



INSTITUT FRANÇAIS DE BIOINFORMATIQUE

Webinaire Inserm Clouds-Santé, 4 octobre 2024

Services numériques cloud IFB-Biosphère pour les applications en Biologie

Christophe Blanchet

**Institut Français de Bioinformatique - IFB
CNRS UAR360I & PRABI-AM, LYON, FRANCE**



Développer un service national pour fournir

- une infrastructure de base de services bioinformatiques aux communautés françaises des sciences de la vie ;
- une formation en bioinformatique pour les scientifiques ;
- un soutien aux communautés dans les programmes de recherche en biologie, santé, agronomie et environnement ;
- un accès à des technologies de pointe, des expertises de haut niveau et des compétences mutualisées.

Développer une vision stratégique pour maintenir la recherche française au plus haut niveau d'expertise

- Axe d'innovation : bioinformatique intégrative
- Être utilisé comme levier pour des projets de recherche nationaux et internationaux ambitieux en sciences de la vie et leurs applications

Assurer la représentation internationale de la communauté bioinformatique française

IFB INSTITUT FRANÇAIS DE BIOINFORMATIQUE

Contact

À propos Services Formation Innovation ELIXIR-FR Ressources **L'IFB recrute**

Bienvenue à l'Institut Français de Bioinformatique

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

En savoir +

- Calcul & Stockage**
Clusters • Cloud
National Network of Computing Resources (NNCR)
EN SAVOIR +
- Accompagnement**
Projets de recherche • Communautés • Guichet d'orientation
EN SAVOIR +
- Formation**
Formation professionnelle & initiale
EN SAVOIR +
- Innovation**
Bioinformatique intégrative
EN SAVOIR +
- ELIXIR-FR**
Noeud français du réseau européen d'infrastructures bioinformatiques
EN SAVOIR +
- Catalogue de ressources**
Outils, données, expertises, services en bioinformatique
EN SAVOIR +
- Remerciements**
Pour vos publications
EN SAVOIR +
- Roadmap**
Feuille de route • KPI
EN SAVOIR +

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

Le périmètre de l'IFB

Une infra numérique distribuée

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

20 plateformes membres
8 PF contributrices
8 équipes associées
>400 experts (~200 FTE)

