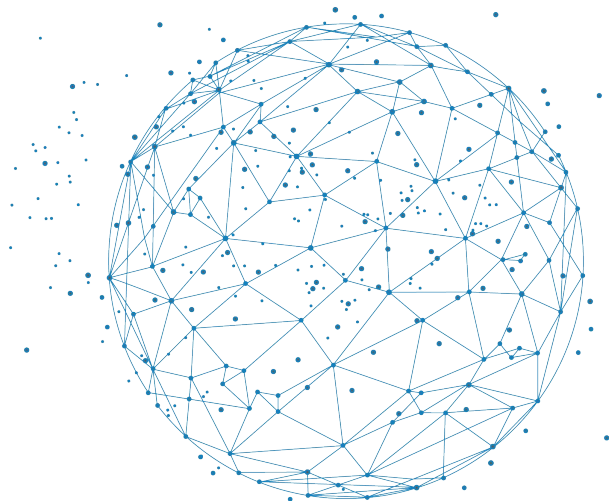




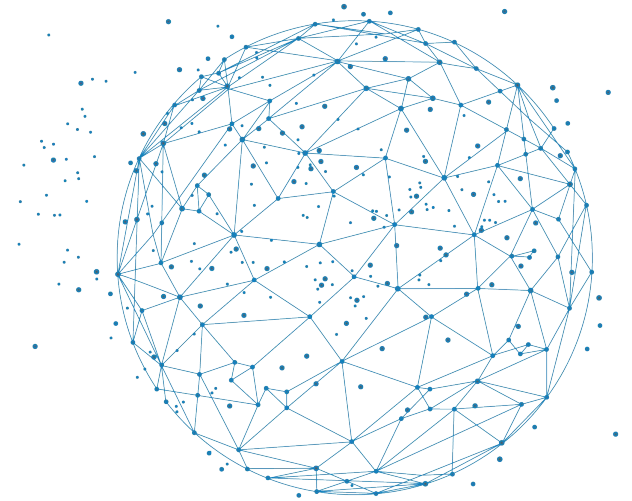
2024 - Fréjus



École thématique en bioinformatique Intégrative

Hélène Chiapello, Olivier Sand
& Lucie Khamvongsa-Charbonnier

Introduction





Buts

- **Intégrer différents jeux de données**, qui proviennent de différents niveaux d'analyse dans la cellule : (meta)génomés, (meta)transcriptomes, (meta)protéomes, métabolomes, réseaux d'interactions...
- **Extraire les informations pertinentes** à partir de ces différents jeux de données, pour ensuite les interpréter en utilisant des modèles intégratifs

> Un domaine de recherche en bioinfo

> L'intégration de données nécessite de maîtriser **des approches bioinformatiques qui résultent de différents domaines**

- Analyses multivariées
- Réseaux
- Web sémantique

Une **forte demande** de formation en bioinformatique intégrative

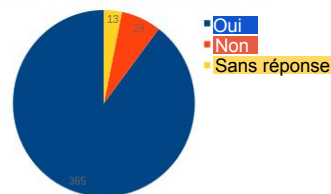
- pour les biologistes et les bioinformaticiens
- pour tous les métiers et niveaux
- dans différents domaines d'application : biologie, santé, agronomie, ...

Stratégie :

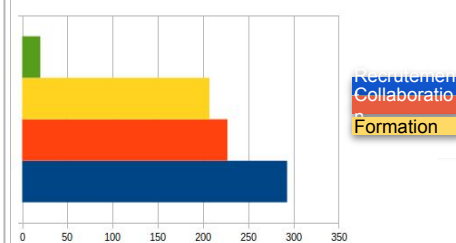
- former en priorité
 - des bioinformaticiens/biostatisticiens
 - des futurs formateurs
- mutualiser des ressources pédagogiques

Enquête IFB 2019 auprès des équipes/unités en France (N=407)

