

<https://www.france-bioinformatique.fr/>

Welcome to the French Institute of Bioinformatics

The French Institute of Bioinformatics is setting up a collection of computer and data analysis needs to deal with the COVID-19 health crisis and is offering the support of its human and IT resources to back up the teams directly involved.

The IFB is the National Bioinformatics Infrastructure that provides support, deploys services, organizes training and carries out innovative developments for the life sciences communities.

COVID-19

Compute & Storage
Clusters • Cloud
MORE INFO

Support
Research projects • Communities • Orientation desk
MORE INFO

Training
Professional & initial training
MORE INFO

Innovation
Integrative bioinformatics
MORE INFO

ELIXIR-FR
French node of the European bioinformatics infrastructure network
MORE INFO

Resource catalogue
Tools, data, expertise, services in bioinformatics
MORE INFO

News
Newsletter • Social Media
MORE INFO

Roadmap
KPI
MORE INFO

Partners: CNRS, Inserm, INRAE, cea, Institut Français de Bioinformatique

National network of computing resources (NNCR)

- Depuis 2017, l'IFB a déployé un réseau national de ressources numériques mutualisées.
- 2 infrastructures complémentaires: **Cloud** et **Cluster**
- Services répartis sur 12 sites
 - 2 serveurs nationaux (IFB-core-cluster & IFB-core-cloud)
 - 10 serveurs régionaux
- Mutualisation des efforts humains : **NNCR task force**
 - Chaque plateforme partage des ingénieurs (10% à 50% ETP)
- Bénéfices partagés
 - Eviter la redondance des efforts
 - Gestion collective des problèmes
 - Partage d'expérience et des solutions
 - Procédures robustes et transposables
 - Forte émulation des ingénieurs
 - Rompres la solitude de l'ingénieur info dans son labo bio



8 cloud sites
 > 8,000 vCPUs
 > 1.5Pb storage
 > 36 TB RAM

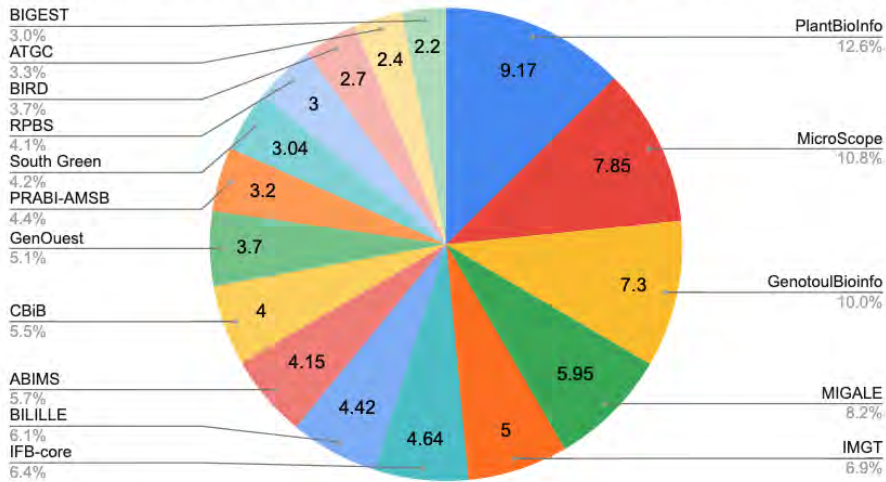
6 clusters
 > 16,000 vCPUs
 > 16Pb storage
 > 92 TB RAM

Plateforme	Localisation	Calcul (#CPU HT*)	Stockage (#TB)	RAM (#GB)
ifb-ore-cloud	Lyon (CC-IN2P3)	3936	408	20 408
AuBI	Clermont-Ferrand	200	20	800
GenOuest	Rennes	600	350	2600
PRABI	Lyon	696	144	3300
BiRD	Nantes	860	150	2 500
BIGEst	Strasbourg	1024	500	4 000
BILILLE	Lille	192	0	768
CBP-PSMN	Lyon	768	24	1920
Total fédération de clouds		8276	1596	36296

Plateforme	Location	Compute (#CPU HT*)	Storage (#TB)	RAM (#GB)	RAM/cor e (#GB)	GPU (#Cards)
ifb-core-cluster	Orsay (IDRIS)	4300	2400	20008	4,65	9
ABiMS	Roscoff	2600	2500	10600	5,50	2
GENOTOUL	Toulouse	6224	8000	36500	5,56	1
GenOuest	Rennes	1866	2300	11616	6,23	7
BiRD	Nantes	902	600	6800	6,79	-
MIGALE	Jouy en Josas	1016	350	7000	6,89	-
Total IFB clusters		16908	16550	92524	35,64	19