Des ressources mutualisées

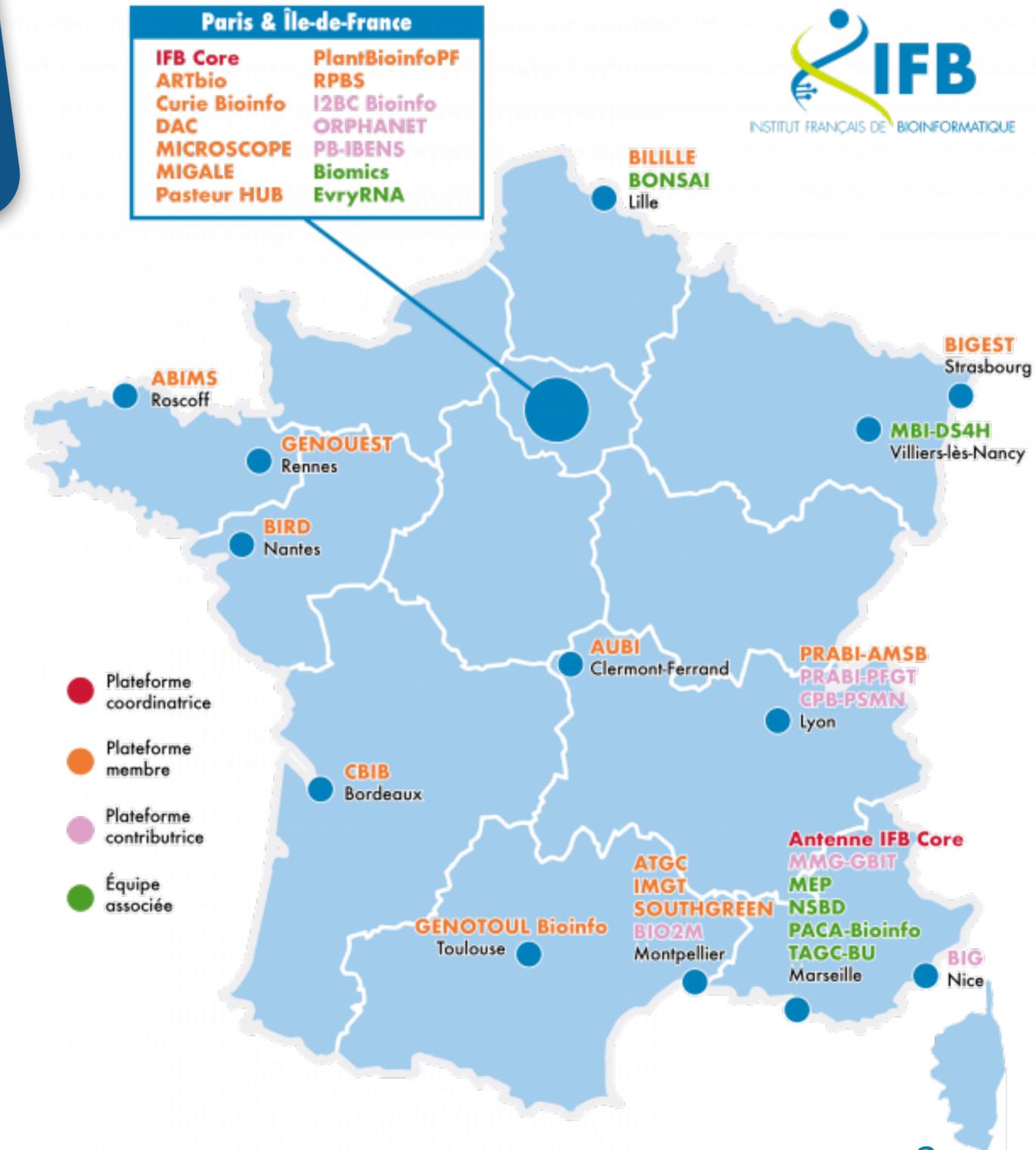
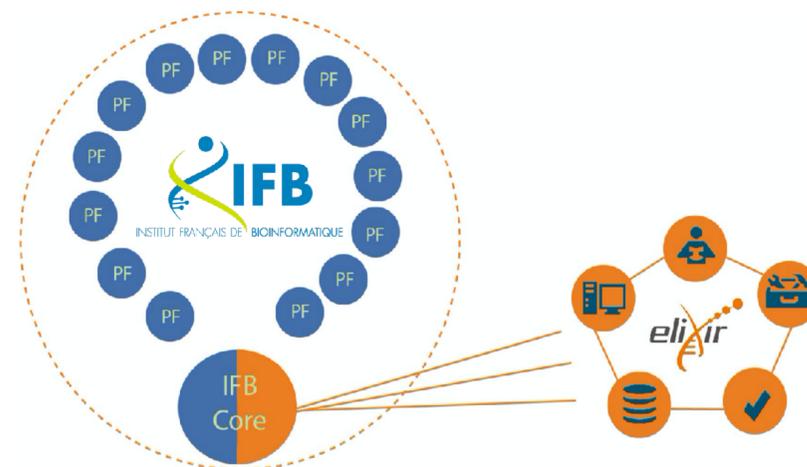
- financements d'infrastructure:
PIA1 RENABI-IFB (22.8M€), PIA3 MUDIS4LS (16.5M€)
- missions articulées autour de "Task Forces" mutualisées entre les plateformes

Atteindre toutes les communautés

- biologie, santé, agriculture, environnement

Coordination: IFB-core (UAR 3601)

- multi-tutelle (CNRS, Inserm, INRAE, CEA)
- mission : coordination et gestion des moyens mutualisés
- noeud français d'ELIXIR (ELIXIR-FR)



Une infrastructure distribuée (NNCR)

- 9 sites
- 30,000+ cores-HT, 165 TB RAM, 14 PB storage

Des technologies complémentaires : cluster et cloud

- 7 batch **clusters** (SLURM, OnDemand, Module, Apptainer)
- 8 **clouds** (Openstack, Docker)