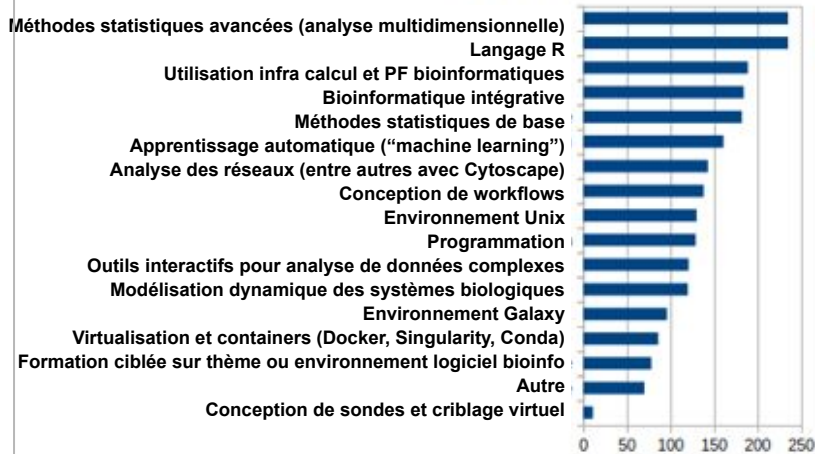
Votre unité / équipe éprouve-t-elle des besoins en bioinformatique ?



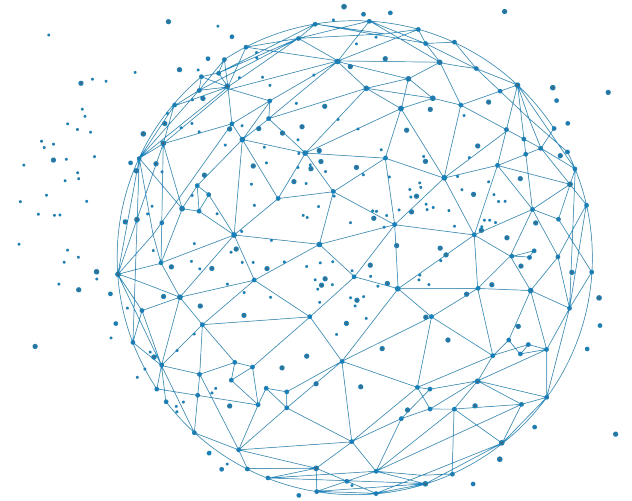
Comment couvrir ces besoins ?



Vos besoins en matière de formations à l'analyse bioinformatique



Présentation rapide de l'IFB



IFB – une fédération d'équipes et de plateformes

Missions

- Services
- Vision stratégique
- Représentation internationale (ELIXIR...)

36 plateformes et équipes associées

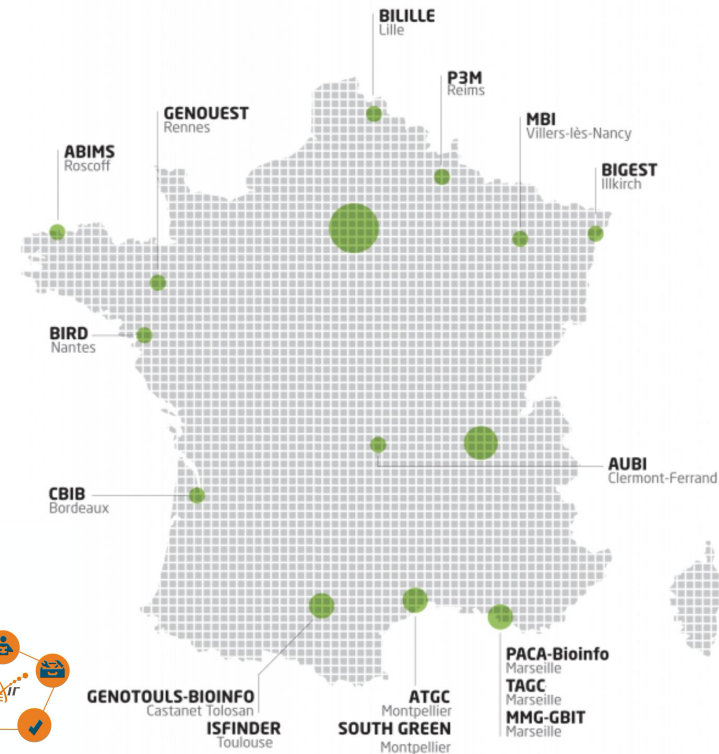
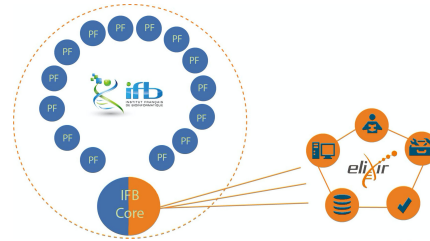
- En lien avec de nombreux instituts et universités

