

Offre de services en bioinformatique

Ce diaporama est accessible ici : <https://goo.gl/89r1t3>

*Jacques van Helden
& les groupes de travail NNCR*

Institut Français de Bioinformatique (IFB)



Structure de l'exposé

- Services de l'IFB: un réseau national de ressources bioinformatiques
 - Un réseau de ressources humaines
 - Calcul et stockage
 - Environnements logiciels
 - Bases de données
 - Catalogue des ressources et guichet
- Les défis
 - Principe FAIR: données, logiciels, processus d'analyse, environnement
 - La course aux technologies
 - (évolution des technologies -> des logiciels, interfaçage, intégration, biohackathons)
 - Mutualisation des ressources entre infrastructures
 - Interdisciplinarité
 - Bioinformatique pour la santé, l'agronomie, l'environnement, la recherche fondamentale

Le plan d'action de la feuille de route IFBv2

WP1. Un environnement distribué de services en bioinformatique

Christophe Blanchet & Jacques van Helden

A1.1 Réseau national de ressources informatiques (NNCR)

Julien Seiler & Christophe Blanchet & Olivier Collin

A1.2 Software and data environment

Gildas Le Corguillé & Jacques van Helden

A1.3 Support aux bases de données

Claudine Médigue & Guy Perrière

A1.4 Catalogue des ressources

Jacques van Helden & Hervé Ménager

A1.5 Accès aux usagers

Christine Gaspin & Erwan Corre

A1.6 Mutualisation des services inter-infra

Christophe Bruley & Jean-François Dufayard

A1.7 Guichet d'orientation et de consultation

Jacques van Helden & Ivan Moszer

WP3. Ouverture internationale + industrie

Anne-Françoise Adam-Blondin & Claudine Médigue

A3.1 IFB, nœud français d'ELIXIR

Anne-Françoise Adam-Blondin et Claudine Médigue

A3.2 Partenariat avec l'industrie

Patrick Durand et Victoria Dominguez Del Angel

WP4. Formation et diffusion

Jacques van Helden & Morgane Thomas-Chollier

A4.1 Formation

Jacques van Helden & Hélène Chiapello

A4.2 Actions jointes avec SFBI + GDR BIM

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

A4.3 Communication & Valorisation

Claudine Médigue & Victoria Dominguez Del Angel

WP2. Innovation: bioinformatique intégrative

Claudine Médigue & Olivier Sand

A2.1 Projets pilotes inter-infrastructures

Claudine Médigue & Etienne Thévenot & Olivier Sand

A2.2 Appel à défis: lever les verrous scientifiques et technologiques

Ivan Moszer & Jacques van Helden

A2.3 Interopérabilité entre ressources

Marie-Dominique Devignes & Alban Gaignard

WP5. Gouvernance

Claudine Médigue, Jacques van Helden

A5.1 Structures de gouvernance et de coordination

Claudine Médigue & Jacques van Helden

A5.2 Système de gestion de qualité

Claudine Médigue

A5.3 Modèle économique

Christine Gaspin

Offre de services

- A1.1 + A1.2. Réseau national de ressources computationnelles (*National Network of Computational Resources, NNCR*).
 - Infrastructure distribuée
 - Environnement logiciel
 - middleware cluster + cloud
 - outils et workflows
 - accès aux données
 - Déploiement de ressources bioinformatiques développées en France (outils, bases de données).
- A1.3: Appui aux bases de données à haute valeur d'annotation
- A1.4. Catalogue des ressources françaises en bioinformatique
- A1.5. Accès aux usagers (individus, équipes)
- A1.6. Mutualisation des ressources entre infrastructures nationales de support à la recherche (un type particulier d'usager)
- A1.7. Guichet de conseil et d'orientation

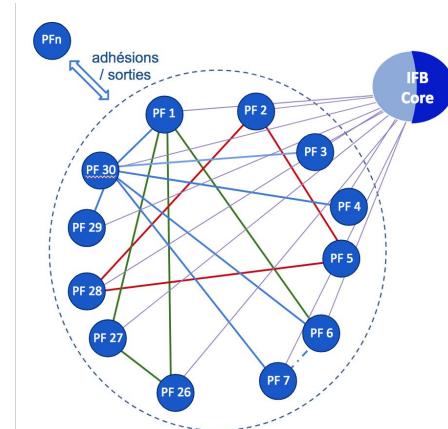


Organisation des services de l'IFB: un réseau national de ressources en bioinformatique

National Network of Computing Resources (NNCR)

Un réseau d'expertises mutualisées

- Chaque action est mise en oeuvre par un groupe de travail (task force).
- Ressources humaines mutualisées
 - Chaque plateforme impliquée mutualise un certain % d'ETP d'une personne pour assurer les tâches communes.
 - Elle bénéficie en contrepartie de tous les % d'ETP mis au pot commun par les autres plateformes.
- Avantages
 - Motivation des agents
 - Evite la redondance des efforts (conception, tests, installation, ...)
 - Partage de l'expérience
 - Solutions conçues d'emblée pour être transposables sur plusieurs serveurs.
 - Elaboration collective de solutions robustes.
 - ...



