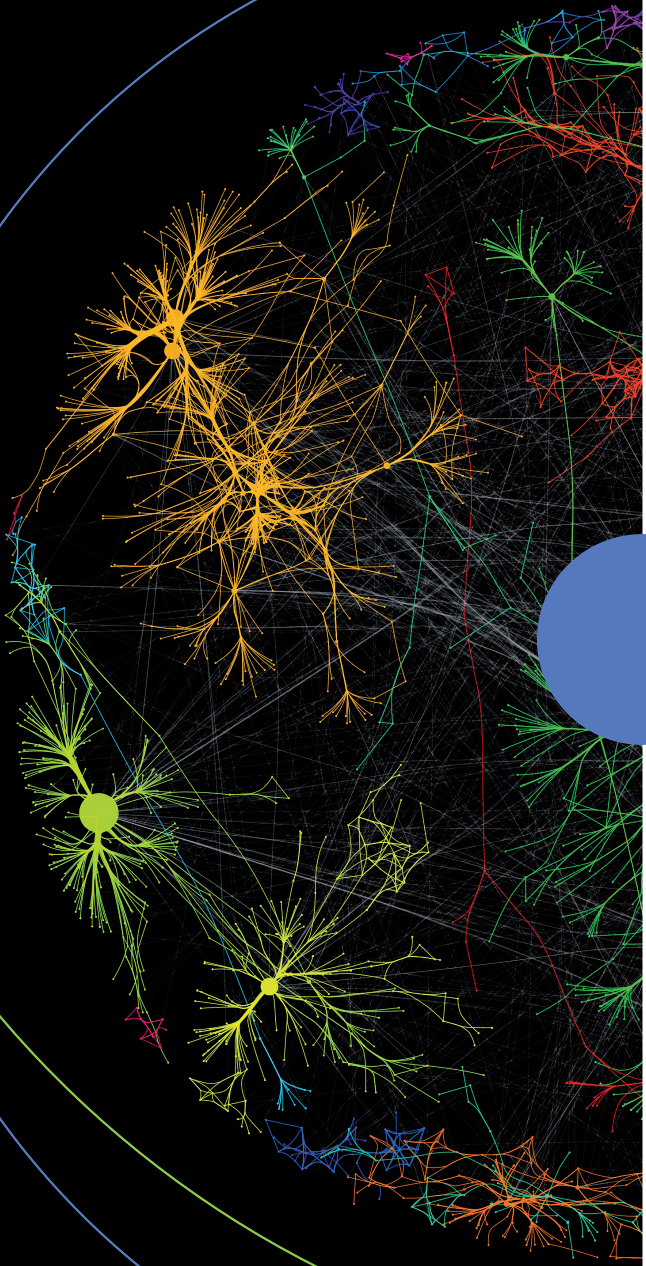
2017-2018

ACTIVITY REPORT



INSTITUT FRANÇAIS DE BIOINFORMATIQUE (IFB)

FRENCH INSTITUTE OF BIOINFORMATICS

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

www.france-bioinformatique.fr

Credits and acknowledgements

Produced by Victoria Dominguez Del Angel,
Olivier Sand and Jacques van-Helden in June 2019.

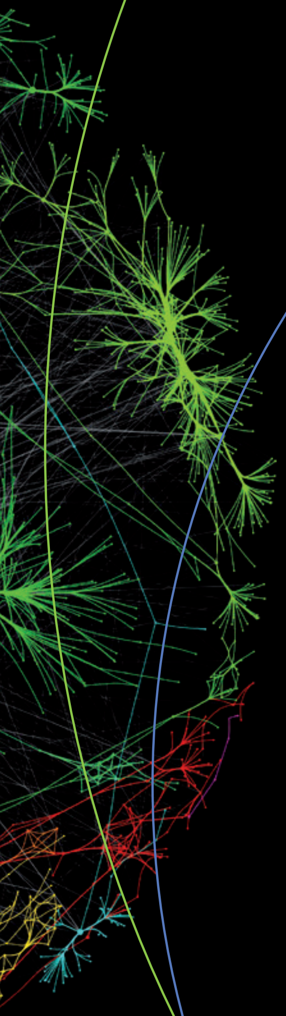
With special thanks to all of those who contributed to this report and the development of IFB, most notably Work Package and Action leads and members of the IFB-Actions