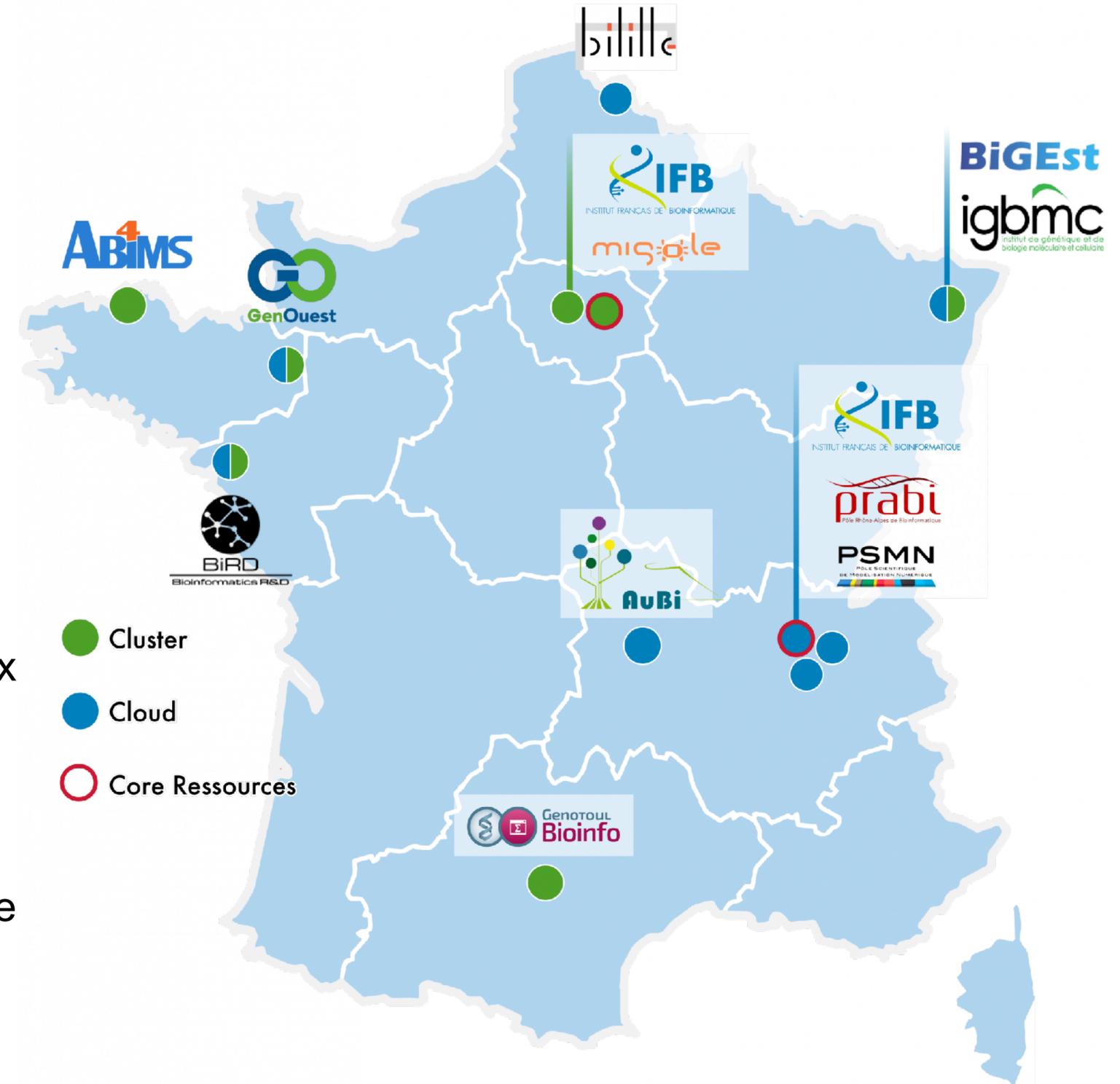
Gestion par un réseau collaboratif d'équipes techniques

- Par les ingénieurs des plateformes régionales et IFB-core
- Partage des connaissances
- Partage des développements
- Assistance mutuelle en cas de panne
- Collaboration avec les centres de calcul nationaux et régionaux

Stockage des données

- Pour les besoins de projets
- Pas de certification pour les données sensibles (HDS...)
- Développements en cours pour des environnements virtuels de bulles sécurisées (PIA3 MUDIS4LS, PEPR CLOUD4SAMS)

