

# Programme du 6 novembre – journée du GDR

9h15	Accueil
9h30-10h15	Introduction, nouvelles du GDR
10h15-11h05	Biologie intégrative: Quelques stratégies statistiques Guillem Rigaill (LaMME et <a href="#">IPS2</a> ) et Marie-Laure Martin-Magniette (MIA Paris et <a href="#">IPS2</a> )
11h05-11h20	Pause
11h20-12h10	Biologie intégrative: workflows scientifiques pour la reproductibilité computationnelle <a href="#">Sarah Cohen Boulakia</a> (LRI, Université Paris Sud)
12h10-13h00	Peer Community in: Un système public et gratuit de recommandations de preprints fondé sur le peer-reviewing <a href="#">Thomas Guillemaud</a> (INRA Sophia Antipolis)
13h-14h	Buffet déjeunatoire
14h-14h30	Intervention de l'INS2I, échanges Adeline Nazarenko
14h30-15h20	Biologie Intégrative: Facilitating complex life science data integration and reuse. <a href="#">Olivier Dameron</a> (IRISA, Université de Rennes 1)
15h20-15h40	Pause
15h40-17h	L' <a href="#">Institut Français de Bioformatique</a> : bilan et projets. Quelle synergie avec le GDR BIM ? <a href="#">Jacques van Helden</a> , co-directeur de l'IFB

# Institut Français de Bioinformatique (IFB)

## Une infrastructure nationale de support à la recherche

<https://frama.link/ifb-gdr-bim-2019-11-06>



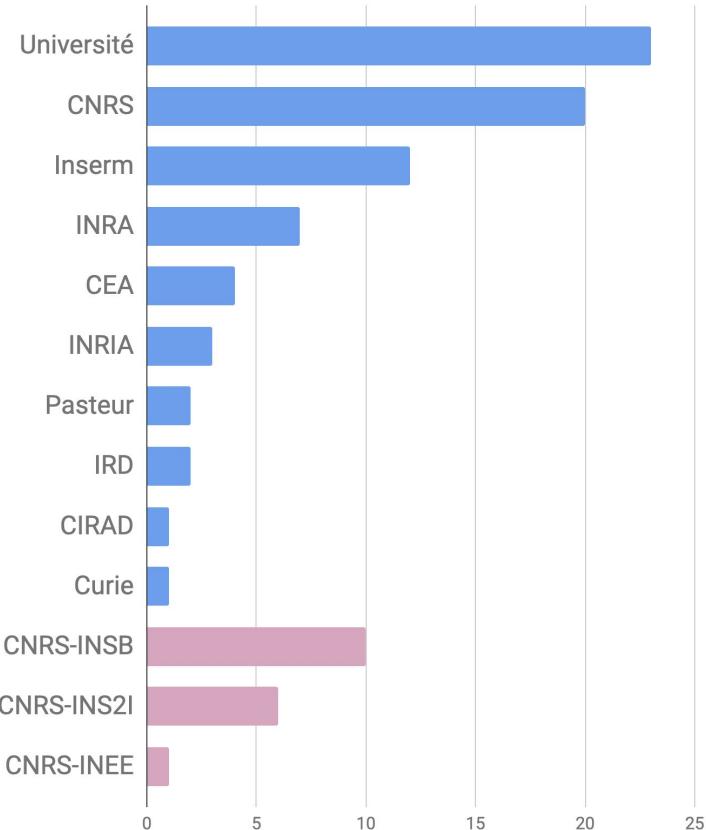
*Jacques van Helden*

# Services

# L'Institut Français de Bioinformatique

- 29 plateformes de services en bioinformatique
  - + plateforme coordinatrice **IFB-core**
    - **Proximité:** couverture de l'ensemble du territoire national
    - 400 personnes, 200 ETP
- Tutelles
  - IFB: CNRS, INRA, Inserm, CEA, INRIA
  - + tutelles plateformes: Universités, Pasteur, IRD, CIRAD, Curie, IGR.
- Couverture, expertise et applications
  - tous les **domaines de la bioinformatique**
  - tous les **domaines du vivant** (biologie fondamentale, médecine, agronomie, environnement, biotechnologies)
- Noeud français de l'ESFRI de bioinformatique ELIXIR.

Tutelles des plateformes IFB



## Calcul et stockage

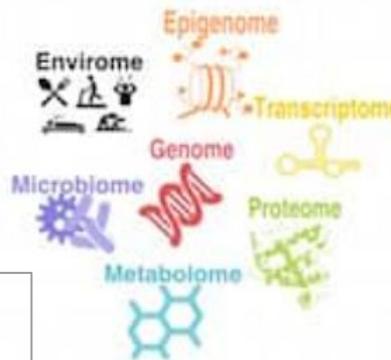
- national network of computing resources (**NNCR**)
- high-performance computing (**HPC**)
- solutions **cloud + cluster**
- politique de protection des données (**GDPR**)

## Développement et déploiement logiciel

- développement de logiciel spécialisé
- packaging (**conda**)
- virtualisation (machines virtuelles, containers)
- bonnes pratiques d'ingénierie logicielle
- code ouvert et collaboratif

## Accompagnement de projets

- de la conception aux livrables
- plans de gestion des données (**DMP**)
- conception de **workflows**
- science des données



## Le guichet d'orientation

- un **point d'entrée unique** pour tous les services et plateformes IFB
- un réseau de 400 experts couvrant tous les **domaines de la bioinformatique** (séquence, structure, réseaux, ...)
- tous les **domaines de recherche** : biologie fondamentale agriculture, santé, environnement, biotechnologies

## Formation

- écoles thématiques (NGS & autres omiques, phylogénie, biostat, programmation, ...)
- tutorat “Bring Your Own Data” (BYOD)
- webinaires
- MOOC
- diplôme universitaire
- adaptation à l'évolution de la demande

## Bioinformatique intégrative

- multi-omique, structure, imagerie, ...
- approches statistiques
- réseaux multiplexes

# Une diversité d'usagers avec des besoins différents

- Granularité institutionnelle
  - Individus
  - Équipes de recherche
  - Plateformes de services
  - Formateurs (formation initiale ou continue)
  - Infrastructures nationales de recherche productrices de données
  - Organismes (EPST, IT, ITMO)
- Diversité disciplinaire
  - Biologistes
  - Bioinformaticiens
- Infrastructure
  - Comptes et espaces projets individuels
  - Espaces projets partagés
  - Utilisation de ressources informatiques pour des formations
  - Déploiement de ressources développées par les bioinformaticiens
- Expertise
  - Conception de projets en amont (schémas des analyses, DMP, ...)
  - Accompagnement de projets (prestations/collaborations)
  - Evaluation de la FAIRification
- Formation
  - Des biologistes
  - Des bioinformaticiens
- Bonnes pratiques
  - Science reproduitible
  - FAIR
  - Standards internationaux (ELIXIR)
- Mutualisation des ressources humaines
  - “Task force” mutualisée
  - Support communautaire

# Fédérer les forces vives

- Feuille de route 2018-2021 (voir trombinoscope)
  - **Responsables d'action** sur l'ensemble des plateformes (engagement volontaire, lettres de missions)
  - **CDD IFB-core** accueillis sur la plateforme régionale la plus appropriée à leur mission
- Groupes de travail
  - Inscription volontaire
  - Réunions régulières en visio + présentielle occasionnelles
- “Task force” spécialisée mutualisée du NNCR
  - Implication des statutaires
- Support communautaire
  - <https://community.cluster.france-bioinformatique.fr/>
- Guichet central
  - [contact@groupes.france-bioinformatique.fr](mailto:contact@groupes.france-bioinformatique.fr)
  - réception et suivi des demandes
  - orientation vers compétences des plateformes
  - indicateurs

## WORK PACKAGE LEADERS AND IFB-CORE STAFF IN 2018

### 2017-2018 workplan

#### WP1. National network of services in bioinformatics (NNCR)

Christophe Blanchet<sup>1</sup> & Jacques van Helden<sup>2</sup>  
Physical Infrastructure:  
compute and storage  
Julien Seize<sup>3</sup> & Christophe Blanchet<sup>1</sup>  
Software and data environment  
Gildas Le Corguillé<sup>4</sup> & Olivier Collin<sup>5</sup>  
Support to Databases  
Claudine Médigue<sup>6</sup> & Guy Perrére<sup>7</sup>  
Catalogue of French resources in  
bioinformatics  
Jacques van Helden<sup>2</sup> & Hervé Ménager<sup>8</sup>  
Access to end-users  
Christine Gaspin<sup>9</sup> & Erwan Core<sup>10</sup>  
Shared services with other national infra  
Christophe Bruley<sup>11</sup> & Jean-François Dufayard<sup>12</sup>  
Consulting and orientation desk  
Vincent Lefort<sup>13</sup> & Ivan Moszer<sup>14</sup>

#### WP2. Innovation: towards integrative bioinformatics

Claudine Médigue<sup>6</sup> & Olivier Sand<sup>15</sup>  
Pilot projects in integrative bioinformatics  
Claudine Médigue<sup>6</sup> & Etienne Thévenod<sup>16</sup>  
& Olivier Sand<sup>15</sup>  
Call for challenges and unlocking of  
technological and scientific barriers  
Ivan Moszer<sup>14</sup> & Jacques van Helden<sup>2</sup>  
Catalyzing interoperability and integration  
between resources  
Alban Gaignard<sup>17</sup> & Marie-Dominique Devignes<sup>18</sup>

#### WP3. Links with international networks and industry

Anne-Françoise Adam-Blondon<sup>19</sup>  
& Claudine Médigue<sup>6</sup>  
IFB as the French node of the European  
ELIXIR network  
Anne-Françoise Adam-Blondon<sup>19</sup>  
& Claudine Médigue<sup>6</sup>  
Partnership with industries  
Patrick Durand<sup>20</sup> & Yves Vandennebrouck<sup>21</sup>  
& Victoria Dominguez Del Angel<sup>22</sup>

A map of France where each dot represents a staff member's location. The dots are color-coded by region: Nord-Pas-de-Calais (orange), Hauts-de-France (light blue), Picardie (yellow), Centre (green), Bourgogne-Franche-Comté (purple), Franche-Comté (pink), Auvergne-Rhône-Alpes (red), Rhône-Alpes (blue), Provence-Alpes-Côte d'Azur (teal), and Occitanie (yellow). Each dot contains a small portrait of a staff member and a number indicating their ID.

Laurent Bouri (Catalogue + Web)<sup>27</sup>  
Mélanie Buy (Pilot-projects)<sup>29</sup>  
Etienne Cameren (Pilot-projects)<sup>29</sup>  
Nicole Charnière (NNCR)<sup>30</sup>  
Victoria Dominguez Del Angel<sup>22</sup>  
(ELMR, NNCR)<sup>22</sup>  
Hervé Ménager (Pilot-projects)<sup>31</sup>  
Jonathan Lorenzo (NNCR)<sup>22</sup>  
Olivier MERZEAU (Communication)<sup>32</sup>  
Sylvaine Milanes (Consulting and orientation desk)<sup>33</sup>  
Guillaume Postic (Pilot-projects)<sup>35</sup>  
Valentin Saint-Léger (Admin & management)<sup>36</sup>

#### WP4. Training and outreach

Jacques van Helden<sup>2</sup> & Morgane Thomas-Chollier<sup>23</sup>  
Training  
Jacques van Helden<sup>2</sup> & Hélène Chapello<sup>24</sup>  
Join actions with SFB1 and GDR BIM  
Morgane Thomas-Chollier<sup>23</sup> & Hélène Touzet<sup>25</sup>  
Communication & Valorisation  
Claudine Médigue<sup>6</sup> & Victoria Dominguez Del Angel<sup>22</sup>  
  
**WP5. Management and governance**  
Claudine Médigue<sup>6</sup> & Jacques van Helden<sup>2</sup>  
Governance and coordination bodies  
Claudine Médigue<sup>6</sup> & Jacques van Helden<sup>2</sup>  
Quality management system  
Claudine Médigue<sup>6</sup>  
IFB Economic model  
Christine Gaspin<sup>9</sup>  
  
**Communities**  
Bioinformatics for health  
David Salgado<sup>26</sup> & Ivan Moszer<sup>14</sup>  
  
**IFB-core staff 2018**  
Laurent Bouri (Catalogue + Web)<sup>27</sup>  
Mélanie Buy (Pilot-projects)<sup>29</sup>  
Etienne Cameren (Pilot-projects)<sup>29</sup>  
Nicole Charnière (NNCR)<sup>30</sup>  
Victoria Dominguez Del Angel<sup>22</sup>  
(ELMR, NNCR)<sup>22</sup>  
Hervé Ménager (Pilot-projects)<sup>31</sup>  
Jonathan Lorenzo (NNCR)<sup>22</sup>  
Olivier MERZEAU (Communication)<sup>32</sup>  
Sylvaine Milanes (Consulting and orientation desk)<sup>33</sup>  
Guillaume Postic (Pilot-projects)<sup>35</sup>  
Valentin Saint-Léger (Admin & management)<sup>36</sup>