For more information about IFB please contact contact@groupes.france-bioinformatique.fr

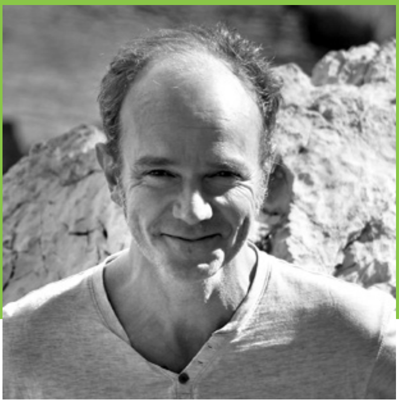
Art direction and design: Emmanuelle Maupoux

IFB is funded by the Programme d'Investissements d'Avenir (PIA), grant Agence Nationale de la Recherche number ANR-11-INBS-0013.

_TABLE OF CONTENTS



_Foreword from IFB directors



Bioinformatics is a key component of life sciences – transversing several disciplines, thematic domains and applications and intervening at each step of research projects: experimental design, raw data processing, extraction of relevant information, interpretation and representation of the final results. Indeed, technological revolutions foster a drastic change of scale in the current way of deciphering biological mechanisms and designing applications for health, agriculture, the environment and biotechnology.

Research projects are conceived with a holistic perspective – relying on multiple omics approaches (genomics, transcriptomics, proteomics, metabolomics, metagenomics etc.) to characterise all the elements of biological systems and the interactions between them. Furthermore, interpreting this molecular-level data requires its integration with other types of massive data (images, phenotypes, medical data etc.) breaking the boundaries between the different scales of analysis (molecule, cell, organism, population and ecosystem). Bioinformatics addresses the multiple challenges of life sciences: deploying adapted infrastructures to ensure the transfer, storage and computation of massive data; developing innovative approaches for the integration of heterogeneous data; ensuring open and reproducible science; and addressing the societal challenges raised by multiple applications of these technologies.

The Institut Français de Bioinformatique (IFB) is one of the 22 French national research infrastructures that have been created to address these challenges in the context of the Programme Investissement d’Avenir (PIA) launched by the Research Ministry in 2010. Granted a €20 million budget, IFB began to operate in summer 2013. Following the midterm evaluation report (autumn 2016), a new work plan was prepared under the co-leadership of Claudine Médigue (Research Director at the Centre National de la Recherche Scientifique) and Jacques van Helden (Professor at Aix-Marseille University). The plan was submitted for external evaluation, and the revised version was definitively validated in May 2018. The current report describes the results achieved by IFB in the context of this restructuration.

The recruitment for the new actions started progressively from July 2018 and is still ongoing. During the intermediate period when the IFB-core team worked actively on the restructuration (January 2017-May 2018), IFB platforms pursued their individual activities. The report highlights the current progress of the 2018-2021 work plan, as well as the achievements of the 398 engineers and researchers working on the 31 IFB bioinformatics platforms that span the national territory.

IFB activities are classified into two main objectives: deploying a high-performance service in bioinformatics optimised for the needs of the life sciences communities and developing innovative bioinformatics resources to address strategic challenges in health, the environment and agriculture.

Services to the communities are ensured by the **National Network of Computational Resources for bioinformatics (NNCR)**, which regroups high-computing systems spread over the national territory and offers a computing and storage architecture suited to the needs of bioinformatics and equipped with a software environment (middleware, generic and specialised tools), ensuring access to biological data. The NNCR is supported by a **mutualised task force** built on the joined expertise of the regional platforms of IFB.

The main **innovation axis** addressed by the 2018-2021 work plan is focused on **integrative bioinformatics**. IFB started to develop an innovative bioinformatics service to address the challenging task of integrating data of heterogeneous nature and rapidly evolving tools in the context of research projects in the environment, agronomy and health. Several pilot projects in integrative bioinformatics are ongoing and have stimulated unprecedented cooperation between 10 National Infrastructures for Life Sciences and two cohorts. IFB has also co-organised, with Aviesan Multi-Organism Thematic Institute for Genetics, Genomics and Bioinformatics (ITMO GGB), a one-day workshop titled *Challenges and Perspectives in Integrative Bioinformatics*, which regrouped ~100 participants. Moreover, we set up a new two-months training programme on integrative bioinformatics (one month for courses + one month for a personal project on an IFB platform), which started in February 2019.