*Une infrastructure physique
dotée d'un environnement logiciel spécialisé*

Deux types d'accès complémentaires : cluster et cloud

Cluster

Une **ressource persistante** pour des **travaux de longue haleine**

- Modalités d'usage
 - Environnement de base: ligne de commande Unix
 - Possibilité de connexion graphique
 - Espaces de stockage pour projets individuels ou d'équipe ou un groupe de formation
 - Persistance des comptes et des données
- Offre logicielle : des **bouquets logiciels** regroupés en modules Unix
 - **Pré-composés** : RNA-seq, ChIP-seq, variants, BLAST, EMBOSS, ...
 - **A façon** pour formations, projets de recherche, ... (software on demand)

Cloud

Une **ressource volatile** dont vous avez la **maîtrise absolue**

- Session typique
 - Démarrer sa machine virtuelle (VM)
 - L'adapter à ses besoins (l'usager est admin de sa machine!)
 - Faire les analyses
 - Récupérer les résultats
(ne pas oublier !)
 - Eteindre la machine (libérer les ressources pour les autres usagers)
- Offre logicielle : une **diversité de machines "prêtes à l'usage"** (appliances)
 - Génériques
 - Spécifiques méthodo (serveur R, Galaxy)
 - Spécifiques application biologiques (protéomique, métabolisme)

Capacité actuelle de l'infrastructure physique

Fédération de clusters

Node	Compute #CPU HT*	Storage #TB	RAM #GB
ifb-core-cluster <i>Orsay (IDRIS)</i>	4 000	400	20 008
ABiMS <i>Roscoff</i>	1 928	2 000	10 600
GENOTOUL <i>Toulouse</i>	6 128	3 000	34 304
GenOuest <i>Rennes</i>	1 824	2 300	7 500
BiRD <i>Nantes</i>	560	500	4 000
Migale <i>Jouy en Josas</i>	1084	350	70 000
Total	15 524	8 550	146 412

Fédération de clouds

Node	Compute #vCPU HT*	Storage #TB	RAM #GB
ifb-core-cloud <i>Lyon (CC-IN2P3)</i>	3936	408	20 408
ifb-genouest-genostack <i>Rennes</i>	600	150	2 600
ifb-prabi-girofle <i>Lyon</i>	448	144	1 500
ifb-bird-stack <i>Nantes</i>	160	50	448
ifb-bistro-iphc <i>Strasbourg</i>	85	50	1 500
Total	5 229	802	26 456

Sites cloud en préparation : [RPBS](#), [Genotoul-bioinfo](#), [CBP-PSMN](#) (mésocentre généraliste)

* HT:
hyper-thread
(2 files par CPU)

Node	Compute (vCPU HT)	Storage (Tb)	RAM (Gb)
Grand total	20 753	9 352	172 868

RAINBio - Appliances Bioinformatiques dans le Cloud

The screenshot shows a web interface for a bioinformatics appliance catalogue. At the top, there's a header with the RAINBio logo, a search bar containing 'biosphere.france-bioinformatique.fr', and various navigation icons. Below the header, the main title 'RAINBIO - APPLIANCES BIOINFORMATIQUES DANS LE CLOUD' is displayed, followed by a subtitle 'Catalogue des appliances bioinformatiques dans le cloud, filtrez-les en utilisant les termes présents dans l'ontologie EDAM, ou en langage naturel.' A navigation bar at the top left includes 'App Store', 'Appliances', 'Outils', and 'Topics'. The main content area is a grid of appliance cards, each with a title, a brief description, and a list of technologies or tools it uses. The cards are color-coded: green for general bioinformatics, purple for specific tools like BioPipes and Bistar, blue for operating systems like CentOS and Debian, and orange for specialized courses like COURS ENS Lyon NGS 2018 and EBAME 2018. Some cards also mention specific fields like 'Bacterial genomics' or 'PathoTRACK'. At the bottom of the grid, a note says 'Le code couleur reste le même pour une même appliance.'

<https://biosphere.france-bioinformatique.fr/catalogue/>

27 "appliances"

- **Techno** : bioconda, docker (compose, swarm), ansible, clusters (SGE, Slurm), workflows (snakemake, nextflow), bureau virtuel...
- **Serveurs Web** : R studio, Jupyter Notebook, Galaxy (+ celles dédiées)
- **Spécialisation bioinfo**: génomique, annotation, alignement de séquences, réseaux métaboliques, écologie microbienne, structure des protéines, métabolomique, profil d'expression, ...

