

L'Institut Français de
Bioinformatique :
Centre de référence
thématique en **Science**
Ouverte pour la **biologie** et la
santé

Marie-Dominique Devignes, Equipe CAPSID, LORIA
Co-responsable Mission Science Ouverte à l'IFB



INSTITUT FRANÇAIS DE BIOINFORMATIQUE



INRAE



Inserm



A. Exposé : IFB et Centre de Ressources Thématique Biologie-Santé

Introduction: IFB et Recherche Data Gouv

I. *Présentation de l'IFB*

II. *Exemples de ressources pour la Science Ouverte à l'IFB*

Conclusion: Perspectives du CRT Biologie-Santé

B. Mise en pratique

I. Utiliser FAIR-Checker

II. Explorer la bio-ontologie EDAM

RECHERCHE DATA GOUV

**ATELIERS DE LA DONNÉE**

Experts de la donnée
en charge de
l'accompagnement
des chercheurs

**ENTREPÔT**

Interface Web
de dépôt de ses données
par le chercheur
+ espace de modération

**CATALOGUE**

Repérage et moissonnage
des données
des entrepôts externes

**CENTRES DE RÉFÉRENCE THÉMATIQUES**

Experts disciplinaires de la donnée

**CENTRES DE RESSOURCES
RATTACHÉS À RECHERCHE DATA GOUV**

- Printemps 2022 : Candidature IFB pour des Ateliers de la Donnée
- Réponse du Ministère : Invitation à devenir Centre de Ressources Thématiques pour la Science Ouverte en Biologie et Santé
- Depuis:
 - Elaboration d'une feuille de route
 - Concertation avec les autres infrastructures nationales en Biologie Santé (INBS)
 - Profils de postes à recruter

I. Présentation de l'IFB

Institut Français de Bioinformatique



- **Infrastructure Nationale de soutien à la recherche**

- créée en 2014 dans le cadre du Programme d'Investissement d'Avenir (PIA2)
- Refondée en 2017 (PIA3)
- prolongée dans le PIA4 (2020)

- **Missions pour les communautés des sciences du vivant**

- assurer un soutien aux équipes de recherche
- déployer des services
- organiser des formations
- réaliser des développements innovants

IFB INSTITUT FRANÇAIS DE BIOINFORMATIQUE

Contact

À propos Services Formation Innovation ELIXIR-FR Ressources

Search here.

L'IFB recrute

Bienvenue à l'Institut Français de Bioinformatique

L'IFB est l'Infrastructure nationale de bioinformatique qui assure un support, déploie des services, organise des formations et réalise des développements innovants pour les communautés des sciences du vivant.

EN SAVOIR +

COVID-19

Calcul & Stockage
Clusters • Cloud
National Network of Computing Resources (NNCR)
EN SAVOIR +

Accompagnement
Projets de recherche • Communautés • Guichet d'orientation
EN SAVOIR +

Formation
Formation professionnelle & initiale
EN SAVOIR +

Innovation
Bioinformatique intégrative
EN SAVOIR +

ELIXIR-FR
Noeud français du réseau européen d'infrastructures bioinformatiques
EN SAVOIR +

Catalogue de ressources
Outils, données, expertises, services en bioinformatique
EN SAVOIR +