As the French node of the **European bioinformatics infrastructure ELIXIR**, IFB contributes to, and benefits from, the exceptional dynamism of this collaborative network to pioneer technological developments, take part in the specification of international standards and promote good practices for open and reproducible science. Noticeably, in November 2018 ELIXIR-FR and the ELIXIR Hub co-organised the first **European BioHackathon**, which gathered 150 bioinformaticians from all around the world to collaborate on 28 hacking projects in five days. The event led to remarkable achievements.

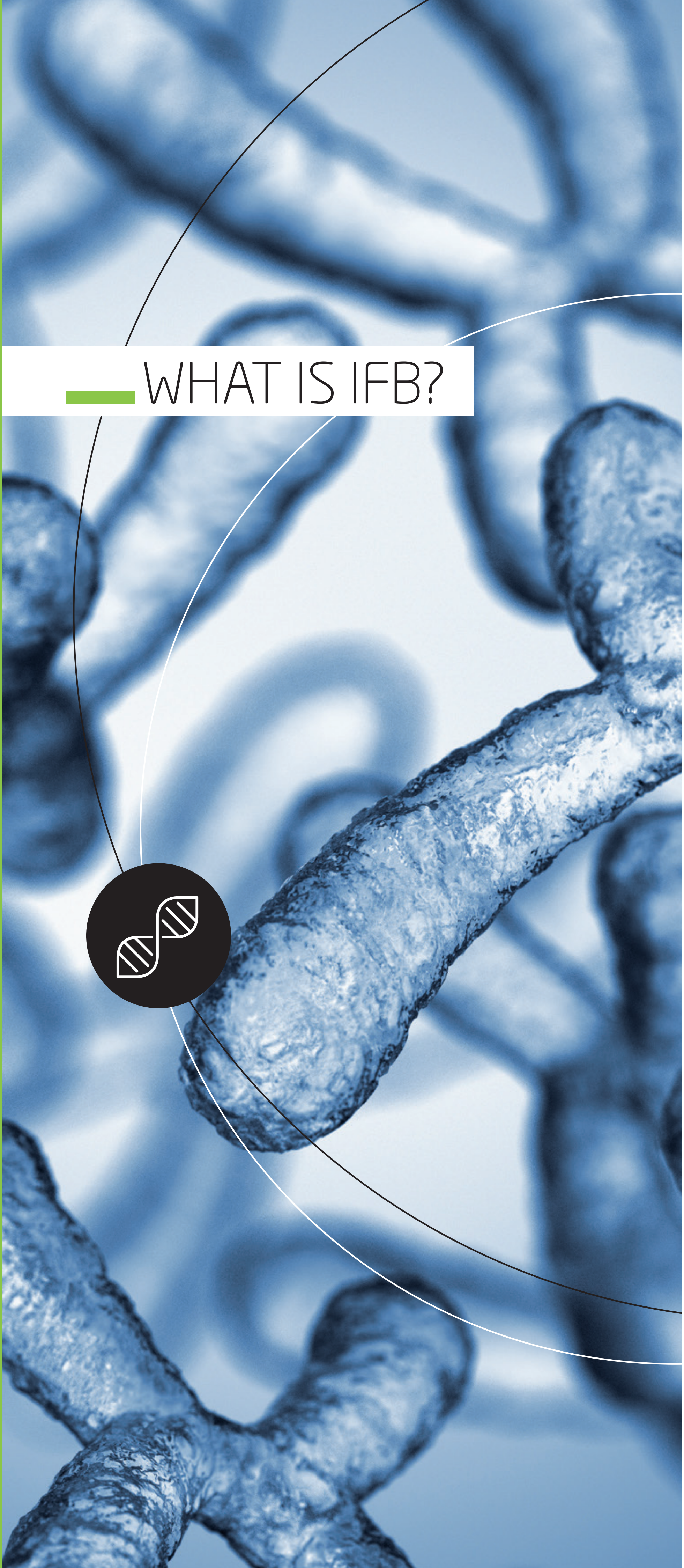
The ongoing transformation of life sciences raises an **increasing demand for training** – issuing, on the one hand, from experimental life scientists who feel the absolute necessity to acquire advanced skills in bioinformatics to stay in line with the rapid evolution of the field and, on the other hand, from bioinformaticians themselves, who need to cope with the permanent emergence of new concepts, approaches and tools.