Formations Hébergées par le Cloud IFB en 2018

Date	Thématique	Lieu	Participants	Editions sur IFB-cloud *
19-20 juin	Biosphère IFB AppDev (Migration StratusLab)	LBBE, Lyon	7	3
25-29 juin	Summer school Genopole	Genoscope, Evry	15	2 *
9-10 juillet	ReproHackathon (GDR MaDICS)	LBBE, Lyon	15	2
11-13 juillet	Summer school Saclay Plant Science 2018 Plant epigenetics and epigenomics	IPS2, Saclay	25	1
12 septembre	Bioinformatique pour le traitement de données de séquençage (CNRS Entreprise)	LBBE, Lyon	12	1
12 sept - 24 oct	Master Biosciences	ENS-Lyon	12	3 *
24-28 sept	ELIXIR Plant Genome Assembling and Annotation (IS WaaS)	Cirad, Montpellier	21	1
29 sept - 06 oct	EBAME Workshop on Computational Microbial Ecogenomics	IUEM, Brest	47	1
30 nov - 5 déc	Master ScV Epigénétique végétale	IPS2, Saclay	18	1
22 oct - 31 jan 2019	Master Meet-U	UPSSaclay/UPMC/Paris Diderot	30	3 *
Déc. 18	DU Création analyse et valorisation des données omiques	Université Paris Diderot	15	1
19 déc - 31 jan 2019	Master AMI2B	Saclay	24	4 *
TOTAL	12		241	

* Pour les masters et écoles, cela correspond au nombre d'années

Offre de service sur le cluster

- **Ressources logicielles pour les chercheurs (biologistes et bioinformaticiens)**
 - SoD : [Software on Demand](#)
 - Organisation par module ([Conda](#) + [Singularity](#))
 - Banques génomiques partagées (bientôt via [BioMaj](#))
- **Espace de déploiement de ressources Web**
 - Serveurs persistants (Galaxy, Phylogeny.fr, RStudio, RSAT, ...).
 - Bases de données
- Réactivité pour répondre aux besoins de formation
 - Espace projets : tuteur/tutoré, espaces lecture-seul
 - Environnements Conda (outils + bouquets thématiques)
 - Un exemple : [7è Ecole de Bioinformatique Aviesan-IFB 2018 \(EBAI18\)](#), Roscoff
 - Demande d'installation : 90 outils, 12 banques de données/6 formats
 - Ouvertures de comptes: 40 participants, 35 formateurs/encadrants
 - Comptes-projet partagés: tutorat basé sur les données des participants (BYOD)
 - Calcul: données des projets analysées sur IFB-core-cluster
 - La ceinture et les bretelles: environnement logiciel et données répliqués sur 3 autres clusters IFB (ABIMS, MIGALE, BISTRO)
 - Forte implication des groupes de travail mutualisés



Soutien aux bases de données

Support aux bases de données

- Ce que nous pouvons proposer : **un soutien logistique pour faciliter la mobilisation des communautés d'experts.**
 - Organisation du déploiement, de la gestion et de mise à jour des bases de données
 - Mutualisation des efforts de déploiement
 - Appui des experts IT du NNCR
 - Développement d'outils d'annotation
 - Structuration de réseaux nationaux et internationaux d'annotateurs
 - formation des experts à l'annotation,
 - événements style « Jamboree »,
- **Visibilité internationale des bases de données françaises**
 - Catalogue français des ressources bioinformatiques
 - ELIXIR (service delivery plan, core data resources)
- Moyens d'action : recrutement d'un CDD.
- Ce que nous ne pouvons pas proposer
 - Participer directement à l'effort d'annotation (c'est le rôle des équipes qui développent ces bases de données)
 - Soutenir financièrement les bases de données (c'est le rôle des EPST et institutions de financement de la recherche)

Identification et accès aux ressources en bioinformatique

Identification et accès aux ressources en bioinformatique

- Ressources couvertes
 - Outils
 - Bases de données
 - Personnes (et leur expertise)
 - Plateformes (et leur expertise)
 - Formations
- Rôles
 - Visibilité des ressources développés en France
 - Identification des expertises françaises
 - pour le conseil
 - pour construire des projets de recherche en amont
 - pour accompagner les projets
 - pour des formations
 - Réutilisabilité des supports de formation
- L'outil de base du **guichet d'orientation et de conseil**
contact@groupe.france-bioinformatique.fr
- Lien avec L'international
- Minimiser la redondance des efforts
 - Synchronisation avec les catalogues ELIXIR
 - Formations: TeSS
 - Outils: bio.tools
- Rationaliser les catalogues
 - Projet ELIXIR: regroupement des catalogues d'outils, packages, workflows, containers.