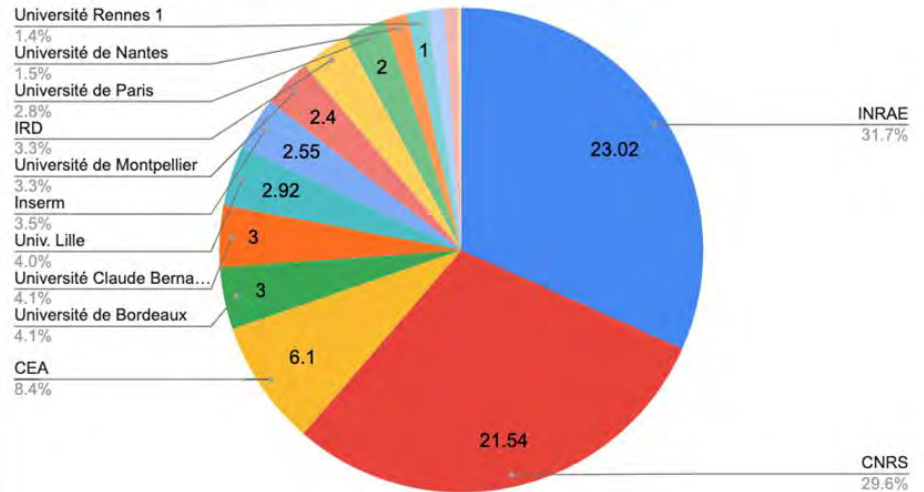


Personnel statutaire par plateforme (toutes tutelles confondues)



Env. 80 ETP

Personnel statutaire de PFs IFB - ETP par tutelle

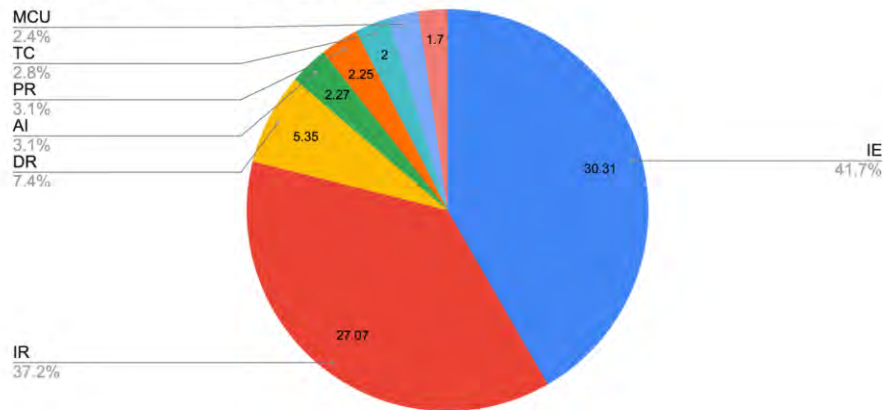




Quels sont les **métiers** présents dans l'infrastructure et impliqués dans les travaux liés aux données ? (chercheurs, enseignants chercheurs, documentalistes, informaticiens, Ingénieurs de recherche, etc.).

- **Ingénieurs** (informaticiens et bioinformaticiens)
- Chercheurs
- Enseignants-chercheurs

Personnel statutaire – ETP par corps



Quelles sont les **fonctions** mises en place au sein de l'IR, dévolues à la prise en charge des données ?

- **CDO (chief data officer)**
 - Poste de stratégie, définit la politique des données
 - Frédéric de Lamotte
 - Animation du **réseau des data stewards**
- **DPD (délégué à la protection des données)**
 - IFB-core: DPD/DPO du CNRS, tutelle gestionnaire
 - Plateformes : DPD/DPO de leurs tutelles
- **Data stewardship center**
 - instance de régulation et de consultation sur les aspects pratiques liés aux données
 - En développement: **centre de référence thématique biologie-santé**