Beyond the first results presented in this report, IFB priorities for 2019 include the following actions: carry out a **broad user survey** addressed to all our target communities, academic as well as industrial users; launch IFB-driven strategic actions bringing integrative bioinformatics to **applications for health and the environment**; deploy user-friendly resources to accompany life scientists and bioinformaticians in the design of their **data management plans (DMPs)**; deepen the extraction of relevant information from big data by **adapting Artificial Intelligence (AI) approaches to the treatment of heterogeneous data** (e.g., molecular + imaging + phenotypic data); develop solutions for **sensitive data** (e.g., by transferring the software to the data rather than the opposite); connect the work of the IFB ethical committee with the other national and international initiatives addressing the **ethical issues at the intersection of biology and data science; and ensure** the sustainability of the national infrastructure by enforcing its recruitment policy and diversifying the sources of income.

The challenges and expectations are huge, and we certainly have a long way ahead before reaching all of them. Nevertheless, one year after the full start of the 2018-2021 work plan, our collaborative network of platforms and people have already passed several important milestones and achieved a significant number of results presented in this report. We would like to thank our funding and partner institutions (the list is on the back cover) and express our heartfelt gratitude to all IFB members – especially the leaders of all the ongoing work packages and actions, as they are playing a key role in the new IFB missions.

Claudine Médigue
and Jacques van Helden
Co-directors of the Institut Français de Bioinformatique (IFB)

WHAT IS IFB?



IFB at a glance

www.france-bioinformatique.fr/about-us

GOALS

The goal of the French Institute of Bioinformatics (Institut Français de Bioinformatique, IFB) is to deploy a national infrastructure of bioinformatics accessible to the communities of medical and life sciences research. The IFB infrastructure includes a physical infrastructure (storage & computing), generic and specialised software environments, biological databases, user support, consulting for the conception and realisation of scientific projects and training.

IFB should also keep up with the rapid evolution of the technologies and needs, promote the adoption of standards and the establishment of reproducible science and anticipate the future developments of the domain. The innovation axis of the 2018–2021 action plan dedicates a special effort to the development of integrative bioinformatics. Moreover, IFB is the French node of the ESFRI ELIXIR, which aims to develop a European infrastructure in bioinformatics.

MISSIONS

① Provide core infrastructure of bioinformatics services to the French life sciences community (academic labs and industries), namely tools and information technology (IT) facilities dedicated to the management and analysis of biological data.

② Furnish training in bioinformatics for life scientists, medical communities and IT.

③ Bring support to biology, health, agronomy and environmental research programmes, and access to high-level expertise and mutualised competences.

④ Develop a strategic vision to maintain French research at the highest level of expertise in the analysis of biological and medical omics data.

⑤ Hand over access to the newest state-of-the-art technology available internationally in the field of bioinformatics to the scientific community and use this as a lever for ambitious national research projects in life sciences and their applications.

⑥ Effectively interlink people and organisations in the field of bioinformatics and data, software, services, publications and projects related to bioinformatics.

⑦ Ensure the international representation of the French bioinformatics community, especially in the context of the European ELIXIR network. IFB is supporting users and partnerships to make their discoveries findable, accessible, interoperable and reusable (FAIR) in order to extract full values from scientific insights.

IFB VALUES, **which align with our commitments to our members and partners, include the following:**

To produce a high level of results with our resources and services.

To have a scientific focus and provide quality services.

To behave ethically with intellectual honesty. This includes our daily work and our relationships with others. We promote a culture of trust and responsibility.

To be an inclusive and participative community. This must be reflected in our way to plan delivering and sharing our services and sciences. We also have a strong link with open and reproducible sciences.

IFB is also attached to the United Nations Sustainable Development Goals[2]. From these 17 Sustainable Development Goals, the IFB missions are strongly related to:

Goal 3: ensure health and well-being.

Goal 4: quality Education. Specifically with a substantial increase in the number of youth and adults who have relevant skills, including technical and vocational skills, for employment, decent jobs and entrepreneurship.

Goal 7: affordable and Clean Energy. In particular, to promote access to clean energy research and technology, including renewable energy, energy efficiency and advanced and cleaner fossil-fuel technology.