De nombreux services

- Calcul
- Stockage
- Formations
- Accompagnement de projets

IFB-core (UMS 3601)

~400 experts
(~200 ETPs)



RÉGION PARISIENNE

EBIO Paris
INSTITUT CURIE Paris
IGR Villejuif
MICROSCOPE Evry
MIGALE Jouy-en-Josas

C3BI

Paris
RPBS Paris
URGI Versailles
ORPHANET Paris
ICONICS Paris
IFB CORE Evry

RÉGION LYONNAISE

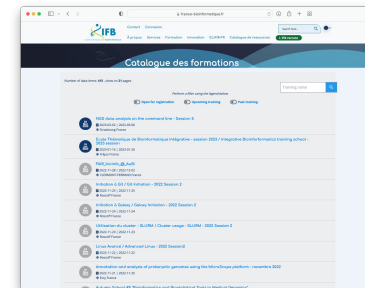
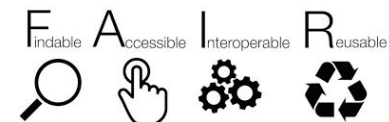
INCA-SLC Lyon
PRABI-HCL Lyon
PRABI-AMSB Villeurbanne
PRABI-Lyon-Grenoble Villeurbanne
PRABI-Lyon-Gerland Lyon



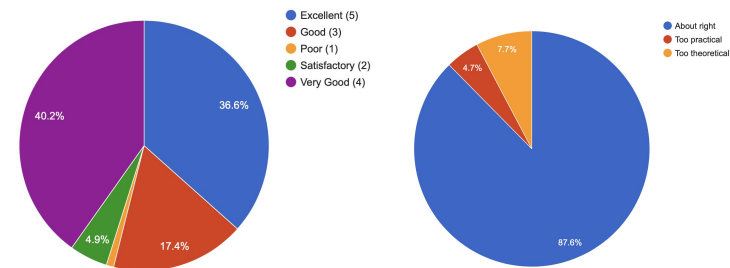
- Des formations régionales, nationales et internationales
- Un catalogue national des formations
- Un cycle de formation autour des principes FAIR en bioinformatique et pour la gestion des données
- Des actions de promotion
 - des bonnes pratiques de formation
 - de la mutualisation de matériel pédagogique (supports, workflows, jeux de données)
- Des actions ciblées pour les formateurs
 - ex : LMS moodle, GT e-formation

	2019	2020	2021
Training events	138	127	176
International	7	16	
Europe/Elixir	3	3	
National	53	76	
Regional	60	19	
University training	35	18	
Professional training	91	207	
Average satisfaction rate	86%	79%	
Learners	2169	2326	4458
Trainers	NA	272	379

**Science
Ouvverte et
principes FAIR**



- ELIXIR-FR Training Coordinator (O. Sand) & deputy (H. Chiapello)
- Adoption de BioSchemas (scraping par TeSS)
- Meilleures annotations des formations dans TeSS
- Curation de TeSS
- FAIR Training Focus Group
(<https://elixir-fair-training.github.io/FAIR-training-handbook>)
- Formation en collaboration avec d'autres noeuds
(<https://fpsom.github.io/2021-06-ml-elixir-fr/>, IA en prep)
- e-Learning WG
- Training Metrics Database (<https://tmd.elixir-hpc.si/user/login>)
- Learning Paths FG



L'infrastructure de calcul de stockage (ou NNCR)

- 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 les redondances
 - Gestion collective des problèmes
 - Partage d'expérience
 - Procédures robustes et transposables
 - Forte émulation des ingénieurs
 - Rompre la solitude de l'ingénieur info dans son labo bio



Hardware

- 6 SLURM Clusters
- 16,900 vCPU
- 92 TB RAM
- 10.5 PB
- 10 GPU cards
- 4935 yearCPU in 2020