Actualités
Newsletter • Réseaux sociaux
EN SAVOIR +

Roadmap
Feuille de route • KPI
EN SAVOIR +

CNRS Inserm INRAE CEA INVESTISSEMENTS D'AVENIR

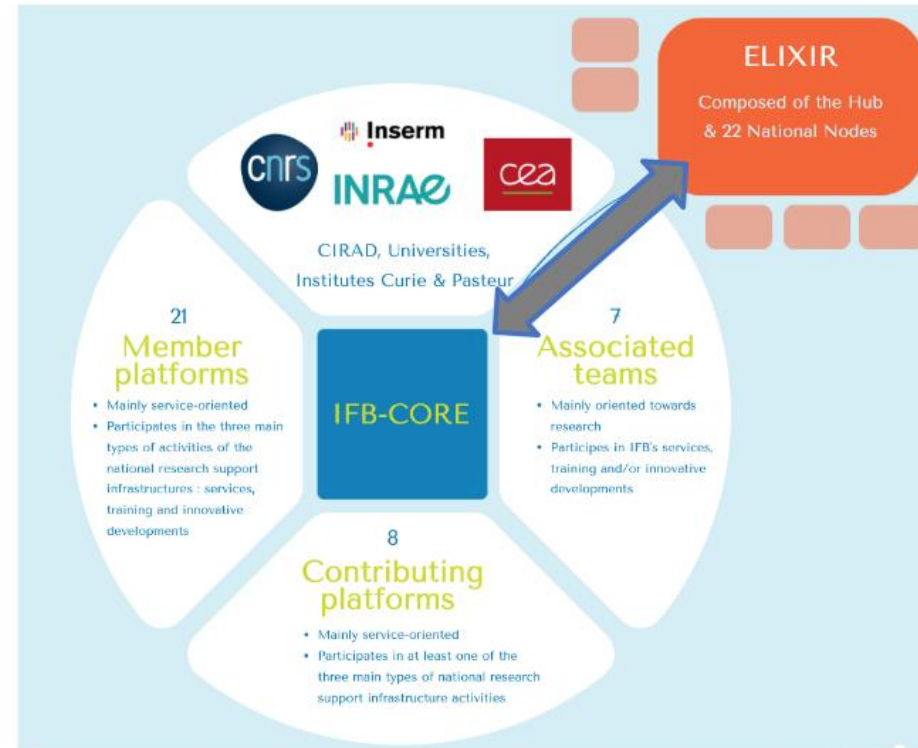
<https://www.france-bioinformatique.fr>

• Acteurs

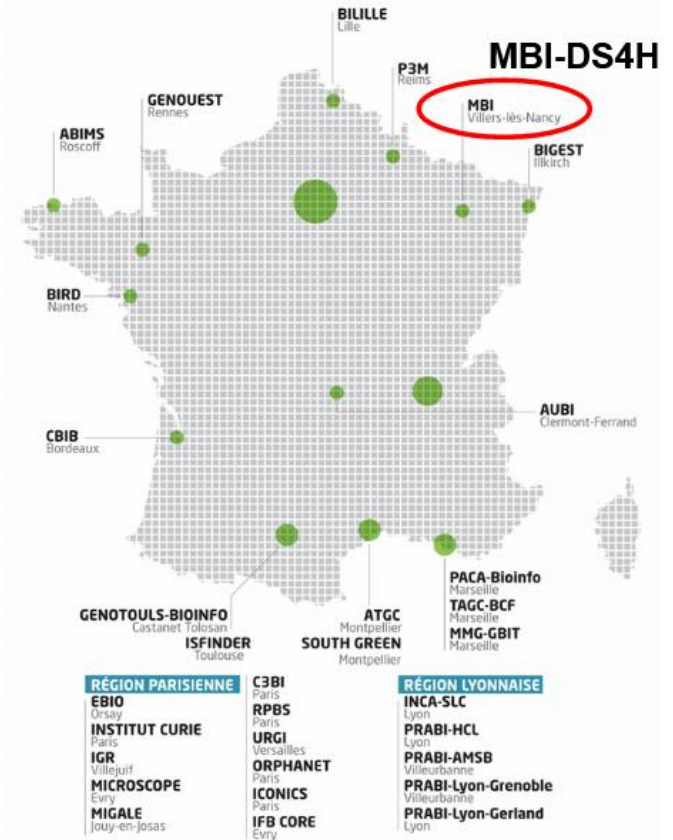
- 29 plateformes et 7 équipes associées dans toute la France
- IFB-Core (UMS3601) : une unité mixte de service
 - *Coordination de l'infrastructure*
 - *Mutualisation des ressources*
 - *Développements méthodologiques*
 - *Soutien aux utilisateurs*
 - *Formation*
 - *Nœud français de l'infrastructure européenne ELIXIR*

• Tutelles

- CNRS, Inserm, INRAE, CEA, Universités, etc.