Goal 9: build resilient infrastructure, promote sustainable industrialisation and foster innovation.

All these values and commitments help us to bring sciences and research closer to societal needs.

---

[1] www.elixir-europe.org/

[2] www.un.org/sustainabledevelopment/sustainable-development-goals/

STRUCTURE OF THE IFB

www.france-bioinformatique.fr/en/elixir-fr

Today, the IFB consists of:

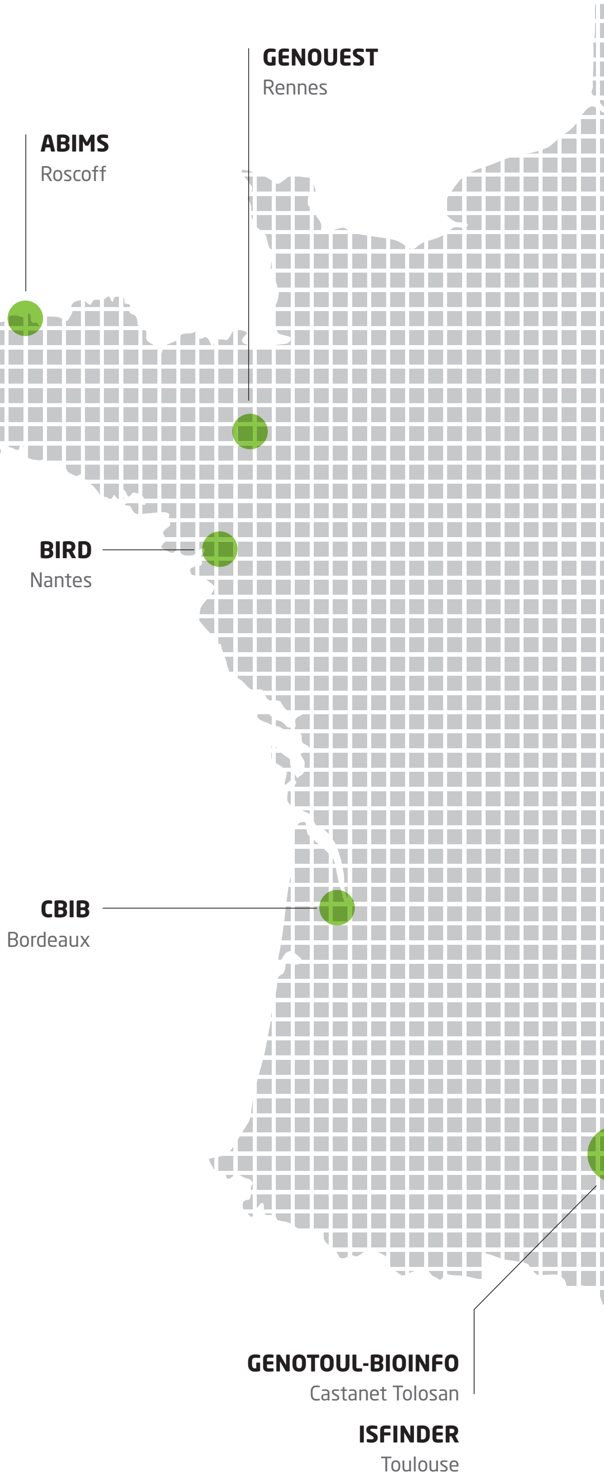
- **Thirty Regional Platforms (PFs)** spanning the entire territory (Figure 2) belonging to the main French research organisations (CNRS, CEA, INRA, INRIA, INSERM, IRD and CIRAD[3]) universities and the Curie and Pasteur Institutes,
- a national hub called IFB-core, located at Genoscope, Evry (since August 2018). It is a joint-service unit depending on five supervisory authorities (CNRS, CEA, INRA, INSERM and INRIA). IFB-core is the French node for the European bioinformatics infrastructure ELIXIR[4].

A detailed description of each platform, their areas of expertise and the services they offer is available in the 'IFB platforms' section of this report and on the IFB website[5]. **There are 388 people working for the IFB.**

Regional PFs provide services and technical support to scientific projects of their local life sciences community. They also provide training, develop code for the life sciences community and share a part of their IT infrastructure (cf. 'Federation of computer clusters and cloud' section). Regional PFs are embedded in bioinformatics research laboratories and are thus able to preserve a critical link between provisions of service and research activities.