# NNCR - Hardware open for the national service

	Servers	Location	Compute (#CPU HT*)	Storage (#TB)	RAM (#GB)	RAM/cor e (#GB)
	ifb-core-cluster	Orsay (IDRIS)	4 300	400	20 008	5.00
	ABiMS	Roscoff	1 928	2 000	10 600	5.50
	GENOTOUL	Toulouse	6 128	3 000	34 304	5.60
	GenOuest	Rennes	1 824	2 300	7 500	4.11
	BiRD	Nantes	560	500	4 000	7.14
	Migale	Jouy en Josas	1 084	350	7 000	6.46
	<b>Total federation of clusters</b>		<b>15 824</b>	<b>8 550</b>	<b>83 412</b>	<b>5.37</b>
	ifb-core-cloud	Lyon (CC-IN2P3)	3 936	408	20 408	5.18
	ifb-genouest-genostack	Rennes	600	350	2 600	4.33
	ifb-prabi-girofle	Lyon	448	144	1 500	3.35
	ifb-bird-stack	Nantes	160	50	448	2.80
	ifb-bistro-iphc	Strasbourg	85	50	1 500	17.65
	<b>Total federation of clouds</b>		<b>5 229</b>	<b>1 002</b>	<b>26 456</b>	<b>5.06</b>
	<b>Grand total (clusters + clouds)</b>		<b>21 053</b>	<b>9 552</b>	<b>109 868</b>	<b>5.29</b>



# IFB work plan 2018-2021

## WP1. Distributed national environment of services in bioinformatics

Christophe Blanchet & Jacques van Helden

### A1.1 Physical infrastructure: compute and storage

Julien Seiler & Christophe Blanchet

### A1.2 Software and data environment

Gildas Le Corguillé & Jacques van Helden

### A1.3 Support to Databases

Claudine Médigue & Guy Perrière

### A1.4 Catalogue of French resources in bioinformatics

Jacques van Helden & Hervé Ménager

### A1.5 Access to end-users

Christine Gaspin & Erwan Corre

### A1.6 Sharing of services with other national research support infrastructures

Christophe Bruley & Jean-François Dufayard

### A1.7 Consulting and orientation desk

Jacques van Helden & Ivan Moszer

## WP3. Links with international networks and industry

Anne-Françoise Adam-Blondin & Claudine Médigue

### A3.1 IFB as the French node of the European ELIXIR network

Anne-Françoise Adam-Blondin & Claudine Médigue

### A3.2 Partnership with industries

Patrick Durand et Yves Vandenbrouck

## WP4. Training and outreach

Jacques van Helden & Morgane Thomas-Chollier

### A4.1 Training

Jacques van Helden & Hélène Chiapello

### A4.2 Joint actions with SFBI and GDR BIM

Morgane Thomas-Chollier (SFBI) & Hélène Touzet (GDR BIM)

### A4.3 Communication & Valorisation

Claudine Médigue & Victoria Dominguez Del Angel

## WP2. Innovation: towards integrative bioinformatics

Claudine Médigue & Jacques van Helden

### A2.1 Pilot projects in integrative bioinformatics

Claudine Médigue & Etienne Thévenot & Olivier Sand

### A2.2 Call for challenges and unlocking of technological and scientific barriers

Ivan Moszer & Jacques van Helden

### A2.3 Catalysing interoperability and integration between resources

Alban Gaignard & Marie-Dominique Devignes

## WP5. Management and governance

Claudine Médigue, Jacques van Helden

### A5.1 Governance and coordination bodies

Claudine Médigue & Jacques van Helden

### A5.2 Quality management system

Claudine Médigue

### A5.3 IFB Economic model

Christine Gaspin

# Déploiement de ressources logicielles

# Site de support communautaire

- Point d'entrée des requêtes spécifiques au cluster IFB
- Extension du support assuré par la task force à l'ensemble des usagers
- Suivi collectif des sujets
- Réponse rapide
- Partage de compétences
- Consolidation des solutions
- Structuration d'un tissu national de compétences

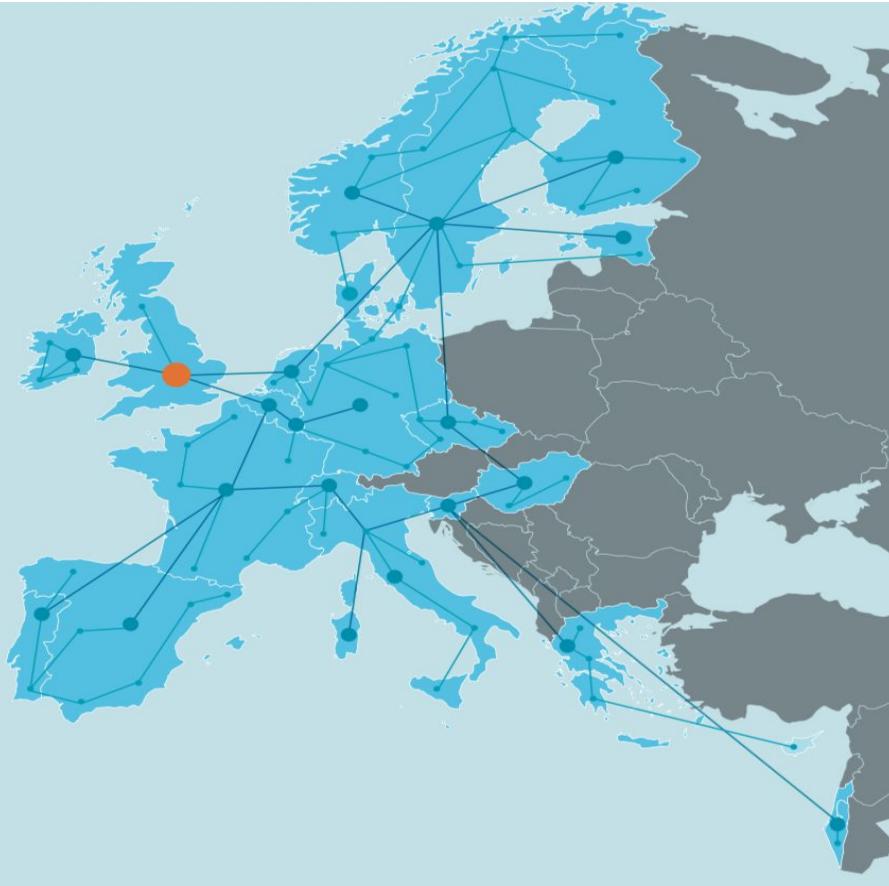
The screenshot shows a web browser window with the URL [community.cluster.france-bioinformatique.fr](https://community.cluster.france-bioinformatique.fr). The page title is "Cluster Community Support". The navigation bar includes links for "toutes les catégories", "toutes les étiquettes", "Récents" (which is highlighted in red), "Non lus (3)", "Top", "Catégories", and a "Créer un sujet" button.

The main content area displays a list of recent topics:

Sujet	Réponses	Vues	Activité
☒ Demande extension de quota / interruption job Support IFB Core Cluster	7	58	3j
☒ Demande installation dernière version Salmon Demande de logiciel ou de banque de données	2	9	4j
☒ Demande d'installation de ABySS Demande de logiciel ou de banque de données	1	7	7j
Templates emails et communication Loungue	5	41	12j
☒ Job sur bigmem ... pending status depuis plusieurs jours Support IFB Core Cluster slurm	6	22	12j
Installation Shasta et isONclust Demande de logiciel ou de banque de données	2	32	12j
☒ Installation bibliothèque R RMySQL ② Demande de logiciel ou de banque de données	5	58	19j
☒ Citation de l'infrastructure M	1	43	19j
☒ Formule pour reconnaître l'ifb dans une publication Support IFB Core Cluster	3	83	19j
☒ Installation scikit-allel Demande de logiciel ou de banque de données conda	2	35	21j
Mémoire / cpus blast Support scientifique slurm, blast	6	182	24j

## L'IFB, noeud français du réseau ELIXIR

# ELIXIR - infrastructure bioinformatique Européenne



=> infrastructure bioinformatique distribuée qui vise à offrir un **support à la recherche en science de la vie, environnement, santé, biotechnologies...**

- <https://www.elixir-europe.org/>
- 23 nœuds (22 pays + EMBL-EBI)
- 200 instituts, >700 experts
  
- Le Nœud Français est porté par l'IFB
  - **Contribution France: ~1M€/an**
  - 2019-2023: la France est le 3ème financeur (après DE et UK)

- Les plateformes constituent l'offre de services d'ELIXIR



- **données** (bases de données)
- moyens de **calcul et stockage**
- **outils** (packaging, workflows)
- **interopérabilité** (standards, identification des objets à travers les plateformes, ontologies dans les domaines biologie-santé, ...)
- **formations** (des biologistes, des bioinformaticiens)

## Platforms

ELIXIR's activities are divided into five areas called 'Platforms'. These are Data, Tools, Interoperability, Compute and Training. The Platforms are managed by Platform leaders and the work is carried out by groups within the Platforms.



### Data Platform

Aims to identify key data resources across Europe and support the linkages between data and literature.



### Compute Platform

Develops services to make it easier to store, share and analyse large datasets.



### Tools Platform

Helps researchers find the best software tools to analyse their data.



### Interoperability Platform

Develops and encourages the adoption of standards to describe life science data.



### Training Platform

Helps scientists and developers find the training they need, and also provides that training.

# Interopérabilité – Articulation IFB / ELIXIR

## Apports de l'infra européenne **ELIXIR**

(y compris **ELIXIR-FR** == **IFB**)

- Etablissement de **standards** ([Bioschemas.org](#))
- **Portabilité** des outils (conda)
- **Workflows** (CWL, Galaxy)
- **Accessibilité** via interfaces conviviales (Galaxy)
- **Virtualisation** (docker, singularity)
- **Ontologies** (EDAM: types de données, formats, opérations, ...)
- **Catalogues** (bio.tools, biocontainers)
- **Formations** à l'interopérabilité
- **Effet-levier** des formations (train the trainer)

## Activités du GT interopérabilité IFB

- Mesure “**FAIRness**” des outils en ligne
- **Catalogue des ressources françaises** en bioinformatique
  - Ressources physiques des plateformes (calcul et stockage)
  - Outils
  - Databases
  - Formations (événements + matériel)
  - Expertises
  - Organisations (plateformes, équipes de recherche)
- **Exposition** des ressources française via Bioschemas and schema.org
- Accessibilité aux usagers (GUI) et programmatique (API) avec l'outil WAVES



Bioschemas



# Communautés ELIXIR

= applications en sciences de la vie en support aux développements technologiques pour :

- ❑ répondre aux besoins de la communauté scientifique
- ❑ guider la coordination et tester les technologies et solutions développées dans un domaine particulier: définition de standards, développement de nouveaux services (à travers les Implementation Studies)

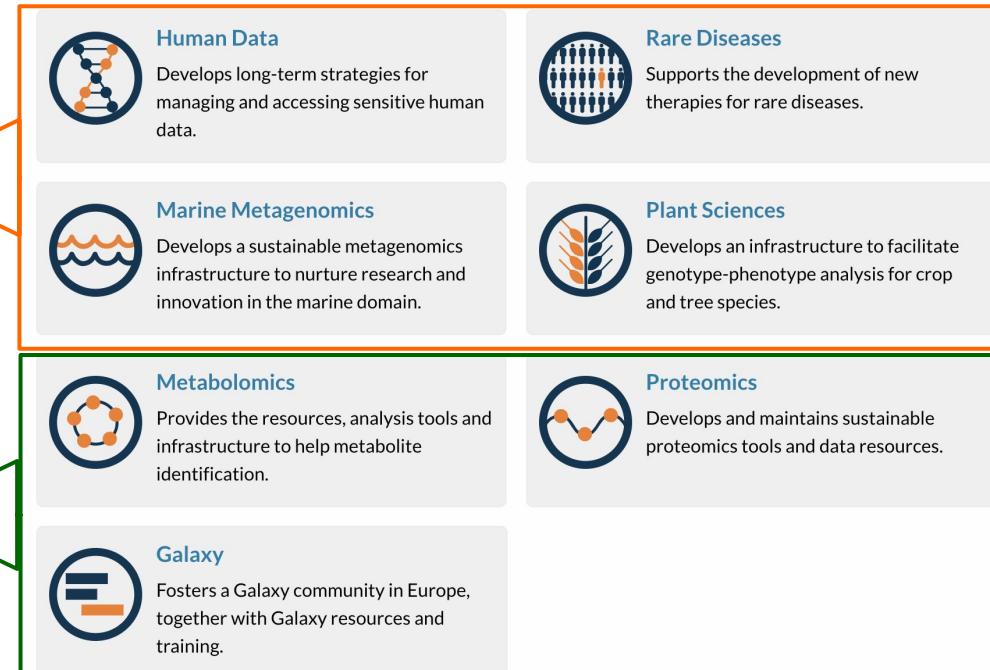
## ■ 4 communautés de la feuille de route

2015-2019: mobilisation forte de ELIXIR-FR dans les communautés **plantes** (FR co-lead depuis 2017), **marine metagenomics**, et **rare diseases**