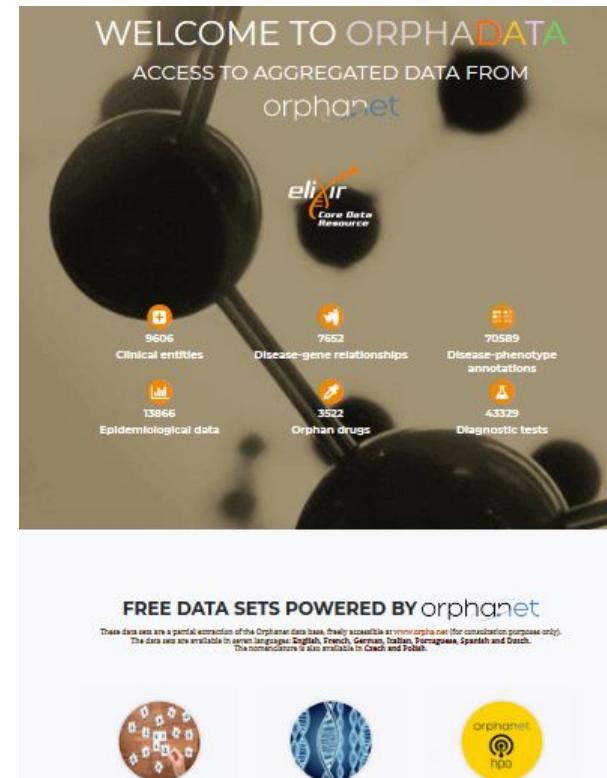
Guichet d'orientation et de consultation – Objectifs

- **Orienter** les utilisateurs vers les bonnes ressources.
 - Réponse aux demandes d'information, mise en route de projets, formation.
 - Orienter les usagers vers les ressources matérielles, logicielles, humaines adéquates.
- **Accompagner** les utilisateurs depuis la conception de projets jusqu'au déroulement des projets.
- **Mesurer** et faire remonter les besoins.
 - Collecter les interactions utilisateurs-IFB qui ne passent pas par le guichet
 - Identifier les thématiques associées aux demandes.
 - Identifier les demandes sans solution disponible -> transfert à l'action “Défis” (A2.2).
- **Optimiser** les ressources de l'IFB
 - Sur base de la mesure des besoins et de la capacité de réponse à l'instant T.
 - Évaluer ce qui fonctionne bien, moins bien, ce qui manque.
- **Promouvoir** les services de l'IFB
 - Assurer le lien entre la demande “de terrain” et les autres actions.

Annotation de ressources dans les catalogues IFB et ELIXIR

- ELIXIR-FR **Service Delivery Plan (SDP)**
 - actuellement 11 services français
 - <https://www.elixir-europe.org/about-us/who-we-are/nodes/france>
- Octobre 2018: appel de l'IFB pour déclarer des ressources françaises (IFB et autres équipes)
 - dans le catalogue IFB;
 - dans le SDP ELIXIR
- 103 ressources soumises
 - L'ensemble des ressources soumises seront référencées sur le catalogue de l'IFB
 - 82 propositions en cours d'évaluation pour le SDP
-

Une **ELIXIR Core Data Resource: OrphaData** (www.orphadata.org)



- Comment offrir un accès “homogène” aux usagers ?
 - Aspects techniques
 - Ressources et technologies complémentaires mais non homogènes
 - Coordination des pratiques par la task force NNCR
 - Aspects financiers
 - Chaque plateforme a ses propres modalités d'accès,
 - Ressources “par défaut” affectée à l'ouverture d'un compte
 - Tarification des services au-delà du “par défaut”
 - Peut dépendre des tutelles et régions
- Chantier en cours
 - Elaboration d'un modèle homogène pour l'ensemble de l'infrastructure, tout en préservant les spécificités institutionnelles et régionales
 - Une première proposition sera soumise aux 31 plateformes IFB, et évaluée en concertation avec les tutelles.

Booster les ressources bioinformatiques : le Biohackathon



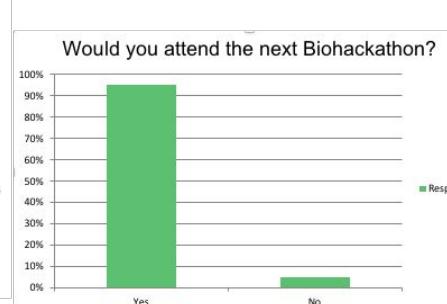
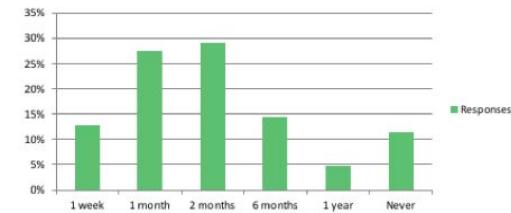
- Principe des hackathons
 - déblocage de verrous techniques;
 - développements collaboratifs basés sur des échanges d'expertises
- Principe d'organisation
 - Soumission de projets
 - Identification d'experts
 - Regroupement des porteurs et experts
 - On enferme tout le monde dans une pièce pendant 2 à 5 jours
 - On passe les pizzas sous la porte
 - Quand les participants envoient la fumée blanche, on rouvre les portes et on publie les résultats (solutions logicielles, cas d'études, comptes-rendus, articles, ...)