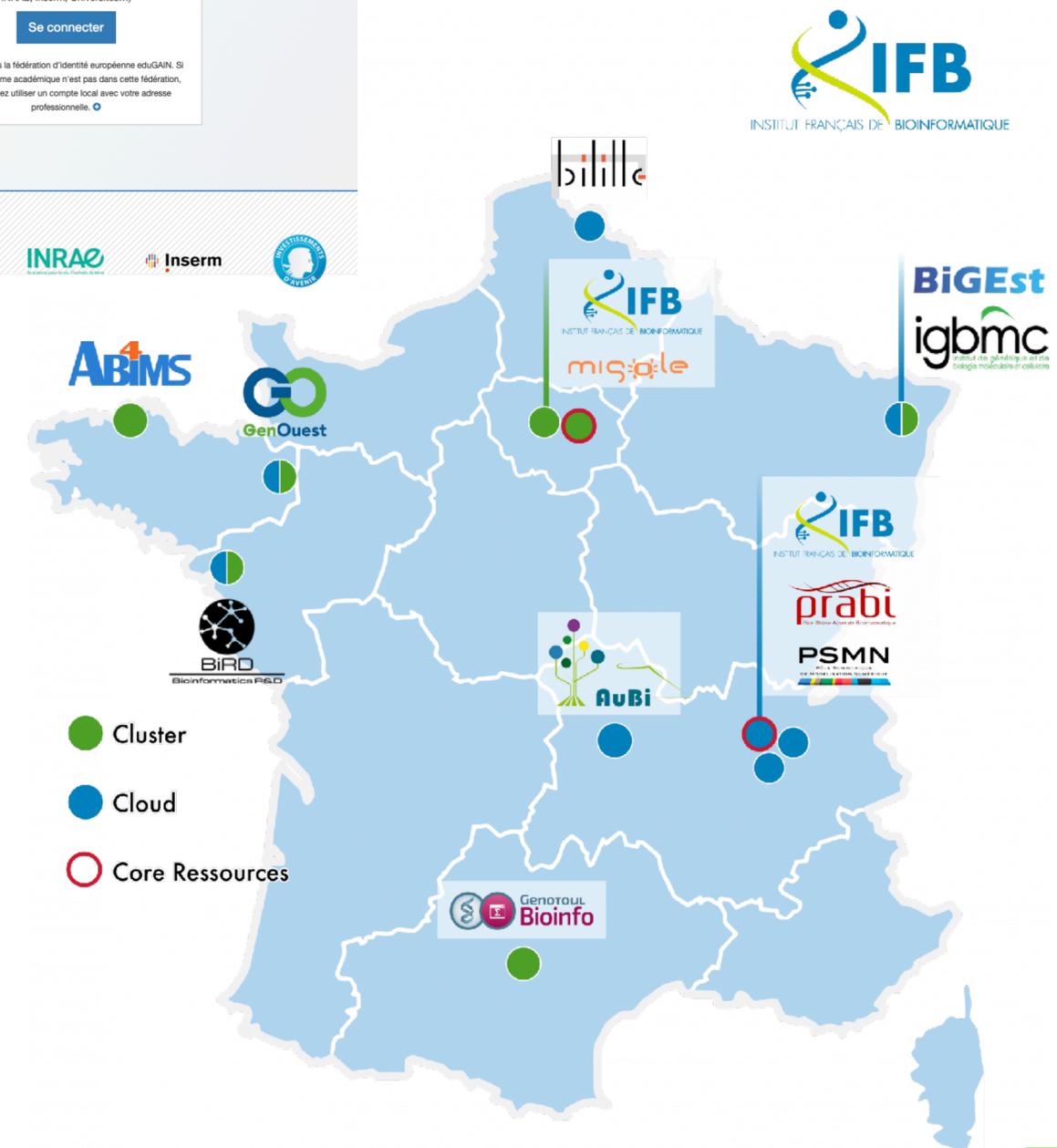
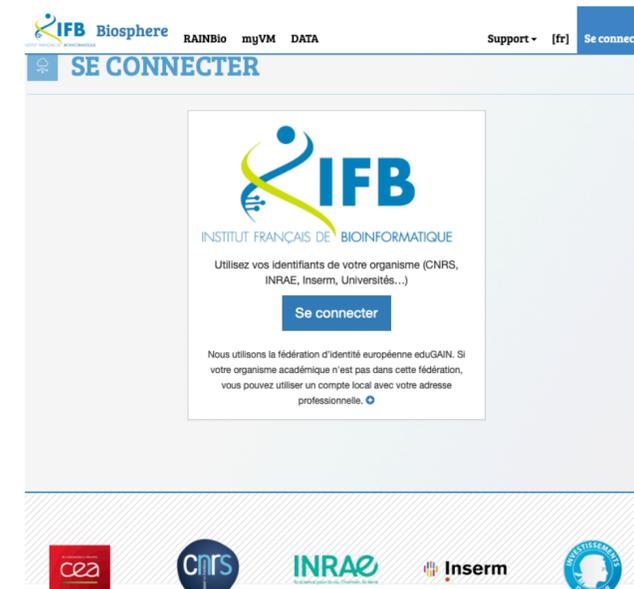
<https://www.france-bioinformatique.fr/calcul-et-stockage/>



Infrastructure Cloud de l'IFB

Une infrastructure informatique pour déployer des environnements bioinformatiques dans des machines virtuelles et des conteneurs pour le traitement des données biologiques.

- Un **portail unifié Biosphère** pour déployer tous les environnements bioinformatiques sur tous les clouds.
- Des **environnements de traitement pré-définis**, prêts à être **déployés en un clic** depuis le **catalogue RAINBio**
- Un **compte unique pour tous les clouds**, avec les **identifiants officiels** des utilisateurs dans leur organisme (fédération d'identités)
- **8 sites cloud** avec 7 000 vCPU et 30 To RAM, des ressources à large mémoire, haute fréquence CPU et un grand nombre de coeurs par CPU.
- Haute-disponibilité avec les différents sites utilisables de la même façon.
- Accès aux bases de données publiques de référence.



Fédération des clouds IFB-Biosphère

L'infrastructure IFB-Biosphère est une fédération de 8 clouds IFB
(partenariats avec 5 centres régionaux et 1 national)