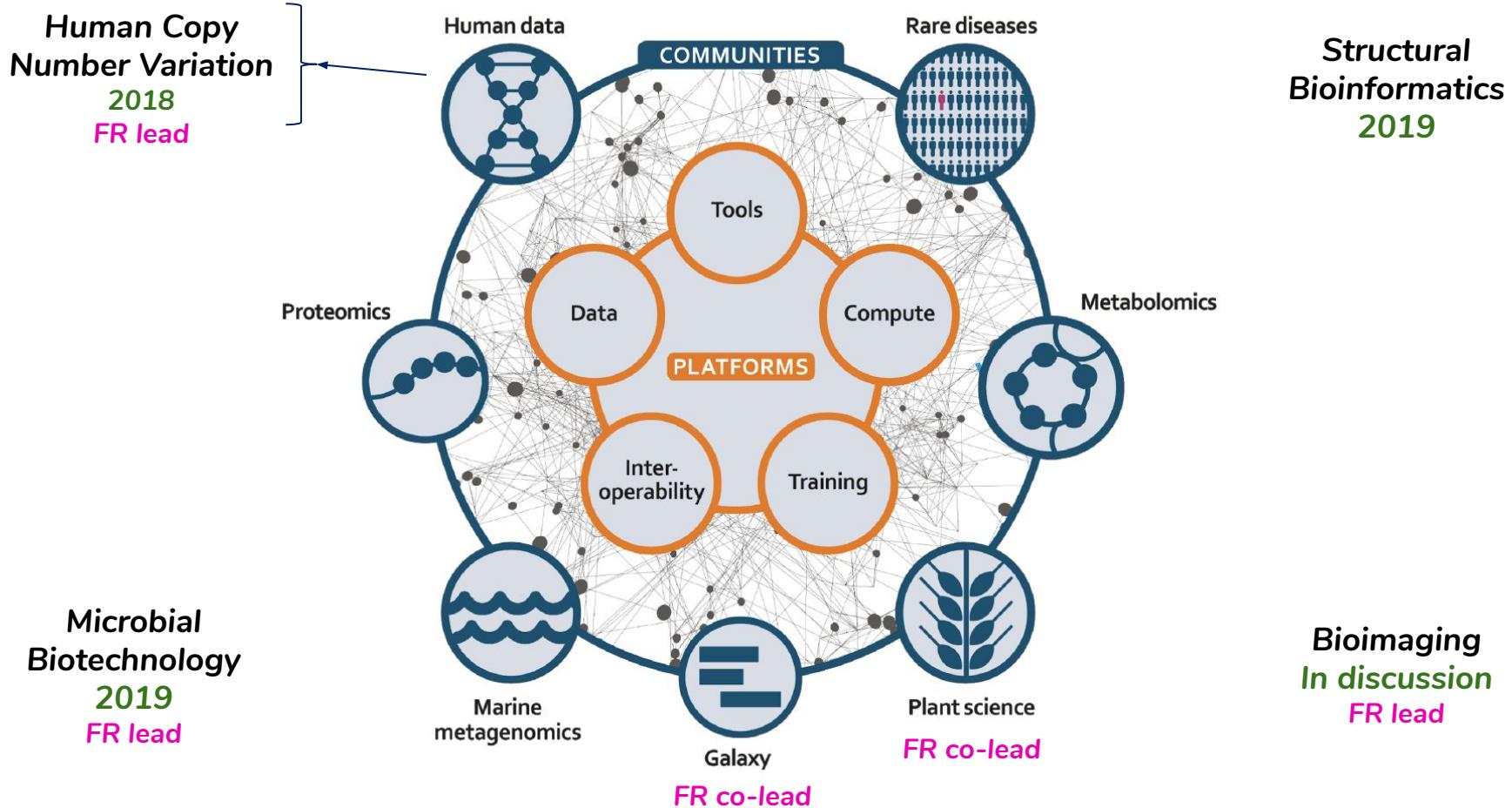
## ■ 3 communautés validées en 2018:

ELIXIR-FR

- participants très actifs dans les communautés **Metabolomics** et **Proteomics**
- co-leader de la communauté **Galaxy**



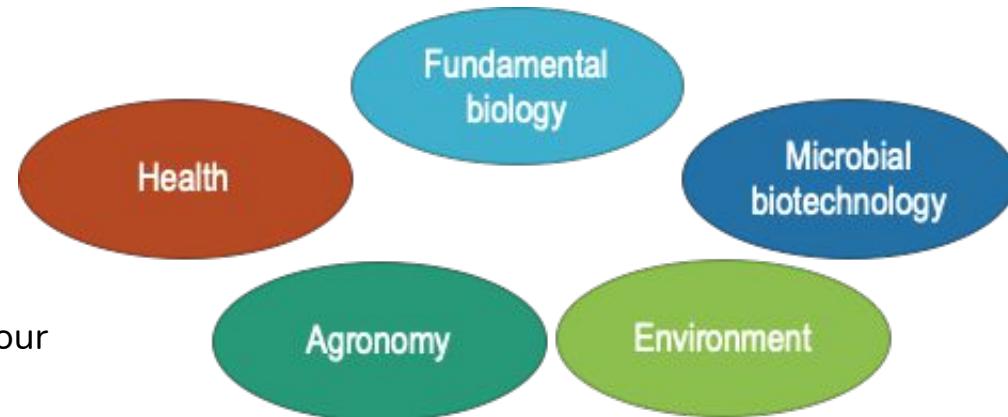
# French contribution to ELIXIR new communities



# Communautés IFB

L'IFB met en place ses propres communautés, avec les mêmes finalités que celles d'ELIXIR

- Organiser des groupes d'**experts** dans des domaines ciblés, **afin d'atteindre des objectifs spécifiques à l'IFB et à ses tutelles**
- Structurer et organiser la communauté française, **coordonner** les initiatives, **rassembler** les ressources
- Organiser une réponse concertée à des requêtes d'usagers et des institutions (tutelles, partenaires)
- Identifier rapidement des **partenaires IFB** pour répondre à des appels à financements nationaux et internationaux



## Communautés IFB

- 5 communautés potentielles identifiées sur base de l'expertise des plateformes
- ouverture à l'émergence d'autres communautés

# The “Bioinformatics for Health” community

## Main achievements



- Contributed to **ELIXIR** activities
  - **INFRADEV3 proposal** (pilot projects), **proposed French partners participation** in the Human Federated Data implementation study and platform actions (DATA platform)
- **Represented the IFB** at the “Journée Recherche et Santé for Clinical phenotyping and Systems Biology”, INSERM and Aviesan (November 22<sup>nd</sup> 2018): talk on “Data science for health and medicine: The French landscape”, C. Béroud
- **Invited to participate in the INSERM transversal project** called “Human Development Cell Atlas” (HuDeCa)
- **Co-organized and participated in an INSERM workshop** on Health Bioinformatics Infrastructures (January 25<sup>th</sup> 2019): talks from C. Médigue, J. van Helden, P. Hupé, H. Chiapello
- **Co-organised a user survey of INSERM laboratories** to collect needs in bioinformatics (software tools, external bioinformatics services, internal capacity building, training programmes, recruitment policies, ...) – to be extended to other labs from governing bodies (CNRS, INRA, INRIA, CEA) – April 2019



# Elixir-FR contributions (technological platforms)

French contribution

ELIXIR  
Training &  
Interoperability  
Platforms



Training  
materials



Crosslinks  
(TeSS,  
Identifiers.org,  
MyExperiment  
etc)



User  
experience

## Tools Platform

An integrated environment for life science software driven by established and emerging ELIXIR Communities

Registry of basic  
information  
& identifiers (bio.  
tools)

Packaging,  
containerisation  
& deployment  
(BioContainers)

Performance  
benchmarking  
& technical  
monitoring  
(OpenEBench)

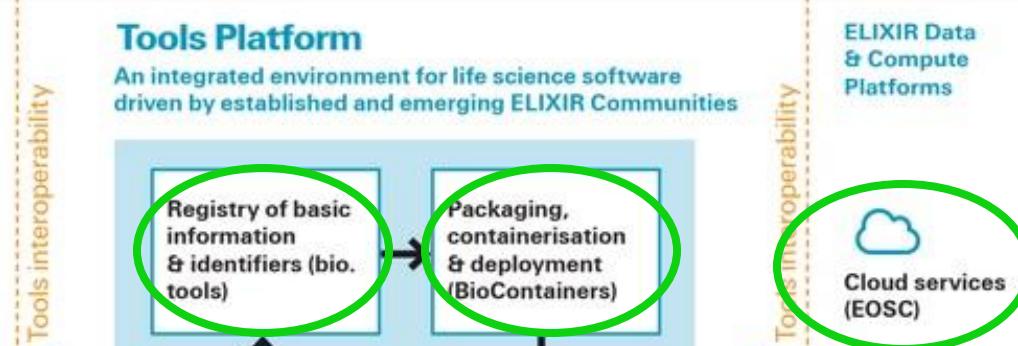
Integration,  
execution  
& interoperability  
(e.g. Galaxy  
platform)



Information  
standards  
(bioToolsSchema,  
CWL, OpenAPI,  
Bioschemas,  
EDAM)

Software  
development  
best practices

(Scientific)  
Community  
support &  
engagement



ELIXIR Data  
& Compute  
Platforms



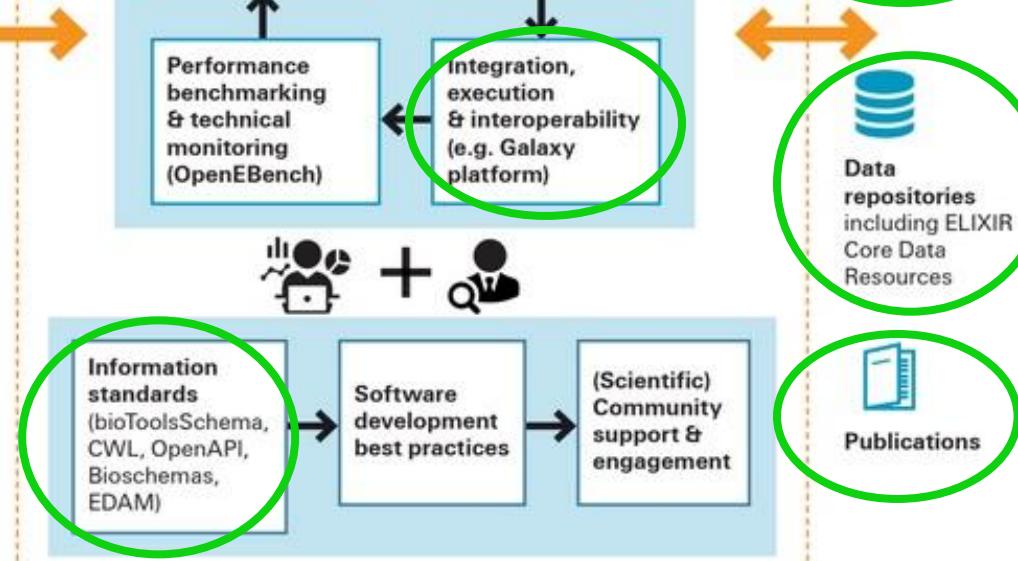
Cloud services  
(EOSC)



Data  
repositories  
including ELIXIR  
Core Data  
Resources



Publications



# Résumé des actions ELIXIR-FR depuis 2017

- Implication d'ELIXIR-FR dans **12 Implementation Studies** acceptées en **2017-2018** (ELIXIR-FR a reçu 466k€ d'ELIXIR pour des IS depuis 2015)
- Participation d'ELIXIR-FR dans **5 Implementations Studies** en **2019**
- **Implication de ELIXIR-FR dans le plan de travail 2019-2023 ELIXIR:** contribution à trois IS stratégiques et à deux WP (dont une en coordination dans la PF interoperability)
- **Participation à des projets Européens acceptés:**
  - H2020 EUCANCan, EOSC Pillard et EOSC Life -> faire travailler ensemble des ESFRI sur leurs services aux utilisateurs.
  - EOSC Life, GenRes Bridge -> opportunités de coordonner les activités d'ELIXIR avec les cluster d'Infra en environnement sur la biodiversité.
  - ELIXIR-CONVERGE : réseau EU autour des plans de gestion des données (DMP)
- **Procédure de sélection des ressources Française** pour la mise à jour du Service Delivery Plan (SDP) du nœud ELIXIR-FR, en tenant compte des critères d'éligibilité d'ELIXIR et des priorités nationales
  - 82 dossiers (sur 103 soumissions) actuellement expertisés

# ELIXIR Service Delivery Plan (SDP)

- Services **ELIXIR**
  - <https://elixir-europe.org/services>
  - 12 services actuellement déclarés
- Critères
  - Communautés d'usagers
  - Liens avec communautés ELIXIR
  - Pérennité
  - Gouvernance
  - Durabilité
- Campagne 2018
  - Appel à candidatures pour faire remonter des services au catalogue IFB + catalogue ELIXIR des services.
  - 103 soumissions, dont 82 pour SDP ELIXIR
  - 25 sélectionnées, à remonter à ELIXIR
  - Accompagnement des autres ressources pour progresser sur les critères non satisfaits

<b>CAZy</b>	The CAZy platform performs the analysis of finished genomes in order to identify the encoded enzymes for the assembly and breakdown of complex carbohydrates.
<b>CRISPRdb</b>	The CRISPRdb database and services are leading international resources for CRISPR sequence analysis.
<b>French Academic Cloud</b>	IFB is currently setting up a federation of Clouds, with a unique entry point, based on a network of distributed regional IT facilities structured around a large national IT facility. The distributed nature of the federation is designed to facilitate data flows (regional PFs are the network entry points in their regions).
<b>GnPLS</b>	GnPLS is an interoperable Information System for plant and pest genomics. It is a powerful multispecies centralized information system with seven linked relational databases.
<b>HOGENOM/ HOVERGEN/ HOMOLENS</b>	A database of homologous genes from fully sequenced organisms (bacteria, archaea and eukarya). It allows you to select sets of homologous genes among species, and to visualize multiple alignments and phylogenetic trees.
<b>IMGT</b>	An integrated knowledge resource specialized in the immunoglobulins (IG) or antibodies, T cell receptors (TR), and major histocompatibility (MH) of human and other vertebrate species.
<b>IS Finder</b>	International Reference Center for Prokaryotic Insertion Sequences (IS) is a public database and a research tool that provides a list of insertion sequences (IS) isolated from bacteria and archaea.
<b>MicroScope</b>	MicroScope is an integrative resource that supports systematic and efficient revision of microbial genome annotation, data management and comparative analysis.
<b>Orphadata</b>	Orphadata provides the scientific community with comprehensive, quality data sets related to rare diseases and orphan drugs from the Orphanet knowledge base, in reusable formats.
<b>Orphanet</b>	Orphanet is the reference resource for information and data on rare diseases and orphan drugs. Orphanet derives from its knowledge base an ontology of rare diseases, information on rare diseases and data on rare diseases.
<b>phyML &amp; Phylogeny.fr</b>	PhyML (Phylogenetic inference using Maximum Likelihood) implements a heuristic algorithm to build phylogenetic trees from the analysis of genetic sequences in a sound statistical framework.
<b>ProDom</b>	This database contains families of homologous protein domains automatically constructed by sequence clustering of all sequences from the UniProt database, which allows for the most extensive coverage of the protein sequence space.