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

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

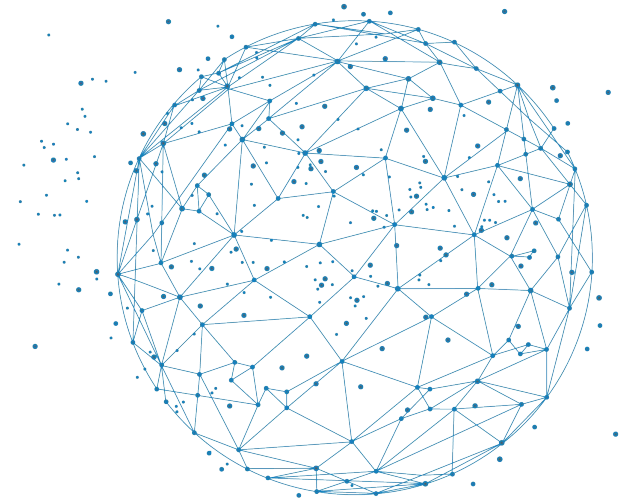
Plateforme	Localisation	Calcul (#CPU HT*)	Stockage (#TB)	RAM
ifb-core-cloud	Lyon (CC-IN2P3)	3936	408	20 408
AuBi	Clermont-Ferrand	200	20	800
GenOuest	Rennes	600	350	2600
PRABI	Lyon	696	144	3300
BiRD	Nantes	860	150	2 500
BiGEst	Strasbourg	1024	500	4 000
BILILLE	Lille	192	0	768
CBP-PSMN	Lyon	768	24	1920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

Plateforme	Location	Compute (#CPU HT*)	Storage (#TB)	RAM (#GB)	RAM/core (#GB)	RAM/core (#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



Démo <https://ondemand.cluster.france-bioinformatique.fr/>

La Bioinformatique Intégrative à l'IFB



Objectifs


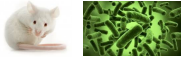















- Développement d'outils et de services permettant d'intégrer des données hétérogènes à haut débit
- Démontrer la pertinence des approches en les appliquant à des **cas d'utilisation spécifiques** reposant sur des données produites par **au moins deux** infrastructures de recherche nationales, des installations de modèles animaux, des cohortes humaines ou des plans nationaux
- En phase avec les **priorités nationales** (santé, environnement, biotechnologies, sécurité alimentaire)
- Obtenir des résultats à relativement **court terme** (18 à 24 mois)

Faits marquants/réalisations

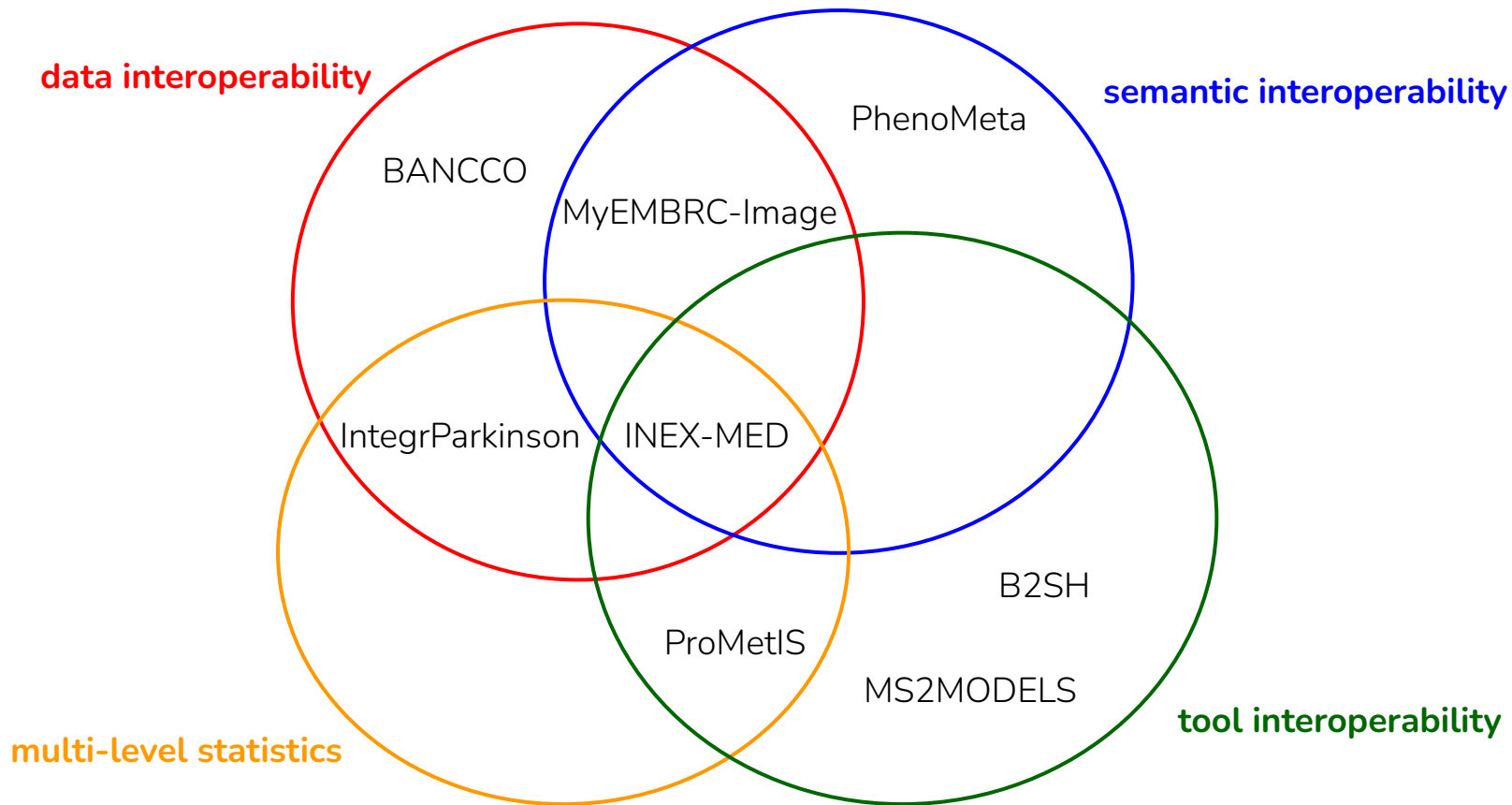
- Sélection de 5 projets bénéficiants chacun d'un CDD sur la durée du projet
- Collaborations entre plusieurs infrastructures nationales de recherche (IFB, France Génomique, MetaboHub, ProFI, FRISBI, FLI, Phenome/EMPHASIS) et deux installations modèles d'animaux et de cohortes humaines (Phenomin, BioBanques)
- Constitution de datasets FAIR
- Packages R, applications web, intégration Galaxy, application Shiny
- Présentations dans des conférences nationales et internationales (posters, démos, exposés)
- Publications dans Nature Communication, J Proteome Res, NAR