Biohackathon 2018, Paris

- But: faire évoluer des ressources bioinformatiques (data, tools, interoperability, training)
- Annuel depuis 2018 au Japon
- 2018: première session en Europe
- Co-organisation **ELIXIR-Hub + ELIXIR-France**
- 29 projets soumis, tous pris en charge
 - Text mining, Structured Metadata, Identifiers, Data Distribution, Data Integration, Validation, Containers, Tool Discovery, Visualisation, Workflows, Training Material, ...
- 150 participants (porteurs de projets + experts)
- 5 jours de hacking
- Résultats:
 - déblocage en quelques jours de verrous qui auraient coûté des mois de travail aux porteurs de projets
 - “Community-building” (réseaux d’expertise, collaborations poursuivies)
 - Taux de satisfaction : 95%
 - ELIXIR-hub relance la collaboration avec ELIXIR-FR pour le Biohackathon 2019



Without the Biohackathon how long would have taken you to reach the same outcome?



Défis actuels de la bioinformatique

Let's go FAIR

- Les données omiques de tous types sont disponibles dans des dépôts internationaux.
- La plupart des labos déposent leurs séquences dans SRA, ENA, GEO, ArrayExpress
- Tout est accessible au public (exception: données patients)
- Questions
 - Qui serait capable de reproduire les résultats de nos (vos) publications à partir de nos (vos) données brutes ?
 - La section “matériel et méthodes” suffit-elle ?
 - Les descriptions des données sont elles précises, complètes, correctes (métadonnées)
 - Les outils sont-ils disponibles ?
 - Les workflows sont-ils disponibles ?
 - L'environnement logiciel est-il transportable ?
- Tout ceci est en principe faisable, mais en pratique ...

Fil rouge de l'interoperabilité = comment faire du “FAIR” ?

FAIR*	Concepts	En pratique pour les données : “Data Management Plan”	En pratique pour les outils : “Software Management Plan”	Exemples de systèmes “FAIR” en bioinfo
F indable (F1 à F4)	Identifiants, Métadonnées, Rechercher...	DOI, URI ; BIOSCHEMAS Catalogues interrogeables,	Idem, BiotoolsSCHEMA	Pour les Données: International <ul style="list-style-type: none">- DataVERSE- FAIRDOM- ISA France (IFB) <ul style="list-style-type: none">- CeSGO
A ccessible (A1 à A2)	Protocoles d'accès ouverts	RDF, LOD JSON	API REST, Web services	
I nteroperable (I1 à I3)	Vocabulaires, langages formels,	Ontologies, web semantique	Containers : Docker, BioConda Workflows : Snakemake,, Galaxy	Pour les Outils: International <ul style="list-style-type: none">- bio.tools- GALAXY France (IFB) <ul style="list-style-type: none">- CoreAPI- Waves
R eusable (R1.1 à R1.3)	Licences ouvertes Provenance, Standards	MIAME, MIAPPE	Licences Creative Commons EDAM : Metadonnées sur les Entrées-Sortie des outils	Rq: Distinction de différents “niveaux” de FAIRification !

Portabilité des ressources

Portabilité des ressources

- Les environnements logiciels développés à l'IFB sont conçus pour être “portables” sur d'autres environnement
- Outils : conda
- Workflows : Galaxy, SnakeMake, CWL, NextFlow
- Virtualisation : machines virtuelles (appliances cloud), containers (Docker, singularity)
- Intérêt
 - Installations locales: plateaux techniques, serveurs d'unités, machines personnelles
 - Données sensibles: déplacer les logiciels vers les données plutôt que l'inverse

Défis de la bioinformatique pour la santé

Enjeux et défis de la science des données en santé

- Volumes de données
 - Capacité de stockage
 - Archivage
- Hétérogénéité des données
 - capacité d'analyse (intégration, interopérabilité)
 - Définition et utilisation d'ontologies communes (HPO, ORDO)
- Défis conceptuels
 - Assemblage de données hétérogènes
 - Créer de nouvelles techniques d'Intelligence Artificielle et de "Machine learning"
- Des attentes ambitieuses
 - Répondre aux problèmes de santé publique
- Spécificité des données humaines
 - Les méso-centres ne possèdent pas les accréditations pour traiter les données de santé
 - IFB: seul Curie possède l'accréditation
 - Agrégation et interconnexion des données de santé sans dévoiler l'identité du patient
 - Éviter l'appropriation/exploitation de ces données par les GAFAM.
- L'IFB est au premier plan pour affronter l'ensemble de ces défis.