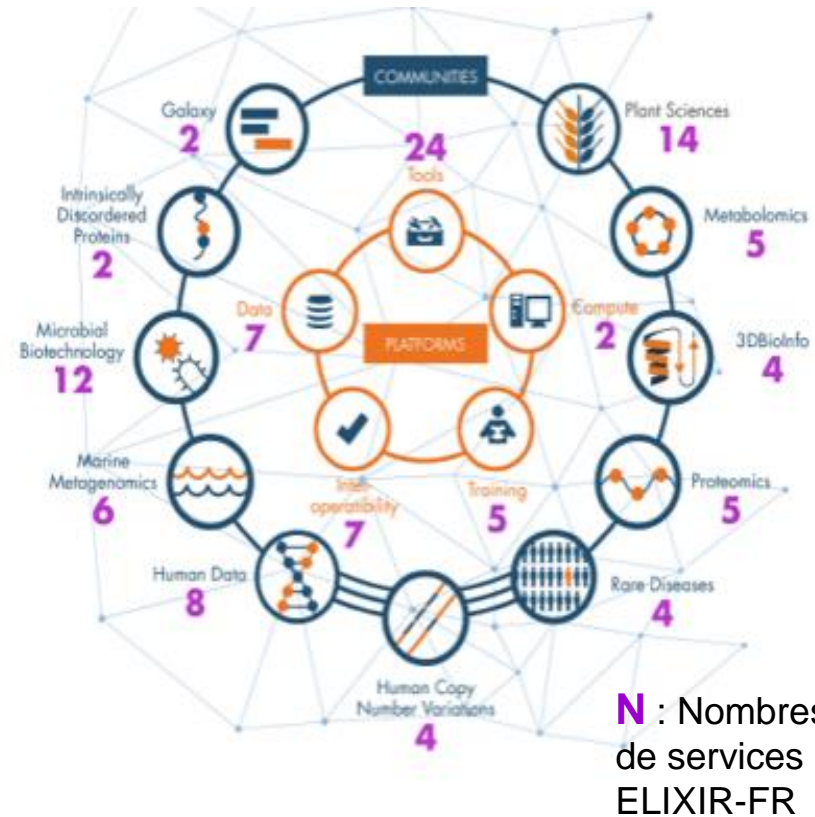


- **Permanents (~ 7 ETP)**
 - 5 IFB-Core (60-100 % ETP)
 - 22 Coordinateurs d'actions IFB (10-30% ETP)
- **Non permanents (~ 25 CDD 2-3 ans)**
 - Affectés à des actions/projets/plateformes
- **Total**
 - Avec les personnels des plateformes:
~ 400 experts pour ~ 200 ETP





- **ELIXIR:**
 - Organization intergouvernementale Européenne
 - Coordination des ressources en bioinformatique de 22 pays
 - <https://elixir-europe.org/about-us>
- **5 plateformes de service (groupes de travail)**
 - TOOLS, COMPUTE, DATA, INTEROPERABILITY TRAINING
- **11 communautés scientifiques**
 - 3D-BioInfo, Intrinsically Disordered Proteins, Galaxy, Metabolomics, Human Data, etc.
 - Une communauté émergente : Data Management (réseau de data stewards)



- Depuis 2017, action Interopérabilité
 - Fournir un accès uniforme à des ressources diversifiées et distribuées
 - Permettre une exploitation intégrée de ces ressources
 - En cohérence avec les objectifs de la Science Ouverte
- Interopérabilité à différents niveaux
 - *Intégration physique des données = dépôts, stockage dans les entrepôts*
 - *Interopérabilité technique = protocoles d'échanges de données, APIs génériques pour accéder aux bases de données et aux outils, containers Docker, Conda*
 - *Interopérabilité syntactique = formats d'échanges pour les données et les méta-données (XML, RDF, OWL, JSON...)*
 - *Interopérabilité sémantique = standards pour décrire le sens des données, terminologies, ontologies (OLS « ontology Lookup Service », BioPortal, AgroPortal...)*
 - *Interopérabilité des outils = description standard des outils, registres (bio.tools à ELIXIR)*
 - *Interopérabilité des workflows = description standard des workflows, de la provenance des résultats au cours de l'exécution (Galaxy, CWL, ...)*

- Depuis 2022, mission transversale Science Ouverte et Interopérabilité (SO&I)
 - Au service des plateformes de l'IFB
 - Au service des utilisateurs extérieurs
- Objectifs
 - **Veille technologique** sur les outils en support à la science ouverte aux niveaux national et international
 - Organiser le **développement** des ressources manquantes pour SO&I en bioinformatique (standards, ontologies, outils logiciels, supports de formations)
 - Favoriser l'appropriation par les **plateformes de l'IFB** des ressources de SO&I existantes et/ou développées à l'IFB et plus largement dans ELIXIR
 - Contribuer à la **formation** en SO&I des communautés scientifiques de biologistes
 - Développer un plan de **communication** sur ces ressources en lien avec le rôle de CRT.