Un portail utilisateur

- Gestion des accès avec la fédération RENATER

Accès sécurisés aux VMs

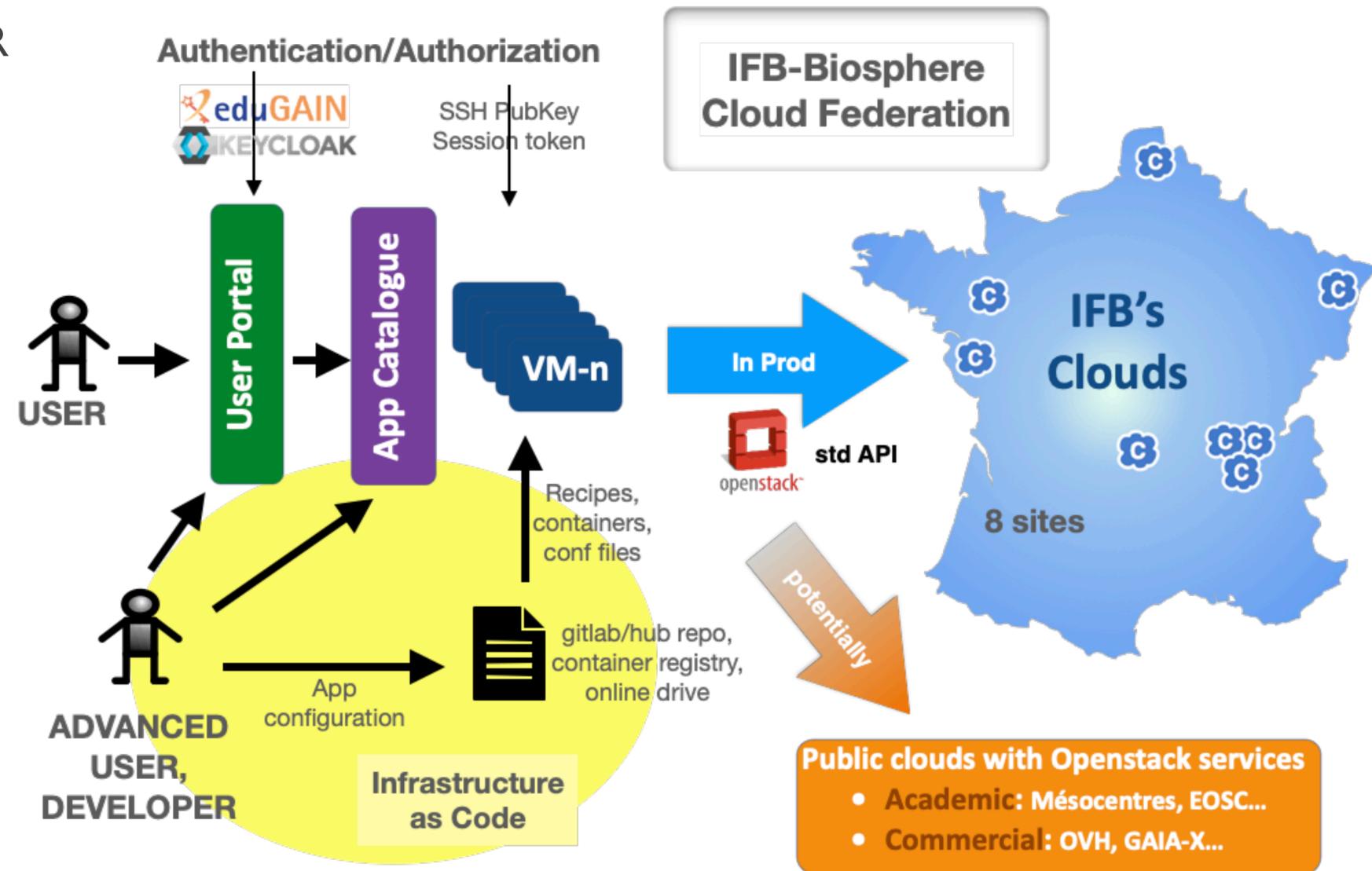
- Clé publique SSH de l'utilisateur
- Tokens uniques de session pour les interfaces Web

Infrastructures avec standards ouverts

- Cloud Openstack, conteneur Docker
- Stockage CEPH, ZFS, NAS

Services aux développeurs

- Configuration des VMs avec dépôts gitlab/hub, délégation droits, gestion asynchrone des modifications (Infrastructure as Code)
- Catalogue de conteneurs pré-définis (déploiement + rapide)



RAINBio - Catalogue des Apps Bioinformatiques

Appliances

- Environnements **prêts-à-emploi**
- Déploiement **en 1 clic**
- **Multi-cloud**, plusieurs VMs
- Interfaces usuelles: SSH, HTTPS, Desktop
- Définies par les développeurs
- Adaptables par les utilisateurs

Apps de base

- Outils génériques : linux, workflows, Rstudio/Shiny, Jupyter...

Apps thématiques

- Génomique, bioimagerie, écologie microbienne, biostatistique, chemins métaboliques, métagénomique, transcriptomique...