## WELCOME TO ORPHADATA

ACCESS TO AGGREGATED DATA FROM

orphanet

Orphadata named ELIXIR  
Core Data Resource



# Elixir Training: Professional skills for managing and exploiting data

**Mission** : to establish an interactive ELIXIR-wide training community across all member states

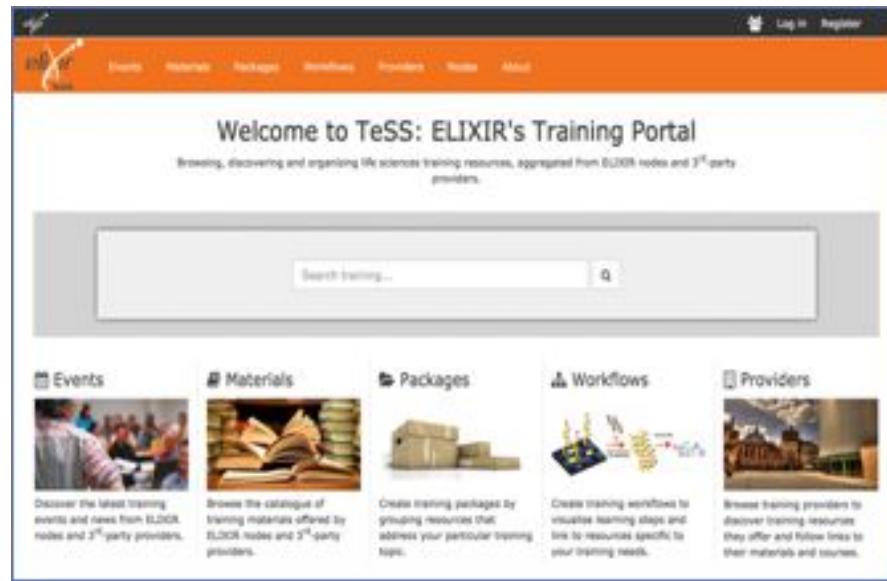
1. Develop and deliver training in selected gap areas for:
  - a. Research (TtR)
  - b. Developers and Infrastructure operators (TtD)
  - c. Trainers (TtT)
  
2. Build sustainable Training Infras : TeSS Training portal, e-learning platform, Best practices guidelines and Metrics, ...



# TeSS Portal

Aggregate, disseminate and coordinate information on ELIXIR's training activities/material and beyond

- 13 ELIXIR Nodes
  - 1023 training materials (126 FR)
  - 48 content providers:
  - 4,934 unique users (2017-18)
  - 308 upcoming events listed
- 
-  Bioschemas integration
  - Linked to other ELIXIR registries: bio.tools and
  - Events annotated with EDAM topics terms



Welcome to TeSS: ELIXIR's Training Portal  
Browsing, discovering and organizing life sciences training resources, aggregated from ELIXIR nodes and 3<sup>rd</sup> party providers.

Search training...

Events Materials Packages Workflows Providers

Discover the latest training events and news from ELIXIR nodes and 3<sup>rd</sup> party providers.

Browse the catalogue of training materials offered by ELIXIR nodes and 3<sup>rd</sup> party providers.

Create training packages by grouping resources that address your particular training topic.

Create training workflows to visualize learning steps and link to resources specific to your training needs.

Browse training providers to discover training resources they offer and follow links to their materials and courses.



# ELIXIR Train the Developer & the Researcher

## ELIXIR-FR 15 Training events (2017 - 2018)

- Elixir Train the Trainers (Pasteur Institute - 2018)
- Biocontainer Hackathon (OCT 2017) => [biocontainers.pro](http://biocontainers.pro)
- Organised the second session of [Workflow4metabolomics](#) workshop (Pasteur Institute)
- Organised an [ELIXIR Plant Genome Assembly and Annotation Workshop](#) (Montpellier). This Workshop is a pilot of Implementation Study WaaS (Workshop as a Services).
- Software Carpentry (unix, git and python) => April 2019

## Capacity building

- Involved in ELIXIR-EXCELERATE **capacity building effort in single-cell transcriptomics** (Single Cell-Finland May 27 2019), Variant analysis (with the Rare Diseases), HPC and the CODATA summer school on [Data Science](#)
- [BH2018](#): Hosted and Co-organized the first <https://2018.biohackathon-europe.org/> (Nov 12 – Nov 16 2018) aux Berges de Seine (South of Paris, France).
- [BH2019](#): Reconduct in Berges de Seine, Nov 18-22, 2019.
- <https://www.biohackathon-europe.org/>

## BIOHACKATHON 2018 PARIS

November 12th - 16th, Campus des berges de Seine



Strengthen the  
interactions with ELIXIR  
Platforms and  
Communities

EVERYBODY IS WELCOME!

REGISTER

LEARN MORE



- ~150 registrations
- 19 Countries

- 29 Hack projects, 11 topics

# {B} {H}

## BIOHACKATHON-EUROPE

PARIS 2019 November 18th - 22th

@BIOHACKATHONEU

Registration will be open on May 11th.

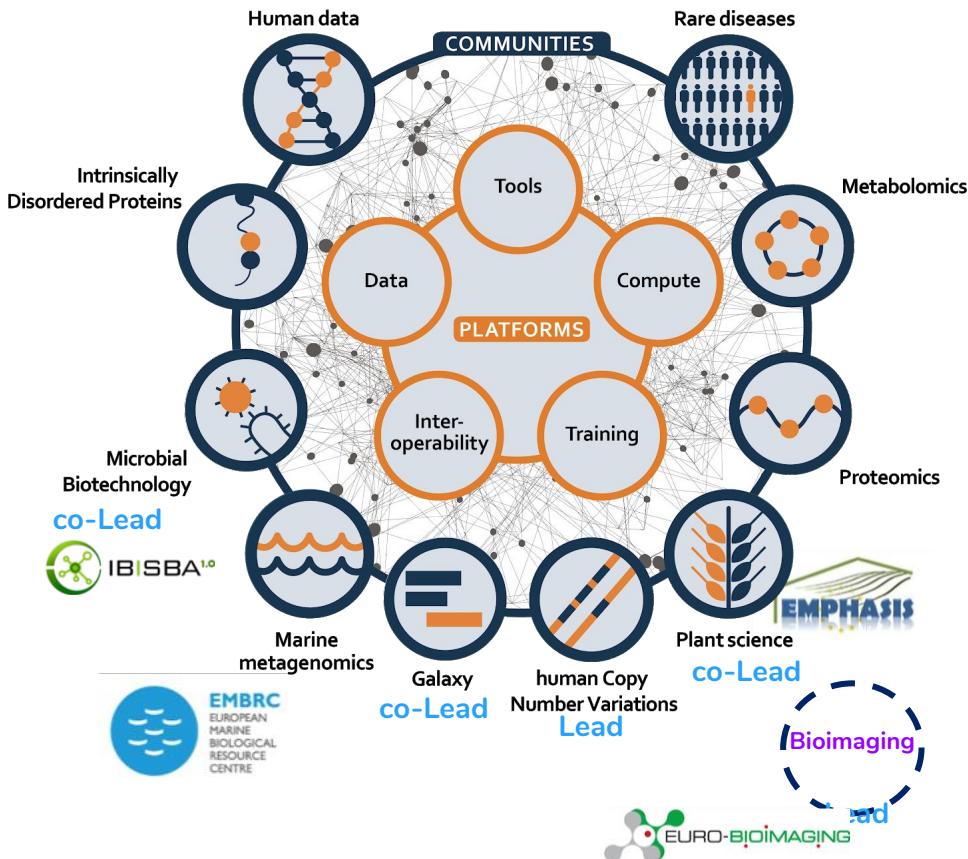
Call for proposals now open

LEARN MORE

SUBMIT A PROPOSAL

# EU integration with ELIXIR

<https://www.elixir-europe.org/>; 23 nodes



8 H2020 projects as leader / co-leader of WP



25 ongoing  
+ 11 completed  
**Implementation Studies**  
with other ELIXIR nodes

(co)-organisation of 6 important trainings/events of which SME 2017, Biohackathon 2018 (+ 2019) in Paris

**Service Delivery Plan**  
FR services eligible to the **ELIXIR catalogue**  
82 submissions  
**25 will be promoted**



# EU integration with ELIXIR



Acronym	Topics	Start	End	Total budget	FR budget
EXCELERATE (coord ELIXIR Hub)	Bioinformatics infrastructure	2015	2019	22 M€	1 M€
CORBEL					
INFRA-EOSC4 Life (coord ELIXIR Hub)	Providing an open collaborative space for digital biology in Europe. France is mainly involved through its coordination of ESFRI or ERICS in life science (BBMRI, EMBRC, ECRIN, ...)	2019	2023	23,8 M€	2,65 M€
INFRADEV / EMBRIC	European Marine Biological Research Infrastructure Cluster to promote the Blue Bioeconomy	2015	2019	9 M€	60 138 €
RIA / EUCANCan	A federated network of aligned and interoperable infrastructures for the homogeneous analysis, management and sharing of genomic oncology data for Personalized Medicine.	2019	2022	6 M€	171 250 €
CSA / GeneRes Bridge	Coordination Structuring Action for the Crop, Forest tree and animal genetic resources. Coordination of the WP about the information Link with the french infrastructure AgroBRC-RARé ( <a href="https://www6.inra.fr/agrobrc-rare/">https://www6.inra.fr/agrobrc-rare/</a> )	2019	2021	3 M€	80,275 € (out of 388 k€ for FR in total)
COST action Integrapé	Co-Coordination of the action about making the data of the grapevine community more open (FAIR).	2018	2022	NA	NA
H2020 INFRAEOSC-05 / EOSC-Pillar	(Grant Agreement preparation) Coordination and Harmonisation of National Initiatives, Infrastructures and Data services in Central and Western Europe	2018 (GA prep.)	2022	6,88 M€	38 700 €

Figure 11 The IFB contribution to EU-funded projects



The European Joint Programme on Rare Diseases (EJP RD) brings over 130 institutions from 35 countries and provide the opportunity to create a comprehensive, sustainable ecosystem allowing a virtuous circle between research, care and medical innovation. This **RD research** project should be improved to overcome fragmentation, leading to efficacious use of data and resources, faster scientific progress and competitiveness, and most importantly to **decrease unnecessary hardship and prolonged suffering of RD patients**. Total Budget : ~120 M

Coordinator: Daria Julkowska, INSERM, France

ELIXIR-CONVERGE 3 years project – From 1 Feb 2020

ELIXIR-CONVERGE is designed to deliver the major, complex and outstanding issues of building and harmonising national data management practices, access to human capital, impact assessments and national roadmap positioning (Total Budget: ~5M , ~130 K FR budget )

# Les correspondants ELIXIR-FR

## Heads of Nodes: *Elixir permanent working groups: Technical Coordinators*

Claudine Médigue et	Compute	C. Blanchet; O. Collin; J. Seiler	
	Interoperability	M-D. Desvignes, A. Gaignard	
	Tools	H. Ménager; O. Sallou	
	Training	V. Dominguez	
Jacques Van Helden	Data	D. Salgado; S. Ricard-Blum	
	<b>Communities</b>		
Deputy HoN:  Anne-Françoise Adam-Blondon	Marine Metagenomics	French coordinator	E. Corre
	Plant	Community coordinator	C. Pommier
	Human data	French coordinator	D. Salgado
	Rare diseases	French coordinator	M. Hanauer
	h-CNV	Community coordinator	C. Béroud
	Human cell atlas		
	Proteomic	French coordinator	C. Bruley
	Metabolomic	French coordinator	E. Thevenot
	Structural bioinformatics	French coordinator	P. Tuffery
	Galaxy	Community coordinator	G. Le Corguillé
	Microbial biotechnology	Community coordinator	V. Fromion