The geographic location of the 30 regional IFB platforms.

Figure 2

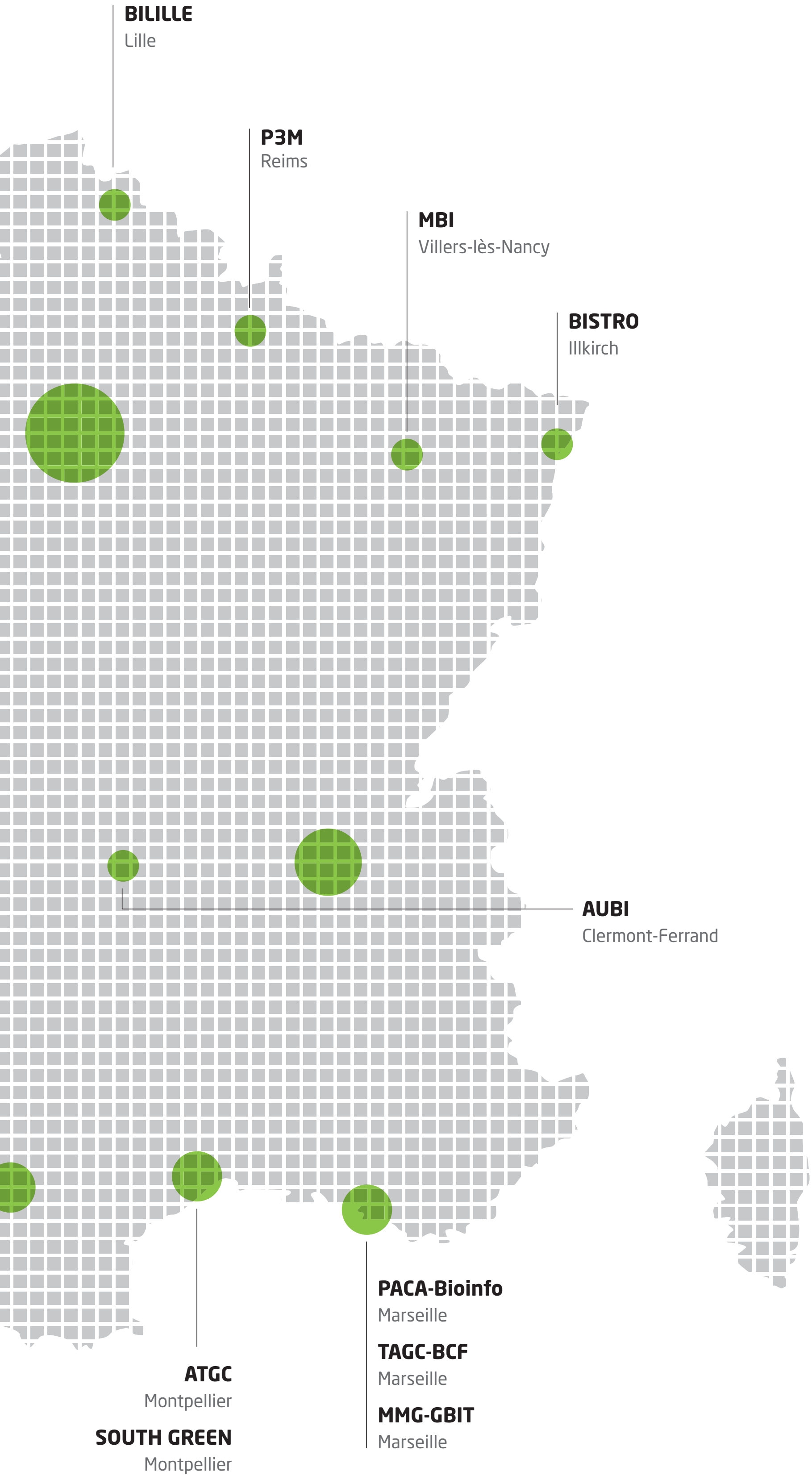


---

[3] **CNRS:** National Center for Scientific Research;
CEA: Alternative Energies and Atomic Energy Commission;
INRA: National Institute for Agriculture Research;
INRIA: National research institute for the digital sciences;
IRD: National Research Institute for Development
INSERM: National Institute for Health and Medical Research;
CIRAD: French Agricultural Research Centre for International Development