The screenshot displays the RAINBio website interface. At the top, there's a navigation bar with 'IFB Biosphere', 'RAINBio', 'myVM', and 'DATA'. Below this is a header section titled 'RAINBIO - APPLIANCES BIOINFORMATIQUES DANS LE CLOUD' with a subtitle 'Catalogue des appliances bioinformatiques dans le cloud, filtrez-les en utilisant les termes présents dans l'ontologie EDAM, ou en langage naturel.' The main content area is a grid of application cards, each with a title, a list of tools, and a brief description. The cards are organized into columns and rows, with some cards having a play button icon. The cards include:

- AnalysesSV**: bcftools, BEDTools, BWA, Jupyter, Matplotlib, pandas; DNA polymorphism, Genetic variation, Genotyping experiment, GWA.
- ANF MetaBioDiv**: Bioconductor, DESeq2, devtools (R), ggplot2, gridExtra; Bioinformatics, Computational biology, Data management, Trans.
- ANF OMERO-FAIRly**: bioconda, CellProfiler, Jupyter, Matplotlib, Numpy, pandas; Microarray experiment, Data architecture, analysis and design, Mathe.
- Askomics**: AskOmics; Data integration and warehousing, Data.
- ateliersUAR2AD**: FastQC, SPAdes; Sequence assembly.
- Bacterial Genomics**: HMMER, InSight, SGE - GridEngine, Ubuntu, Web interface; Protein folds and structural domains, Sequence comparison, Sequenc.
- Bioimaging Desktop**: Bureau virtuel, Icy, ImageJ-Fiji, MATE, Napari, OMERO; Light microscopy, Medical imagin, Informatics, Data visu.
- biom3d**: Biom3d, Bureau virtuel, python3, X2Go, XFCE; Bioinformatics, Biology, Biomedical science, Comput.
- BioPipes**: bioconda, cwltool, Docker, Nextflow, Snakemake; Informatics, Bioinformati, cs, Workflows.
- bistar**: bioconda, Bowtie2, FastQC, Snakemake; Bioinformatics, Genomic, S. Mapping, Sequence alignment.
- Cytoscape**: Bureau virtuel, Cytoscape, X2Go, XFCE; Bioinformatics, Data visualisation, Molecular inte.
- Debian 11**: Ansible, bioconda, Docker; Bioinformatics, Informatic, S.
- Debian 12**: Ansible, bioconda, Docker; Bioinformatics, Informat, ics.
- Decomics**: Bioconductor, ggplot2, Shiny; Bioinformatics, Computational biology, Data management, Data ar.
- DRomics**: Bioconductor, DESeq2, DRomics, R - base, Shiny; Bioinformatics, Compu, tational biology, Data mana.
- EBAME-2022 aDNA**: BAMtools, BEDTools, bioconda, Bowtie2, metaDMG-; Data management, Bioinfor, matics, Sequence alignment, C.
- EBAME-2023 MetaTOR**: Bowtie2, CheckM, hicstuff, Pairs, SAMtools; Sequence alignment, Phylogeneti, cs, Phylogenomics, Taxon.
- EBAME-8 Metabo Modelling**: gapseq, kegg2bipartitegraph, MeneTools, Metage2m; Enzymes, Genotype and phenotype, Small molecules, Data vis.
- EBAME-Anvio**: Anvi'o; Genomics, Omics, Me, tagenomics, Phyloaer.
- EBAME-Quince**: bam-readcount, BEDTools, BWA, CONCOCT, DESeq2; Transcriptomics, Infor, matics, Sequencing, Genomic.
- ETBII Reseaux**: BINGO (Cytoscape), Bureau virtuel, compositions (R).
- formation_CIRI**: DESeq2, FastQC, HISAT2, RStudio, Trimmomatic.
- Formation LBBE-NGS 2024**: BEDTools, Bioconductor, DESeq2, HISAT2, IGV - Inte.
- frangiPANe**: ABySS, bioPython, BWA, CD-HIT, Docker, Jupyter, py.
- Galaxy**: bioconda, Docker, Galax, y portal.

IFB-Biosphère - Apps de base

- **Ubuntu/Debian, CentOS/RockyLinux**
 - ★ Docker +Compose, Ansible(roles), Bioconda
 - ★ Desktop (XFCE + X2GO)
- **BioPipes**
 - ★ Nextflow, SnakeMake, CWLtool, Conda + Bioconda, Docker + Compose
- **R, Rstudio, Shiny**
 - ★ *Images officielles* : tidyverse, verse, ML/cuda, geospatial...
- **Jupyter Notebook**
 - ★ kernels Python, R, Julia
 - ★ *Images officielles* : notebook, R, scipy, datascience, tensorflow...
- **Bioimage**
 - ★ ImageJ-Fiji, Icy, QuPath, Napari
 - ★ connecteur OMERO + plugins
- **Cytoscape**
 - ★ connecteurs NDexBio, PSICQUIC

Maintenues par IFB-core

Infrastructure as Code

<https://gitlab.in2p3.fr/ifb-biosphere/apps>



BIOCONDA®

nextflow

Snakemake



Offre de Services IFB-Biosphère

Services Utilisateurs

- Un portail multi-cloud unique (portail Biosphère).
- Des environnements reproductibles (67) en un clic pour les analyses (catalogue RAINBio).
- Modifiables avec les outils habituels (bioconda, pip, docker, paquets R, paquets système, ansible, scripts shell...).
- Connexion avec les identifiants académiques usuels des organismes et universités (fédération d'identités).
- 8 sites fournissant 7 000 vCPU et 30 To RAM :
 - ★ BigMem (4To), HighFreq (>3,8GHz), ManyCores (>100 vCPU par VM).
- Haute-disponibilité par redondance des sites
- Mise à disposition des données publiques de référence
- Volumes partagés