Mirrors ELIXIR's activities - Monthly meetings

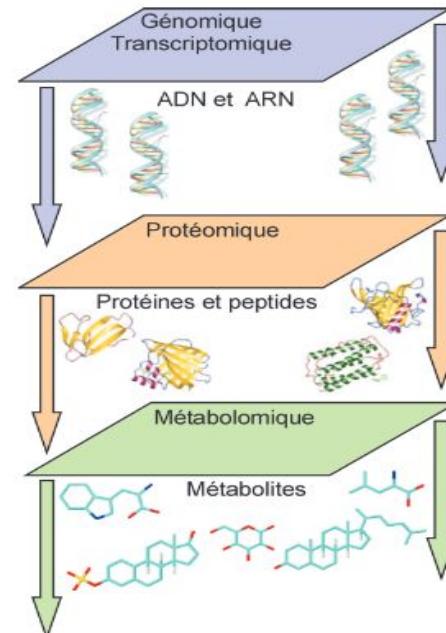
## Axe d'innovation – Bioinformatique Intégrative

# Towards integrative bioinformatics

- Life scientists are supported by **several national research infrastructures**
- **⚠ these are often focused on a specific biological scale or data type**
- Challenge: Promoting **data integration** and **data/tools sharing**

## 3 Actions

- **A2.1 Pilot projects** (2 years), in collaboration with national infrastructures and plans (**in progress**)
- **A2.2 Call for challenges** (3-4 years): lift technological and scientific locks. Develop flagship applications
- **A2.3 Catalyse interoperability** and **bioinformatics resource integration** (**in progress**)

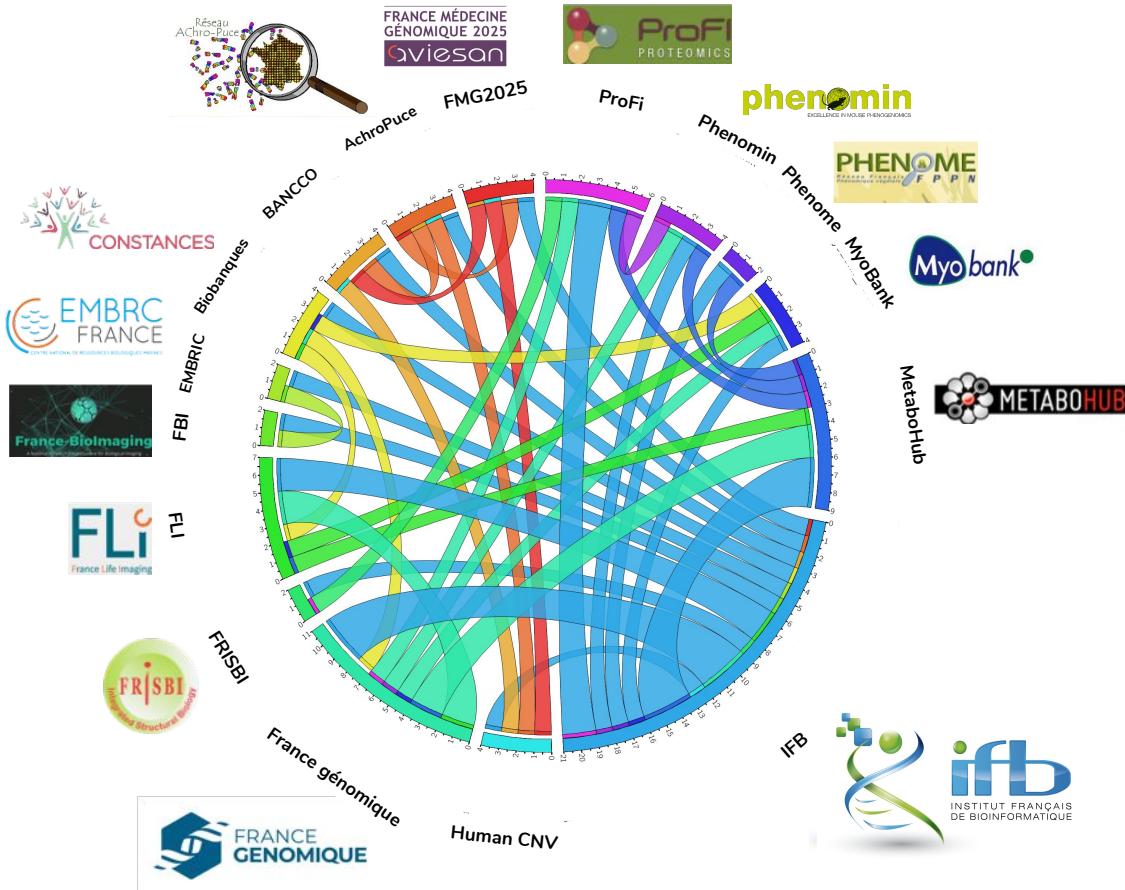


**France Médecine  
Génomique 2025**

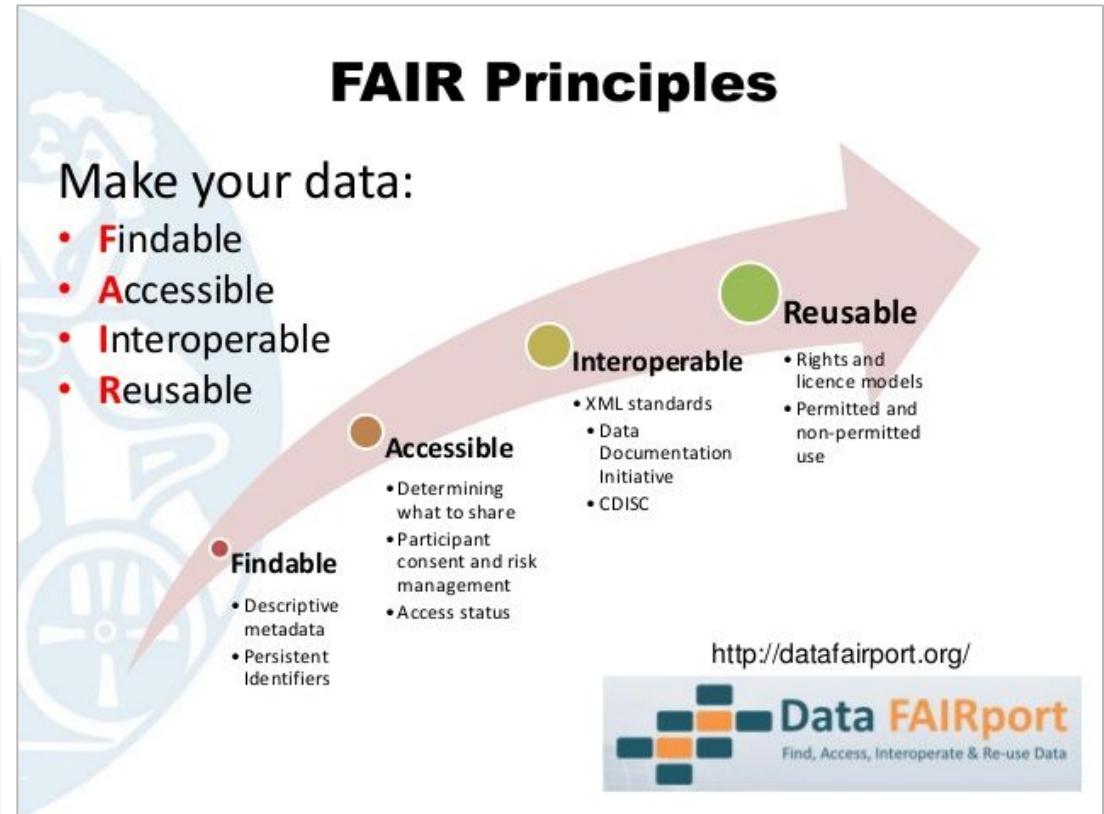
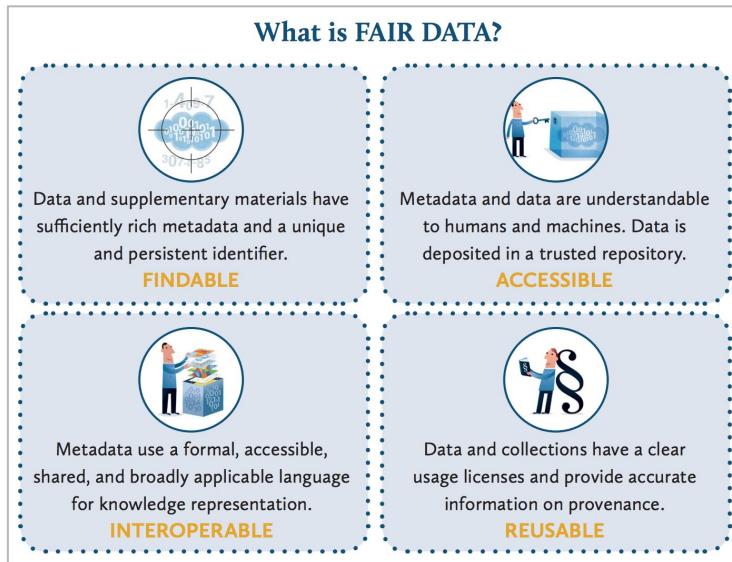


# Innovation : integrative bioinformatics

- Life scientists are supported by 22 national research infrastructures **promoting data integration and data/tools sharing** to enable a global view of complex biological systems.
- Action 1: Pilot-projects
  - Call criteria
    - IFB + at least 2 others national infrastructures;
    - well-defined study case enabling to quickly achieve tangible results;
    - cover national priorities (health, agronomy & environment).
  - Call results: **8 selected projects** (5 in the health domain) **gathering 17 national structures**
- Action 2: development of tools for Interoperability: implementing the **I** of the **FAIR** principles.



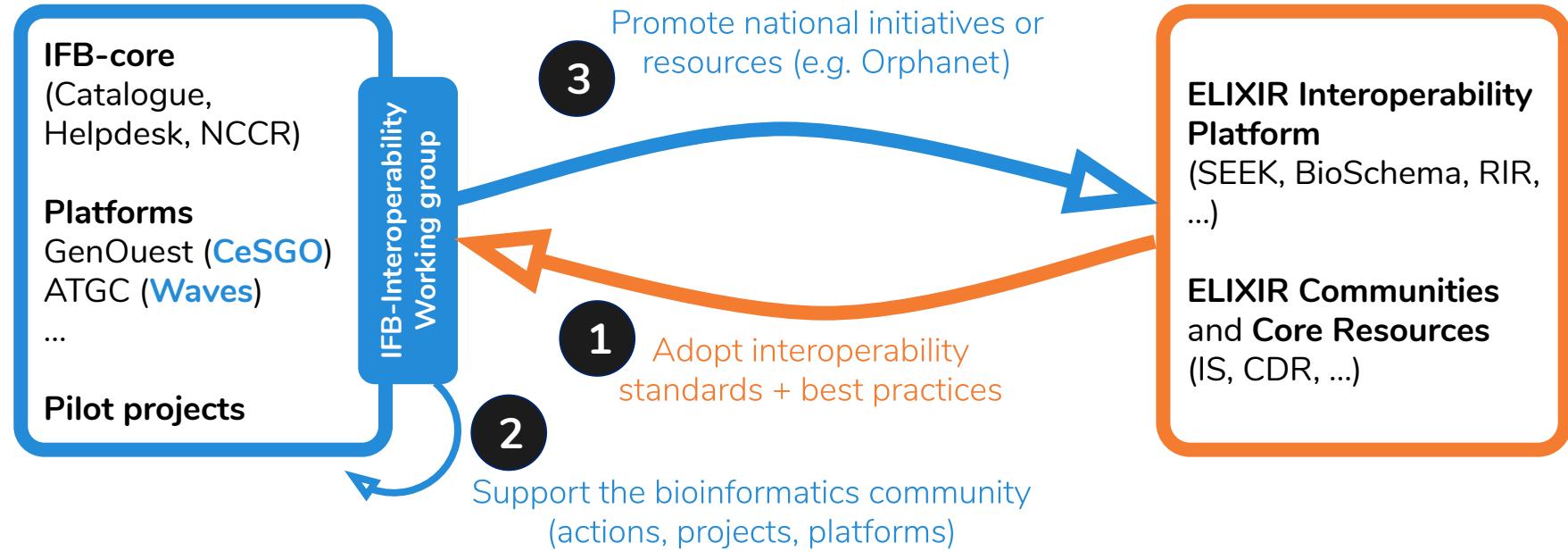
# FAIR Principles



- Wilkinson et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018.
- Source of the figures: slides from Gareth Knight (2015). [Preparing Data for Sharing: The FAIR Principles](#).