Fonctions et instances à développer

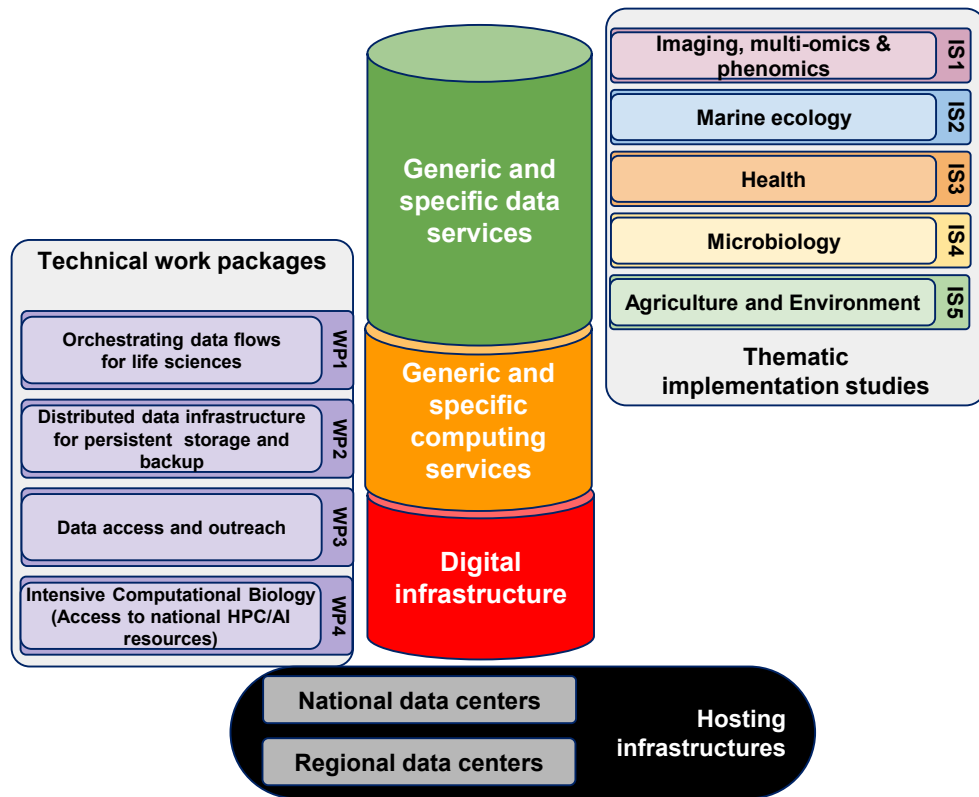
- **Data steward** : un par plateforme
- **Correspondants protection des données personnelles** pour chaque PF concernée
- **Comité des données**
 - en constitution (FdR IFB 2022-2025)
 - associé au comité éthique de l'IFB

Main axes of the MUDIS4LS project

- **Enlarging the national coverage of the NNCR**
- Anchoring in national and regional **data centers**
- Subcontracting with some regional **mesocenters**
- **Data securing** during the projects (mid-term storage)
- **Orchestrating data flows** along the whole life cycle
 - maDMP-based management
 - data brokering
 - linking to electronic lab book
- Data **FAIRification**
- **5 thematic implementation studies** ensuring relevance to address user needs

Organisation

- 39 partner teams
- 14 partner organisms
- 4 national + 7 regional data centers, 4 partner mesocenters
- 170 persons, 3000 PM = 2.5 person•centuries
- Requested funding: 20M€ / **allocated : 16,5 M€**





MUDIS4LS

Mutualized Digital Spaces for FAIR Data in Life and Health Sciences

PIA3 ESR/Equipex+

- **Mutualized Digital Spaces for FAIR Life Sciences**
- Resp. sci.: Jacques van Helden
Resp. Tech: Gildas Le Corguillé & Julien Seiler
- **Budget total : 16.5M€**

ALADIN :

- **Apprentissage actif pour accélérer le développement de biocatalyseurs pour la biotechnologie industrielle**
- Resp. sci et tech: Fayza DABOUSSI (INRAE)
- Implication IFB: hébergement du serveur Galaxy-SynBioCAD sur le NNCR
- **Budget total: 3.9M€**

AO-EMBRC :

- Resp. : Bernard Kloareg (Prof Sorbonne Univ)
- Contribution IFB: FAIRisation et Intégration des données, gestion et analyse des données
- **Budget total: 6.4M€**

PEPR santé numérique

- **Resp.** Hugues Berry (INRIA) et Franck Lethimonnier (Inserm)

DIGPHAT

- “Multi-scale and longitudinal data modeling in pharmacology: towards digital pharmacological twins”
- Resp. : Christophe Battail, CEA, Jean-Baptiste Woillard (MCU/PH)
- Contribution IFB : FAIRisation des données
- En préparation (budget à définir)

EMERGEN

- **Plateforme pour la surveillance et la recherche des infections virales**
- WP3: EMERGEN-DB; resp.: Jacques van Helden
- Financement : Ministère
- **Budget IFB : 1,3M€**



- **Plateforme multi-omique pour la surveillance et la recherche en antibiorésistance**
- Resp.: Claudine Médigue (IFB) et Philippe Glaser (Pasteur)
- Financement : PPR antibiorésistance
- **Budget total (IFB) : 2M€**

PEPR Exploratoire ATLASea

- Resp. Patrick Wincker (CEA) & Hugues Roest Crolius (CNRS)
- Contribution IFB : Gestion et FAIRisation des données
- **Budget total (demandé): 56 M€**

PEPR Agriculture & Numérique

- “e-Infrastructure to boost the use of diversified biological resources”
- Contribution IFB : gestion et FAIRisation des données
- En préparation (budget à définir)

PEPR Produits Biosourcés et Nouvelles Biotechnologies

- “An operating portal for the production of biosourced products”
- Contribution IFB : déploiement logiciel (galaxy-BioProd)
- En préparation (budget à définir)