Services cloud natifs (Openstack)

- Directement auprès des sites

Offre de Services IFB-Biosphère (2)

Services Développeurs

- Infrastructure as Code : configuration VM avec dépôts gitlab/hub, délégation droits, gestion asynchrone des modifications
- Conteneur pré-définis (déploiement + rapide)

Services Administrateurs Cloud

- Tableau de contrôle des ressources déployées pour son cloud
- Monitoring (utilisateurs/groupes, apps)

Services Formations & Projets

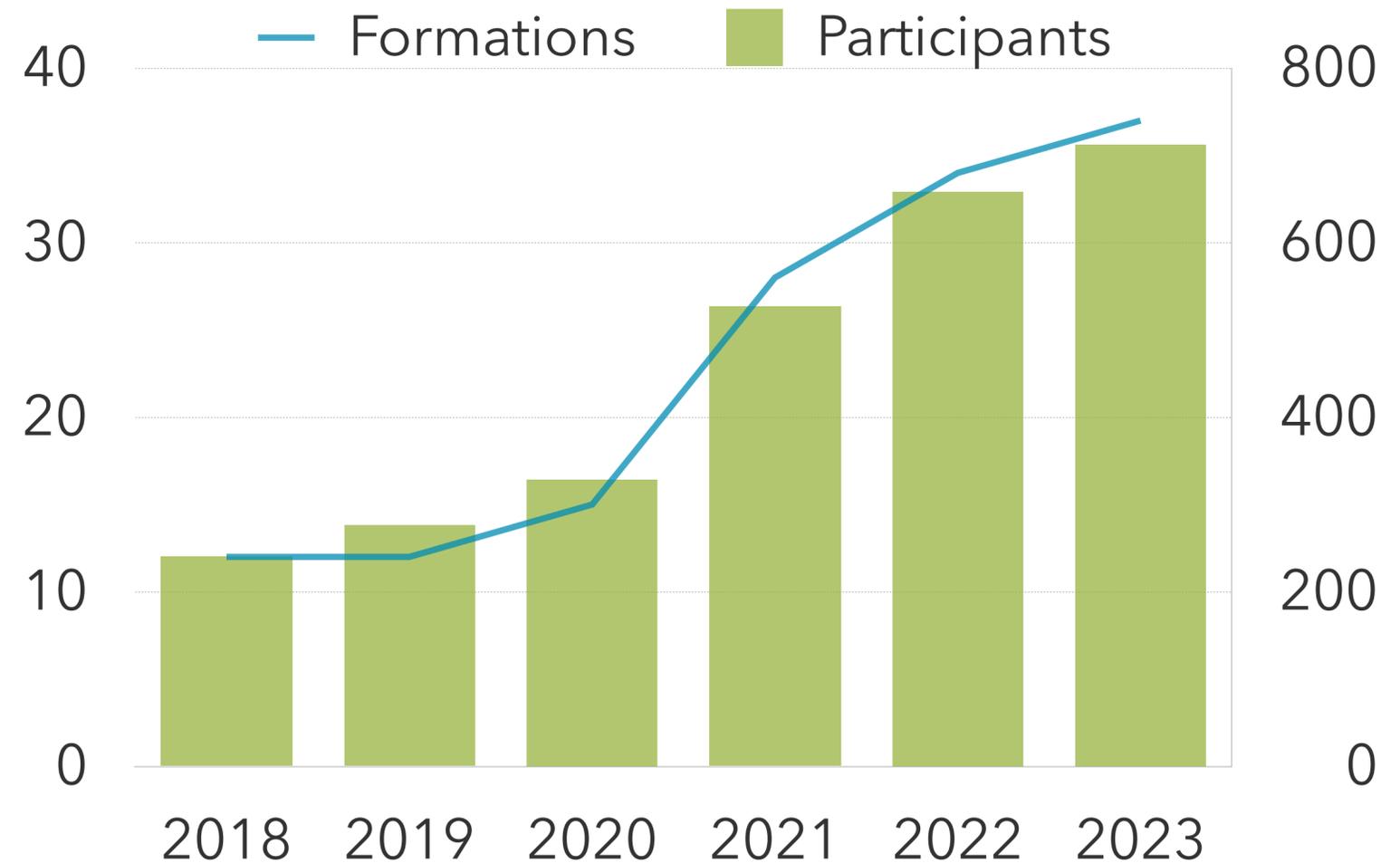
- Allocation de ressources CPU/RAM et stockage
- Délégation de la gestion des membres
- Tableau de contrôle des ressources déployées
- Monitoring pour son groupe (quota, utilisateurs, apps)
- Création d'environnements dédiés (cf. développeurs)