[4] www.elixir-europe.org/

[5] www.france-bioinformatique.fr/



RÉGION PARISIENNE

EBIO
Orsay

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

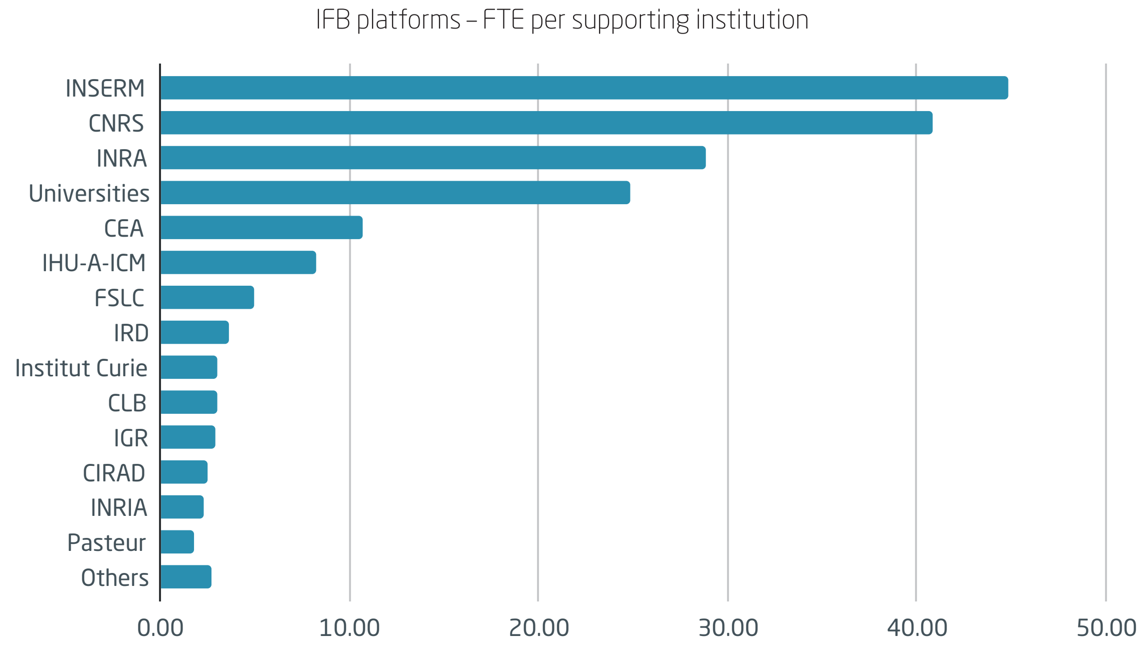


Figure 3 Involvement of IFB supporting research institutions in the staff of the national infrastructure. Note that the Inserm personnel is in large majority assigned to the OrphaNet project, which became an ELIXIR core data resource.

Web Site	www.france-bioinformatique.fr
Computing facilities (compute, storage and software)	contact-nncr@groupes.france-bioinformatique.fr
Training	contact-formation@groupes.france-bioinformatique.fr
Industry	industrie@groupes.france-bioinformatique.fr
General contact address	contact@groupes.france-bioinformatique.fr

THE ADDED VALUE OF A NATIONAL RESEARCH INFRASTRUCTURE

www.france-bioinformatique.fr/en/infrastructure

IFB has a fundamental role in the mutualization of physical, logistics and human resources to address user's needs, to promote technological developments, to enforce technology and scientific watch at the national and European levels (via ELIXIR), to set up training in line with the constant evolution of the domain, and to provide user support.

The IFB work plan is organised around a mutualised task force gathering various experts of the IFB platforms to optimise the sharing of knowledge and competencies and design of robust and innovative solutions and deploy IFB services. **This synergistic federation of bioinformatics regional platforms is clearly an important added value of the IFB infrastructure** relative to the sum of individual platforms (the result is more than the sum of the parts).

IFB also plays a structural role as the national node in the ELIXIR bioinformatics infrastructure. Our contribution to ELIXIR is extremely important for the following reasons:

- gain further competences in order to boost our competitiveness at the international level;
- be source of proposals and interact with the other ELIXIR Nodes for the definition of international standards;
- promote the adoption of internationally defined standards and good practices by the whole French bioinformatics community and;
- increase in efficiency in implementing national services for the large panel of user communities.