# Interoperability working group

- Interoperability: key issue for integrative bioinformatics, part of the “FAIR data” principles



# Bonnes pratiques

- Reproductibilité des analyses
  - Workflows: snakemake, nexflow, CWL, Galaxy
- Reproductibilité de l'environnement logiciel
  - Packaging (conda)
  - Virtualisation (fédération de clouds)
  - Containerisation (Docker, Singularity)
-

# *Formations*

# IFB training missions

- Web: <https://www.france-bioinformatique.fr/en/training>
- Increase bioinformatics capacity and competence in France and internationally
  - Help life-science and medical communities to make best use of bioinformatics resources (infrastructure, tools, databases)
  - Enable bioinformaticians to upgrade their skills and cope with rapid evolution of the domain
  - Design new trainings for new needs
- Promote coordination and collaboration between trainers (IFB platforms + ELIXIR)
- Rely on ELIXIR network and partners to enrich and complement the National training program
- Encourage best practices in training
  - pedagogical approaches
  - reusability of the material
  - feedback mechanisms
  - FAIR practices
- Target audience
  - Life-long training at all professional levels: scientists, engineers, technicians, PhD students and postdocs
  - Academia
  - Private sector

# How to ensure these missions ?

- Build up on the diversity of skills developed by all IFB platforms
  - Platform-specific training activities
  - Collaborations between multiple platforms (e.g. DU-Bli, EBAI, Galaxy, W4M)
- Make benefit of IFB-developed resources (NNCR, containers, training material)
- Develop new modes of trainings: videos, tutorials, MOOC, SPOC...
- Optimize the diffusion the training events
- Promote IFB training trough IFB website, TeSS, Elixir newsletter and so on
- Define and propose for all trainings a consistent set of Key Performance Indicators (KPIs) for the training and its environment (pedagogy, organisation, relevance, theory/practice balance, suitability of the hardware and software, ...)

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

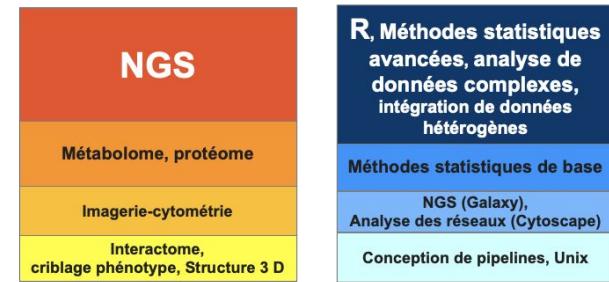
# Meeting user needs

- Jan 2019: collaboration with Inserm\* on a survey of the needs in bioinformatics of research teams
  - Targeted questions
  - Multiple-choice + open-ended answers
- Strong demand for training
  - Data production technologies
    - NGS
    - Other omics
    - Imaging
    - Interactome, phenotype, 3D structure
  - Advanced skills in
    - statistical methods
    - integrative bioinformatics
  - Environments:
    - R and Unix capacitation
    - User-friendly interfaces for omics (Galaxy, Cytoscape)
  - Formats: strong interest for Bring-Your-Own-Data
  - Duration: 1 week to 1 month
- This study is now extended to all the French research institutions



## User survey about expected contents

Technologies de production de données



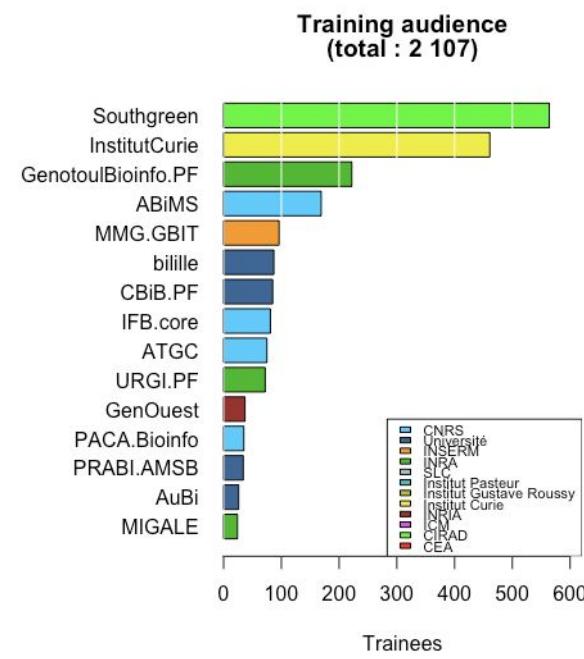
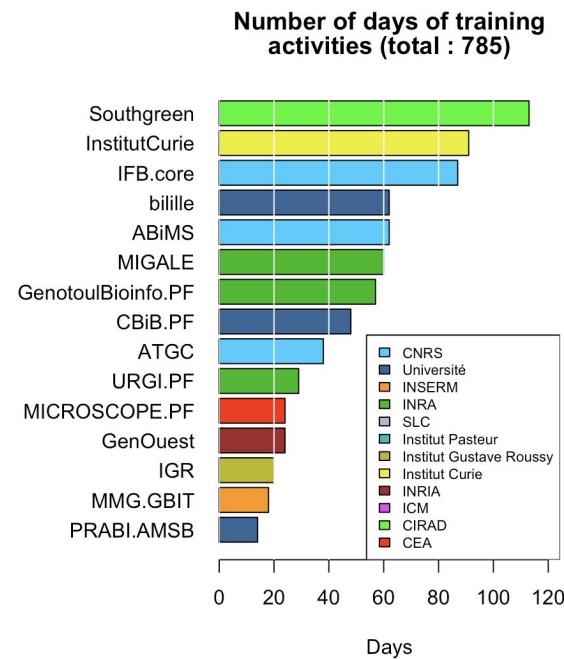
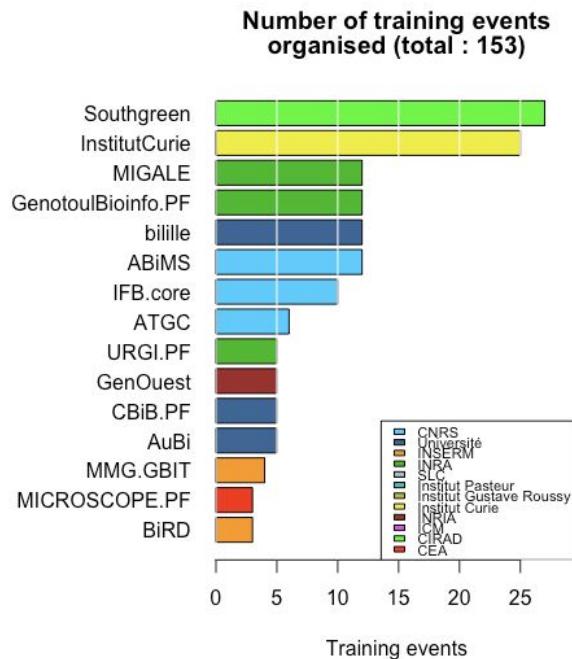
## User survey about duration / format



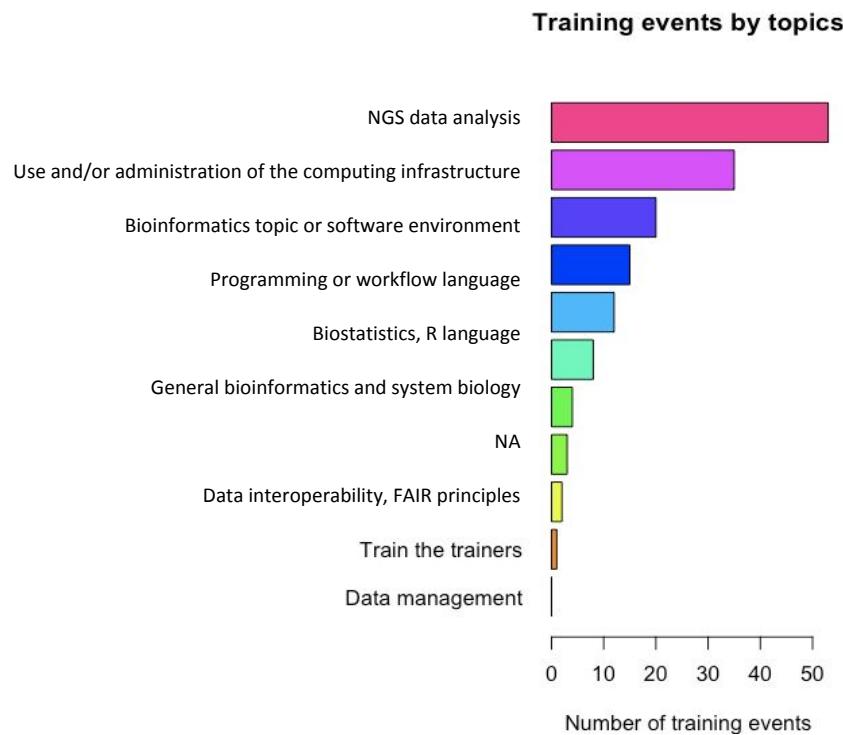
\* Inserm : French institute for health and medical research

# Impact of IFB training activities -2018

153 training events - 785 days of training - 2107 trainees



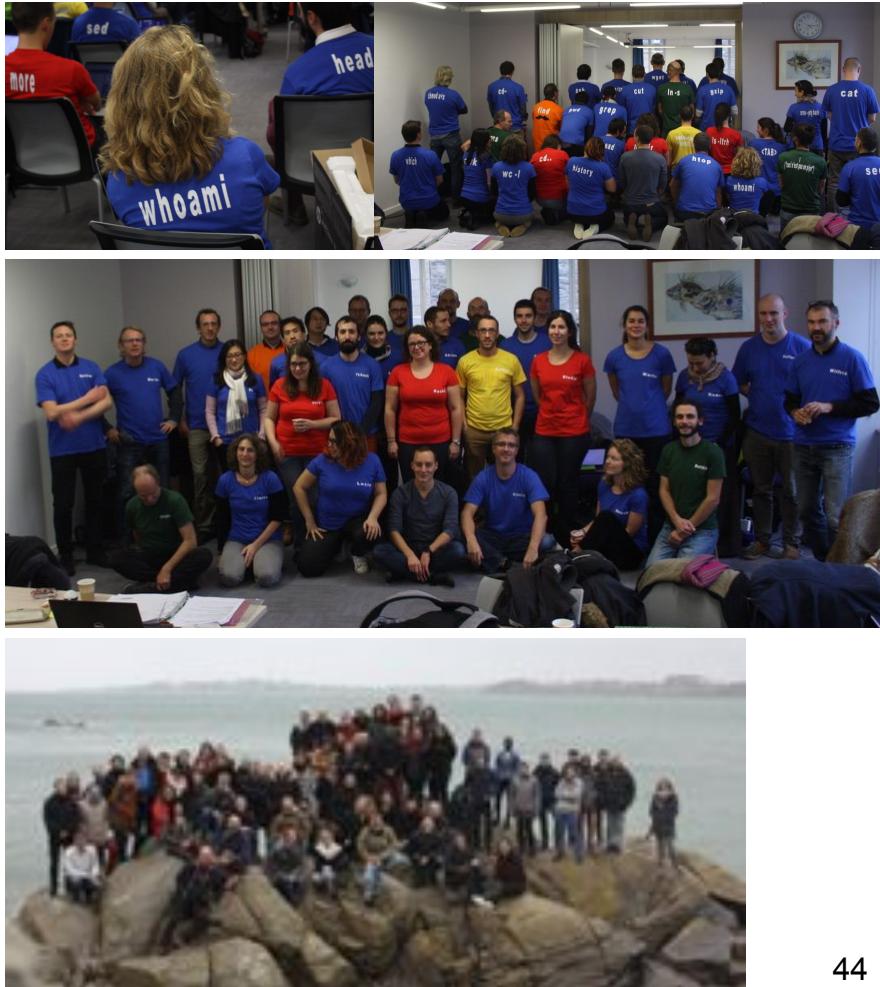
# Training topics handled by IFB platforms in 2017-2018



- Current IFB training focused mainly on NGS data analysis and IT technologies (access, administration)
- Recent emerging topics
  - Advanced use of IT technologies (HPC, virtualization, cloud computing, interoperability)
  - New technologies (long reads, metabolomics, single-cell, Hi-C,...)
  - Integrative Bioinformatics

# Aviesan-IFB school of bioinformatics (EBAI): NGS analysis skills for biologists

- Web:  
<https://www.france-bioinformatique.fr/fr/evenements/EBAI2018>
- Launched in 2013, 1 edition per year, 1 week in Roscoff
- BYOD with **customized tutorship**
- 40 trainees and 30 trainers (teachers + tutors + tech support)
- Evolution to suit biologist demand
  - Initially based on a dedicated Galaxy environment (2013-2016)
  - Since 2017, based on Linux and R
- High success
  - ~120 applications per year, for only 40 places
  - **Excellent satisfaction indicators**
- Accessibility
  - All **teaching material in public**
  - Conda enables trainees to **replicate the software environment** on local systems
- Side benefits
  - Regular **packaging of new tools** (Galaxy, Conda)
  - **Community building** among bioinformaticians and collaboration between IFB platforms (>20 platform and research teams collaborating each year on teaching, tutorship, support)