Services IFB pour répondre à ces défis

Offre générique de services

- Un réseau national de serveurs (calcul et stockage)
- Des environnements logiciels spécialisés et transportables (conda, machines virtuelles, containers)
- Un réseau national des plateformes de services
- Un réseau national d'expertises
 - 16 Groupes de travail du plan d'action IFB
- Un réseau international d'expertises
 - ESFRI ELIXIR
- Innovation
 - Bioinformatique intégrative
 - Interopérabilité
 - **Course aux technologies:** besoin permanent de développement logiciel (translationnel: *nécessite de financer en amont la recherche en bioinformatique*)

Actions spécifiques : bioinformatique pour la santé

- GT IFB **Bioinformatique pour la santé**
 - Contacts: David Salgado et Ivan Moszer
- GT **Métiers et formations en bioinformatique pour la santé**
 - Aviesan, Inserm, FMG2025, SFBI, IFB
 - Contact : Catherine Nguyen
- Des plateformes de services avec **compétences spécifiques en santé** (voir exposé “panorama”)
- Bases de données et outils logiciels spécialisés santé
- Débat sur l'**accréditation** (fin de matinée)
 - Rôle de l'IFB ou structuration interne de l'infrastructure l'Inserm ?
 - Transporter les données vers les logiciels, ou les logiciels vers les données ?

Générique : les “task forces” mutualisées de l’IFB

La force de frappe de l'IFB

- Le plan d'action 2018-2021 de l'**Institut Français de Bioinformatique** repose sur un réseau de plateformes, qui rassemblent une large diversité d'expertise.
- Chaque action est menée par un groupe de travail (“task force” mutualisée) basé sur l'engagement volontaire.
- L'ensemble des groupes de travail regroupent actuellement 133 personnes (dont certaines contribuent à plusieurs groupes).

Task force infrastructure physique (A1.1) et logicielle (A1.2)



- Jacques van Helden
 - Coordination WP1
 - A1.2: environnement logiciel
 - Beta-testeur formation



- Julien Seiler
 - Responsable cluster
 - Resp. A1.1 infra physique



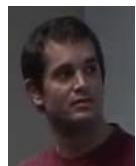
- Gildas Le Corguillé
 - Responsable cluster
 - Resp A1.2 env. logiciel



- Nicole Charrière
 - Administrateur Système



- David Benaben
 - Administrateur Système



- Guillaume Seith
 - Administrateur Système



- Christophe Blanchet
 - Coordination WP1
 - Responsable cloud



- Olivier Collin
 - Responsable cloud



- Jonathan Lorenzo
 - Ingénieur DevOps et intégration cloud



- Olivier Sallou
 - Ingénieur DevOps et administrateur cloud
 - IS Biocontainers



- Jean-François GUILLAUME
 - Ingénieur Systèmes, Réseaux et Virtualisation
 - Responsable Cluster, Cloud, Stockage

Task force infrastructure physique (A1.1) et logicielle (A1.2)



- Bruno Spataro
 - Administrateur Système et réseau
 - Administrateur Cloud



- Stéphane Delmotte
 - Administrateur Système et réseau
 - Administrateur Cloud



- Marie-Stéphane Trotard
 - Administrateur Système et réseau



- Didier Laborie
 - Administrateur Système et réseau



- Céline Noirot
 - Ingénierie en développement bioinformatique



- Olivier Inizan
 - Spécialiste Galaxy



- Jérôme Pansanel
 - Administrateur Système et réseau
 - Administrateur Cloud

- Efflam Lemaillet
 - Administrateur Système et réseau
 - Ingénieur Devops & Cloud

- Chloé Riou
 - Spécialiste BioMaj

- Erwan Corre
 - Ingénieur bioanalyste

- Loraine Gueguen
 - Ingénierie en développement bioinformatique

Groupe de travail soutien aux bases de données (A1.3)



- G. Perrière
 - Responsable action A1.3
- C. Médigue
 - Responsable action A1.3
- V. Navratil
 - Participant au GT
- J.P. Flandrois
 - Participant au GT
- D. Vallenet
 - Participant au GT
- S. Ricard-Blum
 - Participant au GT
- M. Alaux
 - Participant au GT
- M. Rouard
 - Participant au GT
- P. Lemaire
 - Participant au GT



- C. Dantec
 - Participant au GT
- S. Bocs-Sidibe
 - Participant au GT
- M. Mendoza
 - Participant au GT
- D. Lagorce
 - Participant au GT
- G. Vergnaud
 - Participant au GT
- C. Pourcel
 - Participant au GT
- P. Siguier
 - Participant au GT
- D. Salgado
 - Participant au GT
- M. Hanauer
 - Participant au GT
- ...