Support des Formations par IFB-Biosphère

Formations hébergées par le cloud

- 30+ sur 2014-2017
- 132 sur 2018-2023
- Certaines récurrentes depuis 2015

	Formations	Participants
2018	12	240
2019	12	276
2020	15	329
2021	28	527
2022	34	659
2023	37	713
2024-09-24	27	571
TOTAL	165	3315



Se connecter au cloud IFB-Biosphère

1. Se connecter avec son identifiant de son organisme, université...
2. Rejoindre son unité d'affiliation, un projet ou une formation

IFB Biosphere RAINBio myVM DATA Support [fr] Se connecter

SE CONNECTER

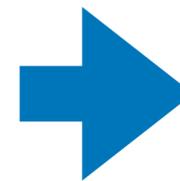
IFB
INSTITUT FRANÇAIS DE BIOINFORMATIQUE

Utilisez vos identifiants de votre organisme (CNRS, INRAE, Inserm, Universités...)

Se connecter

Nous utilisons la fédération d'identité européenne eduGAIN. Si votre organisme académique n'est pas dans cette fédération, vous pouvez utiliser un compte local avec votre adresse professionnelle.

cea CNRS INRAE Inserm INVESTISSEMENTS D'AVENIR



LISTE DES GROUPES

Mes groupes Rejoindre un groupe Créer un groupe

Afficher 10 éléments Rechercher : cnrs

	Nom	Site web	Actions
👤	BioCampus (UAR3426 BioCampus Montpellier), CNRS-UAR3426, INSERM-US009	🔗	+
👤	BISCEm (Biologie Intégrative Santé Chimie Environnement), US 42 INSERM / UMS 2015 CNRS	🔗	+
👤	IGFL (Institut de Génomique Fonctionnelle de Lyon), CNRS UMR5242, INRAE USC1370	🔗	+
👤	Institut Cochin (Institut Cochin U1016 Inserm/UMR8104 CNRS/UMR-S8104), U1016, UMR8104	🔗	+
👤	ISA (Institut Sophia Agrobiotech), UMR1355 (INRA), UMR7254 (CNRS)	🔗	+
👤	LBBE (Laboratoire de Biométrie et Biologie Évolutive), UMR CNRS 5558	🔗	+
👤	LEM (Laboratoire d'Ecologie Microbienne (UMR CNRS 5557 / UMR INRA 1418)), UMR 5557	🔗	+
👤	LMGE (Laboratoire Microorganismes: Génome et Environnement, UMR6023 CNRS-UCA), UMR6023	🔗	+
👤	UFIP (Unité Fonctionnalité et Ingénierie des Protéines), UMR CNRS 6286	🔗	+

Cloud IFB-Biosphère - Bilan 2022-2023

+4,500
vCPUs

	Active Users	Virtual Machines	Apps (new+upd.)	Trainings (trainees)
2022	727	4,800	57 (19)	34 (660+)
2023	903	5,000	63 (14+6)	37 (750+)



Conclusion

L'IFB fournit une infrastructure numérique pour des services bioinformatiques aux communautés françaises des sciences de la vie

- Modalités d'accès
 - ★ accès de base à tous les membres des organismes académiques
 - ★ extension de ressources pour les projets et formations (après évaluation)
- Des ressources de calcul disponibles sur 9 sites
 - ★ technologies logicielles standards cluster et cloud, machines virtuelles et conteneurs
 - ★ matériels HPC/AI : grande mémoire RAM, haute fréquence CPU, grand nombre de coeurs par CPU, GPU
- Un catalogue de services pour les traitements des données biologiques
 - ★ environnements virtuels de recherche développés par les experts
 - ★ modifiables par les utilisateurs
 - ★ interfaces usuelles, Rstudio/Shiny, Jupyter, outils de workflows...
- Des espaces de stockage
 - ★ en mode projet (après évaluation)
 - ★ pas de certification pour les données sensibles (HDS, etc.)

Questions ?

Merci pour votre attention.

Remerciements aux ingénieurs de l'infrastructure cloud IFB-Biosphère

Matéo BOUDET, Guillaume BRYSSBAERT, Micaël CALVAS, Stéphane DELMOTTE, Hervé GILQUIN, Nadia GOUE, David GRIMBICHLER, Jean-François GUILLAUME, Antoine MAHUL, Jérôme PANSANEL, Sébastien PHILIPPOT, Bruno SPATARO, Cyrille TOULET

Ainsi qu'à tou.te.s les collègues de l'IFB et des projets partenaires

Financements

- L'Institut Français de Bioinformatique (IFB) a été fondé par le Programme d'Investissements d'Avenir subventionné par l'Agence Nationale de la Recherche, numéro ANR-11-INBS-0013.
- Ce travail a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre de France 2030 relatifs aux équipements structurants pour la recherche / EQUIPEX+ portant la référence ANR-21-ESRE-0048.



Inserm

INRAE