# Diplôme Universitaire en Bioinformatique Intégrative (DU-Bii)

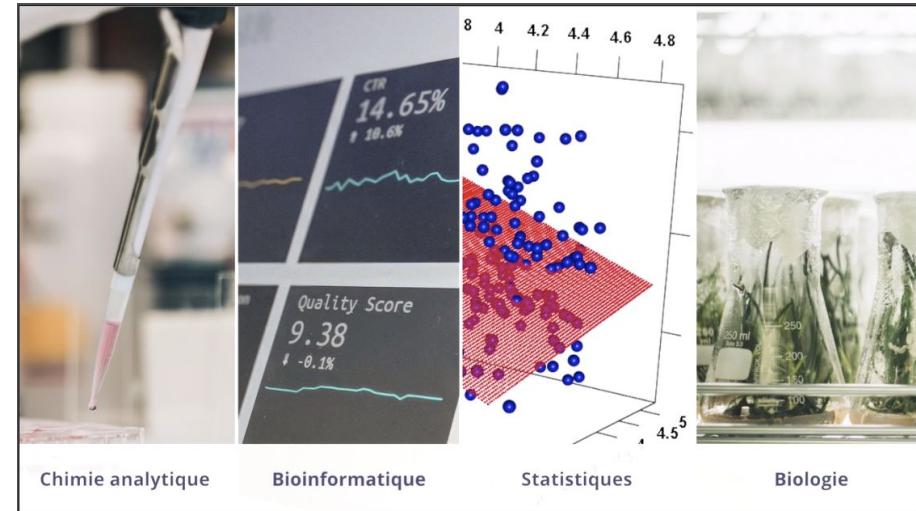
- Web: <https://www.france-bioinformatique.fr/fr/du-bii>
- New University Diploma in partnership with **Paris Diderot University**
- Answer to needs expressed during other training events + surveys in research institutions (INRA, Inserm)
- Audience: **biologists**
  - disposing of multi-omics data for integrative biology projects
  - searching for secondary skills or even professional evolution towards bioinformatics
- First session launched in February 2019: 58 demands, 15 trainees
- 4 weeks of theoretical and practical courses and 4 weeks of **tutored project in an IFB platform (BYOD)**
- Side benefits
  - Conda environments for each teaching module: **reusability, portability of the software environment**
  - Also addresses a **strong demand** from our own bioinformaticians
  - **Training material will be recycled**  
(videos for the Integrative bioinformatics module)

DU-Bii modules
Unix environment
Python programming
Statistics with R
Omics data production
Omics data analysis
Integrative bioinformatics
<b>Tutorship by one or several IFB platforms (BYOD project)</b>



# MOOC Metabolomics

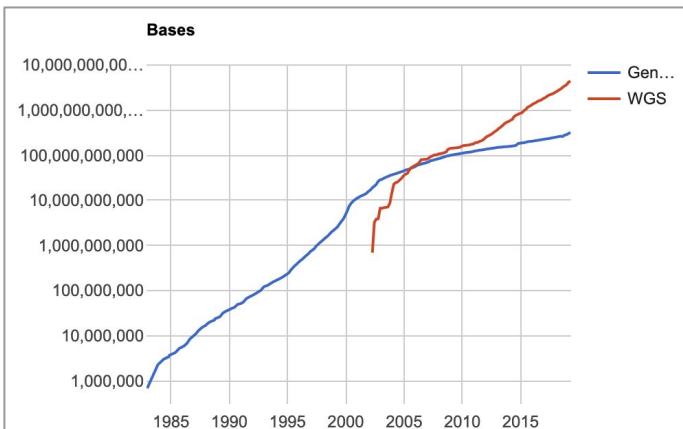
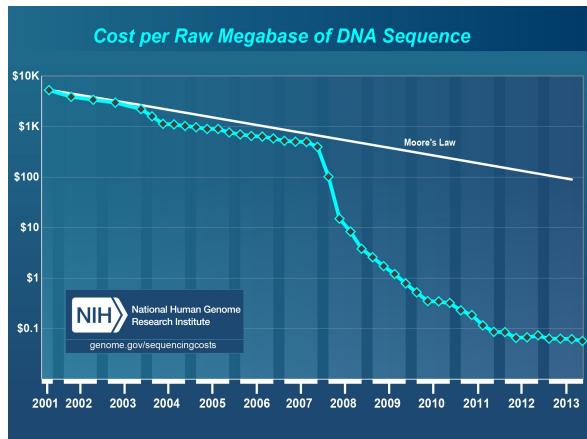
- Web: <https://tinyurl.com/mooc-metabolomique>
- Public : biologists, no prior requirement for bioinformatics
- Goals
  - Master the concepts, tools and methodologies to integrate a metabolomic experiment in a research project
- 20 contributors: teachers, researchers and tech engineers (Univ Rennes 1)
- Format: ~ 20 h / 5 weeks + weekly Quizz + case study
- Attendance (1st Edition, Oct 1 - Dec 5, 2018)
  - 3066 registrations from 77 countries (63% France, 7% Morocco)
  - 602 (20%) took answered at least 1 quiz
  - 214 followed the full training and got the certificate



## *Focus sur les problématiques de données*

# Les défis des données biologiques – La volumétrie n'est pas tout

- **Volumétrie**, en particulier pour données de
  - séquençage (pré- et post-NGS) et imagerie
- **Diversité**
  - Multi-omiques (génome, transcriptome, protéome, métabolome, métagénomique)
  - Images
  - Structures macromoléculaires
  - Réseaux biomoléculaires
- **Multiplicité des sources**
  - Données produites partout (centres de séquençage, d'imagerie, ...), généralement éloignées des centres de traitement
- **Intégration de données hétérogènes**
- **Connection à d'autres types de données**
  - Données de santé, phénotypes, écotypes, climat, ...
- **Complexité**
  - représentation des connaissances biologiques
- **Protection des données personnelles**
  - Couplage données omiques / santé
  - Anonymisation impossible (données de séquençage sont identifiantes, autres données rattachable par recouplements)
  - Consentement “éclairé” dans un domaine en pleine émergence ?



# Stockage, accès, réutilisabilité : types de besoins et ressources

- **Données chaudes : phase d'exploitation**
  - brutes
  - secondaires résultant des analyses
- **Donnée volatile:** pour le calcul, sans persistance
  - Cloud computing: durée de la VM
  - Centres régionaux: durée du calcul
- **Données froides "congelées": archivage**
  - Capacité à récupérer à long terme,
  - Sécurité, garantie légale (5 ans, 10 ans, ...)
  - Récupération pas immédiate (semaine)
- **Données froides accessibles: dépôts de données**
  - Généralement post-projet
  - Objectif: pouvoir réutiliser dans 100 ans
  - **Trouvabilité** : identifiants
  - **Accessibilité** : immédiate, sécurisée
  - **Interopérabilité** : identifiants, références croisées, formats standards, métadonnées, indexation, ....
  - Formats normalisés
  - **Ré-utilisabilité** : figer également l'environnement logiciel (containers)
- **Open Science**
  - Enjeu: partage des données dès la naissance du projet (avant publication)
  - Données chaudes-froides
- **Infrastructures de stockage et calcul**
  - Locales (labos, instituts)
  - Régionales (mésocentres, data centers)
  - Nationales (IDRIS, CC-IN2P3, TTGC, CINES, ...)
  - EPST (ex: [INRA data verse](#))
  - Coordination régions/pays: CINES
  - Européennes (EGI, EUDAT, PRACE → intégration dans EOSC)
  - Compagnies privées (nationales + GAFA)
- **Ce que fait l'IFB**
  - Stockage données chaudes
  - Exploitation
  - Interopérabilité
  - Plans de gestion ([maDMP4LS](#), collab. INIST)
  - Formation
- **Ce qui n'est pas du ressort de l'IFB**
  - Archivage
  - Dépôt
- **Questions**
  - La plupart des ressources disponibles sont orientées avant tout vers le calcul  
→ **infras pour les données ?**
  - Offre pour hébergement de **données "sensibles"** (données de santé) ?

# Traitement des données sensibles

- Données sensibles
  - Données de santé
  - Génomique personnelle
  - Autres types de données personnelles
  - Combinaisons
- Défis
  - Donner accès à ces données
  - Préserver la confidentialité des données personnelles
  - Que faire avec les résultats des analyses?  
? Données primaires (objet de l'étude), secondaires (incidentes)
- Cadre légal français
  - CNIL
  - ANSI
  - Certification pour l'hébergement de données de santé
- Etat de la situation en France
  - Certification HDS
    - Obtenue par des acteurs privés.
    - L'Inserm n'en fait pas partie !
    - L'IFB non plus
- Situation en Europe
  - Projets d'échanges de données:  
1MG, B1MG
  - Difficulté: comment concilier les protocoles techniques et la diversité des législations nationales  
?

- Aujourd’hui, quelles sont les solutions (régionales, nationales, européennes) disponibles pour
  - stockage
  - partage durant la vie du projet
  - archivage long-terme
  - dépôt et accessibilité
- En particulier, comment gérer les **données de santé** (cohortes) ou associées (génomique personnelle) ?
  - Cas d’étude : le projet-pilote HDH recourt à un géant du Web car c’est la seule solution HDS opérationnelle qu’ils ont trouvée à ce jour.
  - Pouvons-nous mettre en place des solutions académiques – nationales ou européennes – pour traiter les données santé ?
  - Distribution (par producteur/type de données) ou centralisation (agrégation) ?
  - Déplacer le logiciel vers la donnée plutôt que l’inverse
- A l’échelle européenne, même questionnement, même perplexité sur les solutions opérationnelles

# Projets - formations pour les données

- Accompagner la transition de l'artisanat à la production “industrielles” des données (nouveaux outils, nouveaux traitements, procédures, standards, modèles de représentation, interopérabilité)
- Développer une **culture des données** partagée entre
  - producteurs de données
  - fournisseurs de services (bioinformaticiens, infrastructures)
  - usagers
- Sujets à couvrir
  - Plan de Gestion des Données (PGD, DMP)
  - Echanges de données
  - Ontologies, Web sémantique
  - FAIR
  - Enjeux légaux (aspects juridiques, droit des données, RGPD, données sensibles, ...)
- Formats des formations
  - Formations courtes pour les usagers et pour les fournisseurs de services
  - Formations diplômantes (DU)
  - Applications aux données des usagers: tutorat BYOD, hackathons, ...
  - Promouvoir la formation aux données en formation initial

## *Articulation avec la recherche en bioinformatique*

# Complémentarités et synergies

- **IFB - INS2I**
- La biologie génère des problématiques fertiles pour la recherche en informatique.
  - Clustal
  - PhyML
  - Phylogeny.fr
- Promotion des approches issues de la recherche méthodologique
  - Packaging
  - Diffusion ressources bioinformatiques (bases de données et outils logiciels) auprès des communautés d'usagers
  - Déploiement de ressources sur serveurs IFB
  - Organisation d'un service autour des outils
- Accès à des réseaux et à des financements nationaux européens
- **IFB - GDR**
  - Actions communes dans la feuille de route IFB
    - JOBIM (collaboration SFBI - GDR BIM - IFB)
    - Contribution au catalogue des ressources françaises en bioinformatique, via la liste des UMRs et équipes de bioinfo
- Un large espace pour des actions communes
  - Promotion des bonnes pratiques de développement
    - standards internationaux
    - interopérabilité
    - perspectives: software management plan
  - Formation des chercheurs (progression en compétences)
    - écoles jeunes chercheurs
    - Formation des moins jeunes
    - Ecoles des 4 saisons

*En savoir plus sur l'IFB*

# Impact of IFB: some indicators

2018  
2017

## Concrete answers to user needs

Training events organised by IFB platforms  
153  
81

User accounts on Unix servers  
14516  
11723

Number of unique users for Orphanet  
12 057 375  
9 645 238

Projects with industries  
32  
-

Participants to IFB training events  
2073  
1100

Supported research projects  
872  
517

Number of unique users for IFB Web tools (without Orphanet)  
331 230  
-

## High user satisfaction

Percentage of satisfied + very satisfied users for followed training  
95%  
-

Percentage of satisfied + very satisfied users for delivered services on projects  
98%  
88%

## Collaborative projects and service delivery

International academic projects (involving foreign academic labs)  
229  
261

Money gathered by IFB platforms on EU and other international projects  
1,810 K€  
520 K€

## Publications

Publications signed by IFB platforms  
131  
129

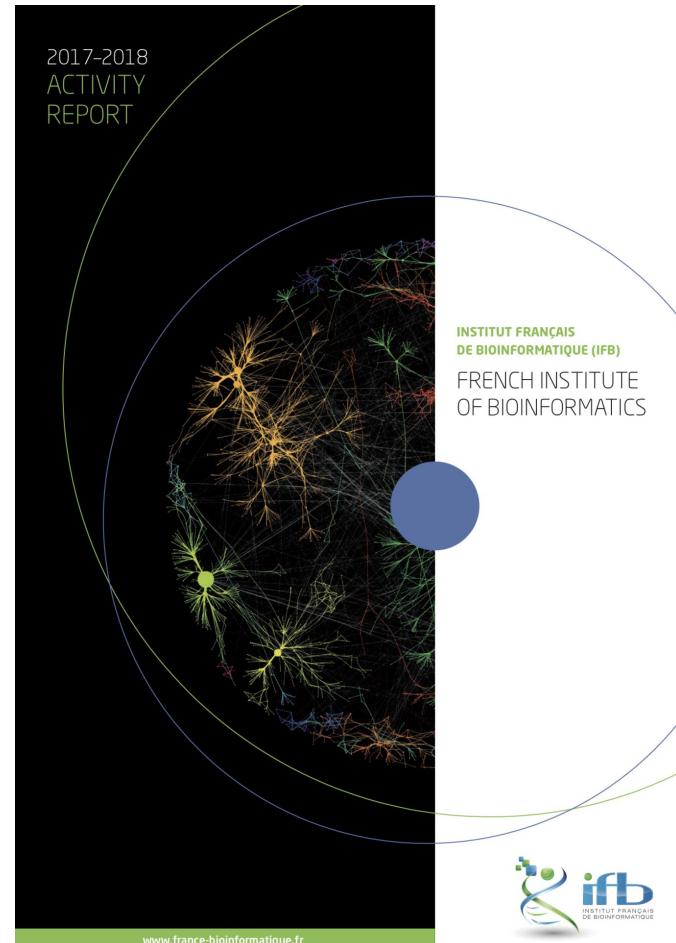
IFB platform as first or last author  
32  
21