GT “Catalogue des ressources bioinformatiques françaises” (A1.4)



- Hervé Ménager
 - Resp. action A1.4
 - Pasteur



- Jacques van Helden
 - Resp. action A1.4
 - IFB-core



- Laurent Bouri
 - Développement du catalogue des ressources bioinformatiques
 - IFB-core



- Sylvain Milanesi
 - Appui au développement du catalogue des ressources bioinformatiques.
 - IFB-core



- Ivan Moszer
 - Lien avec le guichet A1.7
 - ICM



- Alban Gaignard
 - Lien avec l'action 2.3
 - BIRD



- Victoria Dominguez Del Angel
 - IFB core



- Vincent Navrati
 - PRABI



- Guy Perrière
 - PRABI



- Hélène Chiapello
 - MIGALE

Groupe de travail “Accès aux usagers” (A1.5)



- Erwan Corre
 - Resp. action A1.5
- Christine Gaspin
 - Resp. action A1.5
- Audrey Bihoué
 - Participante au GT
- Christophe Blanchet
 - Participant au GT
 - Action 1.1 Cloud
- Olivier Collin
 - Participant au GT
- Valérie Cognat
 - Participante au GT
- Jean François Dufayard
 - Participant au GT



- Vincent Lefort
 - Participant au GT
- Valentin Loux
 - Participant au GT
- Ivan Moszer
 - Participant au GT
 - Action 1.7
- Guy Perriere
 - Participant au GT
- Coraline Petit
 - Participante au GT
- Julien Seiler/Gildas Le Corguillé
 - Participant au GT/
 - Action 1.1/1.2 Cluster



Groupe de travail mutualisation inter-infra (A1.6)



- Christophe Bruley
 - Resp. action A1.6
- Jean-François Dufayard
 - Resp. action A1.6
- Nicolas Pade
 - Coord. **EMBRC-FR**
- Erwan Corre
 - Rep. **EMBRC-FR**
- Jean Salamero
 - Coord. **FBI**
- Charles Kervrann
 - Rep. **FBI**
- Perrine Paul-Gilloteaux
 - Rep. **FBI**
- Etienne Thévenot
 - Rep. **MetaboHUB**



- Denis Milan
 - Coord. **FG**
- Claude Scarpelli
 - Rep. **FG**
- Alain Viari
 - Rep. **FMG2025**
- Régine Trebossen
 - Rep. **FLI**
- Christian Barillot
 - Rep. **FLI**
- Yann Herault
 - Coord. **CELPHEDIA**



Groupe de travail “Guichet d'orientation et de consultation” (A1.7)



- Ivan Moszer
 - Resp. action A1.7



- Jacques van Helden
 - Resp. action A1.7



- **Sylvain Milanesi**
 - Mise en place du guichet d'orientation et de consultation



- Laurent Bouri
 - Appui au déploiement du guichet



- Victoria Dominguez del Angel
 - Participante au GT



- Hélène Chiapello
 - Participante au GT



- Claire Hoede
 - Participante au GT



- Valérie Cognat
 - Participante au GT



- Guy Perrière
 - Participant au GT



- Erwan Corre
 - Participant au GT



- Vincent Navratil
 - Participant au GT



- Alexis Groppi
 - Participant au GT

Groupe de travail Projets-Pilotes (A2.1)



Claudine Médigue

- IFB-core
- Coordinatrice WP2/A2.1



Etienne Thévenot

- MetaboHub (CEA Saclay)
- Coordinateur A2.1/porteur ProMetIS



Olivier Sand

- IFB-core
- Coordinateur WP2/A2.1



Jacques van Helden

- IFB-core
- Coordinateur A2.2



Alban Gaignard

- Plateforme BIRD (Nantes)
- Porteur INEX-MED



Julie Thompson

- Plateforme Bistro (Strasbourg)
- Porteur INEX-MED



Pierre Tuffery

- RPBS (Paris)
- Porteur MS2MODELS



Raphaël Guérois

- FRISBI (Saclay)
- Porteur MS2MODELS



Cyril Pommier

- URGI (Versailles)
- Porteur PhenoMeta



Ivan Moszer

- iCONICS (Paris)
- Porteur IntegrParkinson



Stéphane Lehérecy

- CENIR-ICM (Paris)
- Porteur IntegrParkinson



Faisal Bekkouche

- EMBRC (Villefranche/mer)
- Porteur My-EMBRC-Image



Pascal Roy

- PRABI (Lyon)
- Porteur B2SH



Christophe Béroud

- INSERM (Marseille)
- Porteur BANCCO



David Salgado

- INSERM (Marseille)
- Porteur BANCCO



David Vallenet

- MicroScope (Evry)
- Porteur ProMetIS



Etienne Camenen

- IFB-core
- CDD IntegrParkinson



Guillaume Postic

- IFB-core
- CDD MS2MODELS



Mélanie Buy

- IFB-core
- CDD PhenoMeta



Alyssa Imbert

- IFB-core
- CDD ProMetIS



Maxime Folschette

- IFB-core
- CDD INEX-MED

Groupe de travail Interopérabilité (A2.3)



- Marie-Dominique Devignes
 - Plateforme MBI (Nancy)
 - Co-responsable Action Interop



- Jacques van Helden
 - Plateforme TAGC-BCF (Marseille)
 - Co-responsable Action Interop.
 - Collège de direction