II. Exemples de ressources pour la Science Ouverte et l'Interopérabilité à l'IFB

Catalogue

Waves

maDMP4LS

OpenLink

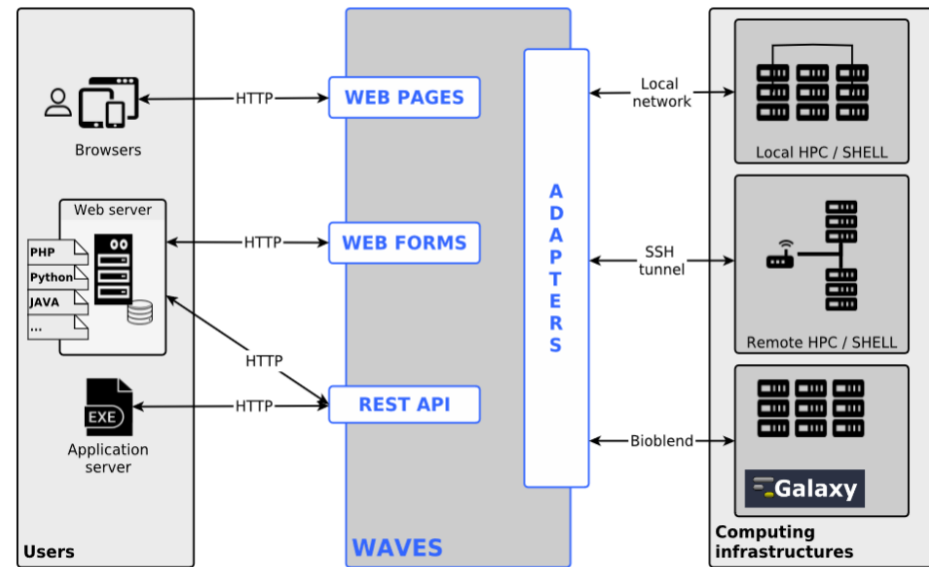
FAIR-Checker

EDAM ontology

- Une base de données collectant les métadonnées de:
 - Outils logiciels disponibles sur les plateformes de l'IFB
 - Bases de données et ontologies hébergées par les plateformes de l'IFB
 - Formations et supports de formation (proposés par l'IFB)
 - Organisations (plateformes labo)
 - Personnes et expertises
- Accessibles via:
 - Le site de l'IFB
 - Une API REST
- Interopérable avec
 - [Bio.tools](#) (registre d'outils d'ELIXIR)
 - [TeSS](#) (catalogue de formations d'ELIXIR)
 - Balises Bioschemas (extension de schema.org)

The screenshot displays the 'Catalogue des outils' (Tools Catalog) website. At the top, there is a navigation bar with the IFB logo and links for 'Contact', 'À propos', 'Services', 'Formation', 'Innovation', 'ELIXIR-FR', and 'Ressources'. A search bar is located in the top right corner. Below the navigation bar, the main heading reads 'Catalogue des outils'. The central section is titled 'Overview of available bioinformatics tools' and includes a search bar with the text 'PhyML'. Below the search bar, there is a grid of tool cards. Each card displays the tool's name, a brief description, and metadata such as 'Licence' and 'Maturity'. The tools shown include acdtool, AgroLD, ANISEED, AnnotQTL, AphidBase, Arabidopsis chloroplast database (AT_CHLORO), ARIA, AskOmics, and Assemble2. The bottom of the page features a green banner with the ELIXIR FRANCE logo.