IFB or platform acknowledged  
98  
115

Internal projects (for the hosting laboratory of the IFB platform)  
388  
245

# En savoir plus sur l'IFB

- Site Web
  - <https://www.france-bioinformatique.fr/>
- Rapport d'activités 2017-2018
  - <https://www.france-bioinformatique.fr/en/activity-report-2018>
- Nous contacter
  - [contact@groupes.france-bioinformatique.fr](mailto:contact@groupes.france-bioinformatique.fr)
- Journées IFB - ELIXIR-FR
  - <https://ifb-ag.sciencesconf.org/>
  - **Journée IFB (2019-11-27)**
    - Ouverte au public (sauf dernière session-)
    - AG
      - Bilan des actions 2018-2019
      - Perspectives
    - Après-midi: panorama des ressources françaises pour les données de biologie et de santé
  - **Journée ELIXIR-FR (2019-11-28)**
    - Positionnement stratégique de la France dans ELIXIR
    - Après-midi: panorama des projets européens pour les données de biologie et de santé





## *Diapos supplémentaires*

## *Bilan et perspectives de la contribution de la France à ELIXIR*

*Strengths*

*Weaknesses*

*Opportunities*

*Threats*



**Data deposition:**  
ENA, EGA, PDBe, EuropePMC, ...



**Data management:**  
Genome annotation  
Data management plans



**Added value data:**  
UniProt, Ensembl, OrphaNet, ...



**Data Interoperability:**  
Standards, Identifiers, Ontologies



**Compute:**  
Secure data transfer, cloud computing,



**Bioinformatics tools:**  
Bio.tools



**Industry:**  
Innovation and SME programme  
Bespoke collaborations



**Training:**  
TeSS, Data Carpentry, eLearning

# Points forts de la contribution française (Strengths WOT)

## Ce qui fonctionne bien

1. **Mise en réseau des expertises** mobilisables pour des actions en France (formations,, ...)
2. Intégration dans des projets de recherche européens, pour l'apport en bonnes pratiques en science reproductive (Data Management Plan, workflows, ...)
3. Mutualisation des efforts pour générer des **ressources logicielles déployables** (Bioconda, Galaxy, Biocontainers).
4. Incitation collective à concevoir des **solutions logicielles plus durables** (d'emblée conçues pour être réutilisables, et testées par partenaires des autres noeuds).
5. Engagement dans des **projets d'envergure** où la France ne se lancerait peut-être pas seule.
6. **Valorisation des ressources bioinformatiques développées en France** (Service Delivery Plan, Core Data Resources).
7. FR très **moteur dans la dynamique “communautés”** d'ELIXIR

## Ce qui va mieux qu'avant

1. Mise en place d'un système pour discuter de façon concertée d'une feuille de route, avec des **mécanismes de prise de décision** impliquant activement 23 partenaires.
2. En cours: **rééquilibrage** des poids respectifs des plateformes ELIXIR, du Hub et des noeuds nationaux.

# Faiblesses de la contribution française (S Weaknesses OT)

## Ce qui ne fonctionne pas bien

1. **Retour financier insuffisant** au regard des efforts de la communauté ELIXIR-FR
2. **Difficulté à comptabiliser l'impact réel** de la participation française à ELIXIR
  - ❑ Le retour financier sous-estime l'apport à la biologie et bioinformatique nationale
  - ❑ Projet OCDE **RI-path** pour estimer l'impact des infrastructures, où ELIXIR est l'infra la plus complexe.
3. ELIXIR-FR l'un des plus gros noeud d'ELIXIR avec de multiples centres d'intérêt -> **difficile de se positionner**
  - ❑ Choix délibéré de couvrir la diversité des activités IFB plutôt que de favoriser une ou deux thématiques
4. Difficulté à suivre la **multiplicité des activités et des informations**
  - ❑ A faire: **recruter une personne à plein temps** habitués à gérer des projets internationaux (aspects techniques et gestion)
5. Manque de présence Française dans la **coordination d'ELIXIR** (plateformes)
  - ❑ Participation permettrait indiscutablement de bien positionner le noeud dans les projets
  - ❑ Discussion entamée avec le hub pour renforcer la participation FR à la coordination
6. **“Stop and go”**: incertitude persistante concernant la poursuite de la contribution française ne facilite pas l'engagement
  - ❑ Absence au démarrage -> difficulté à s'insérer a posteriori dans EXCELERATE
  - ❑ Incertitude actuelle -> difficulté à s'engager dans la feuille de route 2019-2023 en particulier se porter candidat à la coordination de plateformes ELIXIR
7. Difficulté d'**incorporer des équipes extra-IFB** dans des projets via ELIXIR

# Opportunités (SW Opportunities T)

## Ce qui pourrait devenir une force dans le futur

1. Axe d'innovation de la feuille de route IFB (**bioinformatique intégrative**) en phase avec
  - Le nouveau plan de travail d'ELIXIR 2019-2023 : renforcement des collaborations entre les communautés (infrastructures nationales/ESFRI)
  - Les appels européens inter-ESFRI
2. Augmenter l'impact (visibilité, utilisation, service rendu) et la "durabilité" des ressources FR pour l'ensemble des usagers.
3. Utiliser notre positionnement actuel dans les communautés ELIXIR pour
  - interagir de façon plus étroite avec les **plateformes technologiques ELIXIR**
  - augmenter l'impact des activités ELIXIR sur la **communauté scientifique FR**
4. Contribuer à et bénéficier de la stratégie internationale d'ELIXIR.

- **FRELIXIT (sortie d'ELIXIR)**
  - France écartée des constructions européennes alors que l'Europe met les infrastructures au cœur de sa stratégie EOSC
  - France écartée de la stratégie internationale d'ELIXIR (Commission EU, G7, ONU, OCDE)
  - Fragilisation des ressources bioinformatiques françaises dans l'écosystème Européen (difficulté à les défendre dans des projets européens)
- Faiblesse du financement des activités propres du noeud français: grosse perte de vitesse de la communauté française dans le domaine de la bioinformatique

## Ressources européennes

- <https://www.egi.eu/>
- Ressources de calcul et stockage distribuée (grille)
- Environnements logiciels prêts à l'usage via machines virtuelles (cloud computing)

## EGI Services

EGI delivers advanced computing services to support scientists, multinational projects and research infrastructures.

The EGI Services are provided by EGI's **federated cloud providers** and **data centres**.  
The services can be requested by everyone involved in academic research and businesses via the **EGI Marketplace**.

Examples of how researchers use our services can be found in the **EGI Use Cases publication**.  
See also examples of how **EGI supports research infrastructures**.

INTERNAL SERVICES

THE EGI MARKETPLACE

### Compute



#### Cloud Compute

Run virtual machines on demand with complete control over computing resources



#### Cloud Container Compute BETA

Run Docker containers in a lightweight virtualised environment



#### High-Throughput Compute

Execute thousands of computational tasks to analyse large datasets



#### Workload Manager BETA

Manage computing workloads in an efficient way

### Security



#### Check-in BETA

Login with your own credentials

### Applications



#### Applications on Demand BETA

Use online applications for your data & compute intensive research



#### Notebooks BETA

Create interactive documents with live code, visualisations and text

### Storage and Data



#### Online Storage

Store, share and access your files and their metadata on a global scale



#### Archive Storage

Back-up your data for the long term and future use in a secure environment



#### Data Transfer

Transfer large sets of data from one place to another

### Training



#### FitSM Training

Learn how to manage IT services with a pragmatic and lightweight standard



#### ISO 27001 Training

Learn how to manage and secure information assets



#### Training Infrastructure

Dedicated computing and storage for training and education

- <http://www.prace-ri.eu/>
- Partnership for Advanced Computing in Europe
- 



The image shows the homepage of the PRACE website. At the top, there is a blue header bar with the PRACE logo (a circle of stars) and the text "PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE". Below the header is a navigation menu on the left with links to About PRACE, HPC Access, Industry, PRACE Projects and Outcomes, Training and Documentation, Events, Media, Outreach, and Contact. The main content area features a large graphic of a megaphone with sound waves, the text "Apply now!", and information about the "PRACE 20th Call for Proposals for Project Access". It also includes sections for "Call announcements", "Job vacancies", "WHITEPAPERS", "Browse projects", "Subscribe to PRACE", "CODE VAULT", "PRACE MOOC", "Outreach to Universities", and "International HPC Summer School". At the bottom, there is a "LATEST NEWS" section with a "PRACE news" card and a "PRACE Ada Lovelace Award 2020—Call for Nominations is open" card.

PRACE  
PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE

About PRACE  
HPC Access  
Industry  
PRACE Projects and Outcomes  
Training and Documentation  
Events  
Media  
Outreach  
Contact

PRACE 20th Call for Proposals for Project Access  
Open from 20 September to 29 October 2019  
Special Industry Track

Call announcements  
Job vacancies  
WHITEPAPERS  
Browse projects  
Subscribe to PRACE  
CODE VAULT  
PRACE MOOC  
Outreach to Universities  
International HPC Summer School

LATEST NEWS

PRACE news  
Thursday 26 September 2019  
CONTINUE

PRACE Ada Lovelace Award 2020—Call for Nominations is open  
Thursday 26 September 2019

- <https://eudat.eu/catalogue>
- Collaborative Data Infrastructure
- 

Research Data Services, Expertise & Technology Solutions

The screenshot shows the EUDAT website homepage. At the top is a blue header bar with the EUDAT logo and social media links. Below it is a white header with the EUDAT logo and navigation links for SERVICES & SUPPORT, USE CASES, DATA MANAGEMENT, SYNERGIES & POLICY, EVENTS, NEWS & PUBLICATIONS, and GETTING INVOLVED. A teal button on the right says "EUDAT Collaborative Data Infrastructure". The main content area features a banner for the "PRACE Collaborative EUDAT Data Infrastructure SUMMER SCHOOL" from September 23-27, 2019, at ICTP Trieste, Italy. The banner includes the PRACE logo, the EUDAT logo, and the ICTP logo. To the right of the banner is a photograph of a large, ornate building in Trieste. Below the banner are several service cards: "EUDAT Service Catalogue" (red card with a document icon), "I need solutions for ..." (orange card with a bulleted list: Data Access & Reuse, Data Preservation, Data processing and analysis, Sensitive Data Management), "Getting Involved" (dark blue card with a bulleted list: Use EUDAT Research Data Management services, Integrate & deploy EUDAT services for my community, Become an EUDAT service provider), "Policy" (light blue card with a small text blurb about EUDAT's contribution to research data policy), and "Use Cases" (yellow card with a small text blurb about discovering domain-specific and cross-disciplinary research communities).

EUDAT

SERVICES & SUPPORT ▾ USE CASES DATA MANAGEMENT ▾ SYNERGIES & POLICY ▾ EVENTS ▾ NEWS & PUBLICATIONS ▾ GETTING INVOLVED ▾

PRACE Collaborative EUDAT Data Infrastructure

SUMMER SCHOOL

Managing scientific data from analysis to long term archiving

23-27 September 2019, ICTP Trieste, Italy

Hosted by

ICTP International Centre for Theoretical Physics

EUDAT Service Catalogue

I need solutions for ...

- Data Access & Reuse
- Data Preservation
- Data processing and analysis
- Sensitive Data Management

Getting Involved

- Use EUDAT Research Data Management services
- Integrate & deploy EUDAT services for my community
- Become an EUDAT service provider

Policy

Find out how EUDAT is contributing to Research Data policy including the European Open Science Cloud, Open

Use Cases

Click here to discover which domain specific and cross-disciplinary research communities and organisations

# EOSC – European Open Science Cloud

- Vient de démarrer
- Construction progressive
- Regroupe réseaux précédents
  - ELIXIR
  - PRACE
  - EGI
  - ECRIN
  - EMBR-C
  - BBMRI (biobanques)
  - ...
- Chantiers actuels
  - Construction de la gouvernance
  - Interopérabilité sémantique entre données générées / gérées par différentes ESFRIs
  - Intégration de données biologiques, climatiques, environnement, ...

The screenshot shows the eosc-portal.eu website. At the top, there's a navigation bar with links for 'About', 'Governance', 'Services & Resources', 'Policy', 'EOSC in Practice', 'Media', 'For providers', 'My Services', 'Profile', 'Contact Us', and 'Login'. Below the navigation is the 'EUROPEAN OPEN SCIENCE CLOUD' logo. The main content area features the 'EOSCsecretariat.eu' logo, the 'RDA RESEARCH DATA ALLIANCE' logo, and text about the 'The International Research Data Community contributing to EOSC'. It mentions the event 'This EOSC Event at the RDA Helsinki Plenary is an integral part of the EOSC Stakeholder Forum' on '22nd October 2019, 9:00 - 17:00 Helsinki, Finland' and a 'Register Now!' button. To the right is a large image of the Helsinki Cathedral. Below this, there's a section titled 'ACCESS EOSC SERVICES & RESOURCES' with icons for Networking, Compute, Storage, Sharing & Discovery, Data Management, Processing & Analysis, Security & Operations, and Training & Support.