- Khalid Belhajjame
 - LAMSADE (Paris)
 - Reproductibilité des workflows scientifiques
- Sarah Cohen-Boulakia
 - LRI (Orsay)
 - Première co-responsable de l'Action Interop
- Alban Gaignard
 - Plateforme BIRD (Nantes)
 - Projet Pilote INEX-MED
- Hervé Ménager
 - Plateforme Pasteur (Paris)
 - Co-responsable Action 1.4 Catalogue



- Vincent Lefort
 - Plateforme ATGC (Montpellier)
 - Collège de direction
- Olivier Sand
 - IFB-CORE (Paris)
 - Responsable Projets-Pilotes.
- Cyril Pommier
 - Plateforme URGI (Versailles)
 - Plant and Fungi Data Integration



Groupe de travail Formation (A4.1)



- Hélène Chiapello
 - Plateforme: MIGALE/IFB-core
 - Resp. A4.1 formation
 - Coord. DUBii
- Jacques van Helden
 - Plateforme: IFB-core
 - Resp. A4.1 formation
 - Coord. DUBii & EBAI
- Laurent Bouri
 - Plateforme: IFB-core
 - Manager of the web site
- Erwan Corre
 - Plateforme: ABIMS
 - Coord. EBAI
 - Action 1.5
- Joelle Amselem
 - Plateforme: URGI
- Christine Tranchant-Dubreuil, Sarah Gautier, François Sabot
 - Plateforme: SouthGreen



- Yousra Mahmah
 - CDD A4.1
 - Ingénierie de la formation
- Victoria Dominguez Del Angel
 - Plateforme: IFB-core
 - Coord. formation ELIXIR-FR
- David Salgado
 - Plateforme: GMGF-GNiT
- Gildas Le Corguillé
 - Plateforme: ABIMS
- Claire Hoede
 - Plateforme : Genotoul-Bioinfo
- Raluca Teusan
 - Plateforme: BIRD
- Emmanuel Bettler
 - Plateforme: PRABI-LG

Diapos supplémentaires

Pour en savoir plus sur le “FAIR” :

- **Articles de référence**
 - The FAIR Guiding Principles for scientific data management and stewardship. Wilkinson, Dumontier et al., *Scientific Data volume 3*, Article number: 160018 (2016)
 - Developing a Framework for Digital Objects in the Big Data To Knowledge (**BD2K**) Commons : report from the Commons Framework Pilots Workshop. Jagodnik KM et al., *J. Biomed Inform.* 71:49-57 (2017)
 - A design framework and exemplar metrics for FAIRness. Wilkinson et al. *Scientific Data volume 5*, Article number: 180118 (2018)
 - **Principes du FAIR en français**
 - <https://www6.inra.fr/datapartage/Produire-des-donnees-FAIR>
- **Elixir Interoperability Platform ; FAIR-DOM / SEEK**
 - <https://www.elixir-europe.org/platforms/interoperability>
 - <http://fair-dom.org>
- **Récente journée de l’IN2P3 :“Comment FAIR en pratique?” 27 nov 2018 à Jussieu** <https://gt-donnees2018.sciencesconf.org/resource/page/id/3>
- **Initiative GO-FAIR : France-Allemagne-Pays-Bas**
 - A bottom-up international approach for the practical implementation of the European Open Science Cloud (EOSC) as part of a global Internet of FAIR Data & Service
 - <https://www.go-fair.org/go-fair-initiative/>

Pour les geeks: les dessous du cloud

- Déploiement d'**environnements multi-cloud** composés de 1 à N machines virtuelles
 - Définitions basées sur des recettes écrites par les développeurs
 - Scripts **shell**, playbook [ansible](#), environnement [docker](#) simple ou complexe ([docker-compose](#)), environnement [conda](#), workflow [snakemake](#) ou [nextflow](#) (associé à docker et/ou conda), ...
 - Enregistrement d'**images complètes des VM**
 - Déployable en un clic par les utilisateurs
 - Dépôts des **banques publiques de référence** (associés aux clouds, avec BioMaJ)
- Catalogue d'environnements pré-définis : [RAINBio](#) sur [BioSphere](#)
 - 27 "appliances" prêtes à l'emploi, de nombreux outils intégrés
- Services natifs de cloud
 - Directement auprès des sites : [Openstack horizon](#), nova, cinder, manila, heat, ...
- Usage
 - 149 utilisateurs (déc. 2018) + comptes temporaires formations (194) + utilisation opportuniste ELIXIR (55 000 heures, via [EGI](#))
 - **12 formations** soutenues depuis juin 2018



nextflow



RAINBIO

A faire

- Le jour où j'aurai le temps + les infos
 - FMG2025 (inviter Alain sur diapos)
 - Bases de données
 - Orphanet
 - BD CNV
 - Ressources soumises au SDP spécifiques à la santé
 - Un service à la communauté : le développement de logiciels bioinfo
 - Un trombinoscope général de l'IFB (remplacer les multiples trombinoscopes)
 - Un réseau personnes - groupes de travail
 - Anchage des ingénieurs dans les plateformes santé, immersion recherche ?
 - ...