	Metabolomics MetaboHub	Proteomics ProFI	Genomics France Genomique	Images FLI...	Structure FRISBI	Phenotypes Phenomin/Phenome	Clinic/ Environment
ProMetIS							
PhenoMeta							
MS2MODELS							
INEX-MED							
IntegrParkinson							

+ 3 projets supportés par d'autres actions IFB : My-EMBRC-Image (*infrastructure sharing*), B2SH (*interoperability*) and BANCCO (*databases*)





ProMetIS : computing/storage resources

PhenoMeta : ELIXIR

MS2MODELS : computing/storage resources

INEX-MED : computing/storage resources

IntegrParkinson : computing/storage resources

B2SH : training, catalogue, consulting/orientation desk

MyEMBRC-Image : computing/storage resources, sharing of services with other national research support infrastructures, ELIXIR

BANCCO : support to databases, ELIXIR



Coordination : H. Chiapello & J. van Helden (IFB), P. Poulain & B. Cosson (U. Paris Cité)

3 éditions (2019 à 2021) et 54 apprenants, tutorat sur PF IFB

Objectifs pédagogiques

- Former de futurs bioinformaticiens
- Proposer un module innovant dédié à la **bioinformatique intégrative**
- Besoins fort en formation identifiés au sein des **bioinformaticiens IFB eux-mêmes**

Quelles étaient vos motivations spécifiques ?



Acquérir des compétences générales en informatique et statistique	16
Acquérir des compétences générales en analyse de données à haut débit	22
Acquérir des compétences spécifiques en bioinformatique intégrative	23
Apprendre à implémenter des logiciels utilisables par d'autres chercheurs	6
Apprendre à maîtriser la ligne de commande Unix	11
Acquérir des compétences de base en programmation (Python, R)	21
Communiquer avec les bioinformaticiens	11
Me réorienter vers une activité professionnelle en bioinformatique	13
Autres	1

A quel type de données comptiez vous initialement appliquer ces compétences ?



Données produites par mes travaux de recherche	16
Données produites par les collègues de mon équipe	11
Autre	5





INSTITUT FRANÇAIS DE BIOINFORMATIQUE