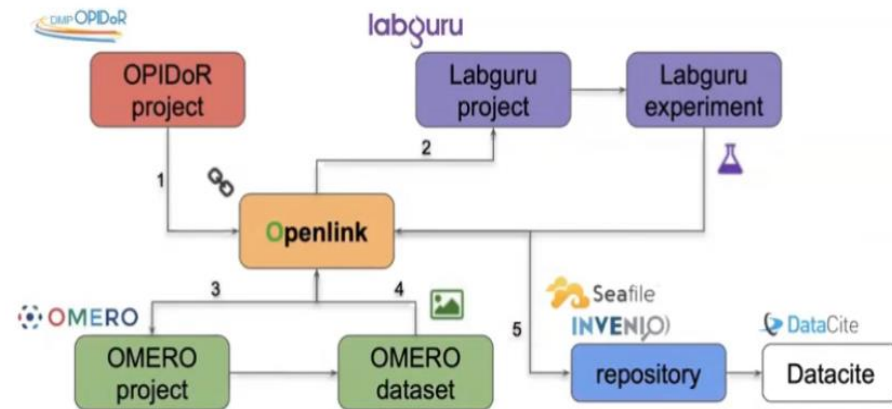
- **BUT** : Intégrer des outils informatiques
- **SOLUTION** :
 - Plateforme générique pour servir des applications
 - Développé en Python (Django framework)
 - Fournisseur d'outils et agrégateur d'outils
- **3 TYPES D'ACCES aux SERVICES BIOINFORMATIQUE**
 - Web pages
 - Web forms
 - REST APIs
- **AVANTAGES**
 - Des adaptateurs pour se connecter aux machines de calcul à l'arrière
 - Les pages web sont automatiquement enrichies avec des métadonnées
- **Exemple** : <https://www.phylogeny.fr/>



Marc Chakiachvili et al. 2019. WAVES: a web application for versatile enhanced bioinformatic services. *Bioinformatics* 35:140-142.

<http://www.atgc-montpellier.fr/waves/>

- **But :**
 - faciliter la conformité FAIR tout au long du cycle des données de la recherche
- **Machine actionable Data Management Plans (IFB-Core)**
 - Contenu DMP structuré, conforme aux standards (modèle OpiDOR)
 - Interfaçage avec les référentiels du domaine pour automatiser la saisie des données et métadonnées
- **OpenLink (IGBMC)**
 - Passerelle entre DMP, cahiers de laboratoires, entrepôts de données et de publications
 - Transfert de métadonnées automatique le long du cycle des données
 - Appliqué aux données d'imagerie comme preuve de concept



- **But**
 - Vérifier et améliorer la conformité FAIR d'une page web
 - Pouvoir tester automatiquement une page web identifiée par un DOI ou un URL
- **Solution : application web avec 2 fonctionnalités**
 - **CHECK** : test des principales métriques FAIR
 - *Avec des recommandations quand le test échoue*
 - **INSPECT** : Inspection avancée des métadonnées incluses dans la page web
 - *Utilisation des technologies du web sémantique (RDF, SPARQL, SHACL)*
 - *Contrôle de qualité pour tester si les termes utilisés sont présents dans des ontologies de référence*

How FAIR is my resource ?

Enter resource identifier (URL/DOI):
<https://doi.pangaea.de/10.1594/PANGAEA.914331>
 The input contains the following DOIs that you can also test: [View recommendations](#)

Progress

Click on results

Radar chart of metrics completion

Principle	Name	Description	Time	Comment	Recommandation	Score	Result	Test
F1	Unique Identifier		0:00:02			1	Success	Check
F1	Identifier Persistence		0:00:02			0	Failure	Check
F1	Data Identifier Persistence		0:00:03			0	Failure	Check
F2	Structured Metadata		0:00:03			1	Success	Check
F2	Grounded Metadata		0:00:03			1	Success	Check
F3	Data Identifier Explicitly In Metadata		0:00:03			0	Failure	Check

Complémentaire à d'autres outils comme FAIR-Evaluator, FAIR Shake, F-UJI (see FAIRassist.org)

<https://fair-checker.france-bioinformatique.fr>

- **Vocabulaire contrôlé hiérarchique**

- Opérations bioinformatiques et concepts de base
- Co-développé ELIXIR-FR et ELIXIR-NW
- Chaque terme a un identifiant unique, est rattaché à un thème et a une place dans l'arborescence de ce thème.

- **Utilisation pour renseigner les métadonnées**

- Lors d'un dépôt dans les entrepôts de données (zenodo, DataVerse)

The screenshot shows the EDAM ontology browser interface. At the top, it says "EDAM ontology" and "EDAM is a simple ontology of well established, familiar concepts that are prevalent within bioinformatics." Below this is a search bar and a dropdown menu set to "1.25 (stable)". The main area displays a hierarchical tree of terms, with "Sequence alignment" selected. To the right, a "Details of term 'Sequence alignment'" panel is open, showing the following information:

Term	Sequence alignment
Definition	Alignment of multiple molecular sequences.
Comment	
Exact synonyms	Multiple sequence alignment msa
Narrow synonyms	
URI	data_0863
Has topic	Sequence analysis
Parents	Alignment
Community usage	
bio.tools	167 times as input (236), 121 times as output (201).
Biosphere	not used by appliances, not used by tools.
TeSS	1 times
Links	Open in AberOWL, BioPortal, OLS or WebProtégé.

<https://edamontology.github.io/edam-browser>





Catalogue des formations

Number of data-items: 7, show on 1 pages.

Training name

Perform a filter using the legend below

Open for registration Upcoming training Past training

-  02 Feb 2023 [Principes FAIR pour la gestion des données d'une plateforme IBISA - session 2023](#)
📅 2023-02-02 | 2023-02-03
📍 Paris France
-  13 Jun 2022 [Principes FAIR dans un projet de bioinformatique - Session 2022](#)
📅 2022-06-13 | 2022-06-15
📍 Paris France
-  22 Mar 2022 [Principes FAIR pour la gestion des données de recherche en sciences de la vie - édition Francilienne - Session 1 \(mars\)](#)
📅 2022-03-22 | 2022-03-23
📍 Paris France
-  [Formation IFB Science Ouverte & PGD - Comment gérer des jeux de données haut débit en sciences de la vie et de la santé - Session 2](#)

- **L'IFB : une infrastructure nationale de soutien à la recherche en Biologie-Santé**
 - Coordination de plateformes de service en bioinformatique
 - Activités ponctuelles en Science Ouverte et Interopérabilité
 - *Conception et développements d'outils manquants*
 - *Formations : FAIR, PGD*
 - Expertise mutualisée au service de la recherche et de l'innovation
- **Rôle de Centre de Ressources Thématiques SO Biologie-Santé**
 - Définir (en coopération avec les communautés, autres INBS et partenaires européens)
 - *Les standards de description des données*
 - *Les standards de diffusion des données*
 - *Les bonnes pratiques associées*
 - *Les entrepôts thématiques de référence*
 - Animer/soutenir
 - *Les communautés pour faire adopter les standards et les bonnes pratiques*
 - *Les interactions (dans RDG) avec les ateliers de la donnée, le catalogue, les entrepôts*
 - Adapter le guichet de l'IFB pour enregistrer les demandes faites au CRT et retourner les réponses



MERCI DE VOTRE COMPREHENSION ET DE VOTRE PATIENCE !

Mission transversale SO&I à l'IFB

- Jean-François Dufayard
- Frédéric de Lamotte
- Alban Gaignard
- Jacques van Helden
- Hélène Chiapello
- Sarah Cohen-Boulakia (-> 2019)
- GT SO&I

Développeurs à l'IFB

- Thomas Rosnet
- Paulette Lieby
- Julien Seiler
- Hervé Ménager
- Thomas Denecker
- Vincent Lefort
- Sylvain Milanese
- Laurent Bourhi
- ...

Partenaires

- INIST (équipe OPiDOR)
- DSW (ELIXIR-CZ)
- EDAM (ELIXIR-NW)
- ...



INSTITUT FRANÇAIS DE BIOINFORMATIQUE



france-bioinformatique.fr



elixir-europe.org



[@IFB_Bioinfo](https://twitter.com/IFB_Bioinfo)



[@InstitutFrançaisDeBioinformatique](https://www.linkedin.com/company/InstitutFrançaisDeBioinformatique)



contact@france-bioinformatique.fr



INRAE

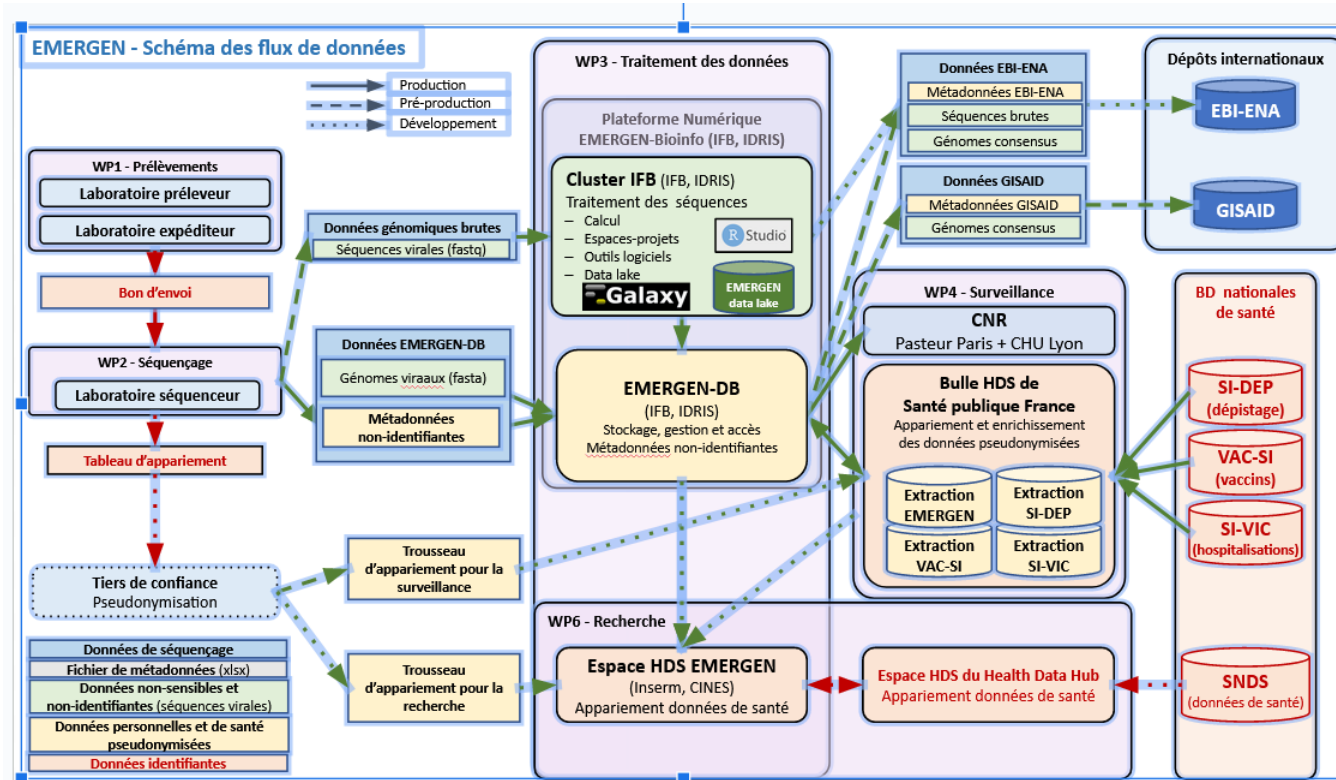


Inserm



DIAPOS COMPLEMENTAIRES

- Projet de surveillance génomique pendant la pandémie Covid-19.
- Circulation des données de séquence complexes car données de santé
- EMERGEN-DB : intermédiaire pour la soumission des séquences (données + métadonnées) aux dépôts internationaux (GISAID, ENA).
 - Mapping de métadonnées
 - Rôle du PGD pour anticiper les choses



- **Introduits in 2016**
 - Article principes de Mark Wilkinson, Michel Dumontier and Barend Mons*
- **Guidelines de haut-niveau**
 - Pour des environnements de publication de données modernes
 - *Dépot manuel et/ou automatique*
 - *Exploration des données*
 - *Partage*
 - *Ré-utilisation*
- **Prévu non seulement pour les données**
 - Mais aussi pour les outils, les workflows
- **Précède l'implémentation**
 - Interprétations variées
 - Effort d'unification



From Australian Research Data Commons

*<https://www.nature.com/articles/sdata201618>

Interoperability FAIR principles



From Australian Research Data Commons

1) <https://www.nature.com/articles/sdata201618>

- Introduced in 2016 by Mark Wilkinson, Michel Dumontier and Barend Mons (1)
- High-level guidelines for contemporary **data publishing environments**
- With respect to supporting both manual and automated data deposition, exploration, sharing, and reuse.
- Intended not only for data: also tools, workflows
- Precedes implementation -> various interpretations available