At the national level, IFB contributes towards increasing significantly the visibility of bioinformatics activities and resources by acting as an overall coordinator to build a strategic vision of national coherence. In this context, IFB plays an essential role for the French platforms and research laboratories:

- Creation and animation of working groups around user communities connected to ELIXIR communities (plants, marine metagenomics, human data etc.).
- Mobilisation of the community of bioinformaticians and coordination of technological projects funded by ELIXIR. IFB is also able to optimise the French integration in European projects coordinated by the ELIXIR Hub or by research communities which are partners of IFB.
- Integration of national resources in the ELIXIR-FR Service Delivery Plan (SDP) and guidance and support to become 'ELIXIR core resource'.

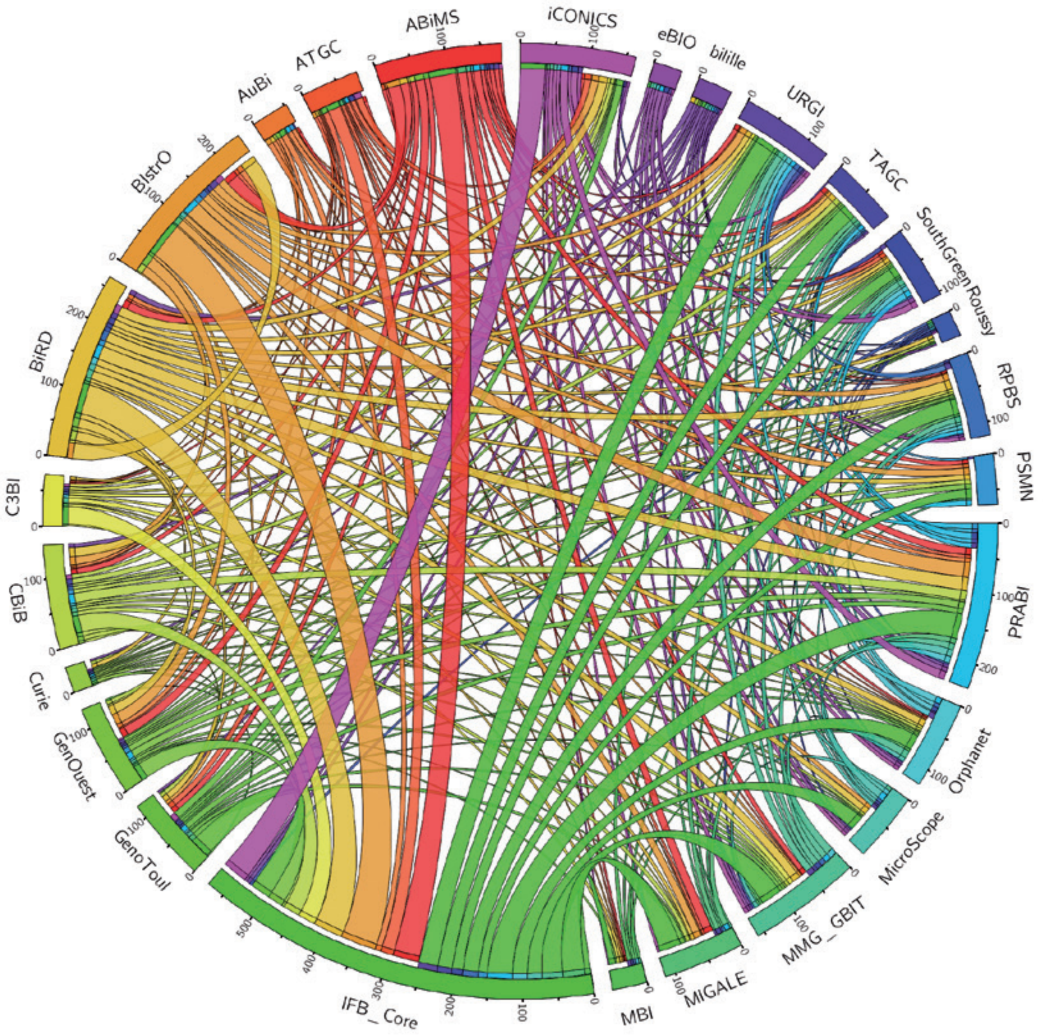
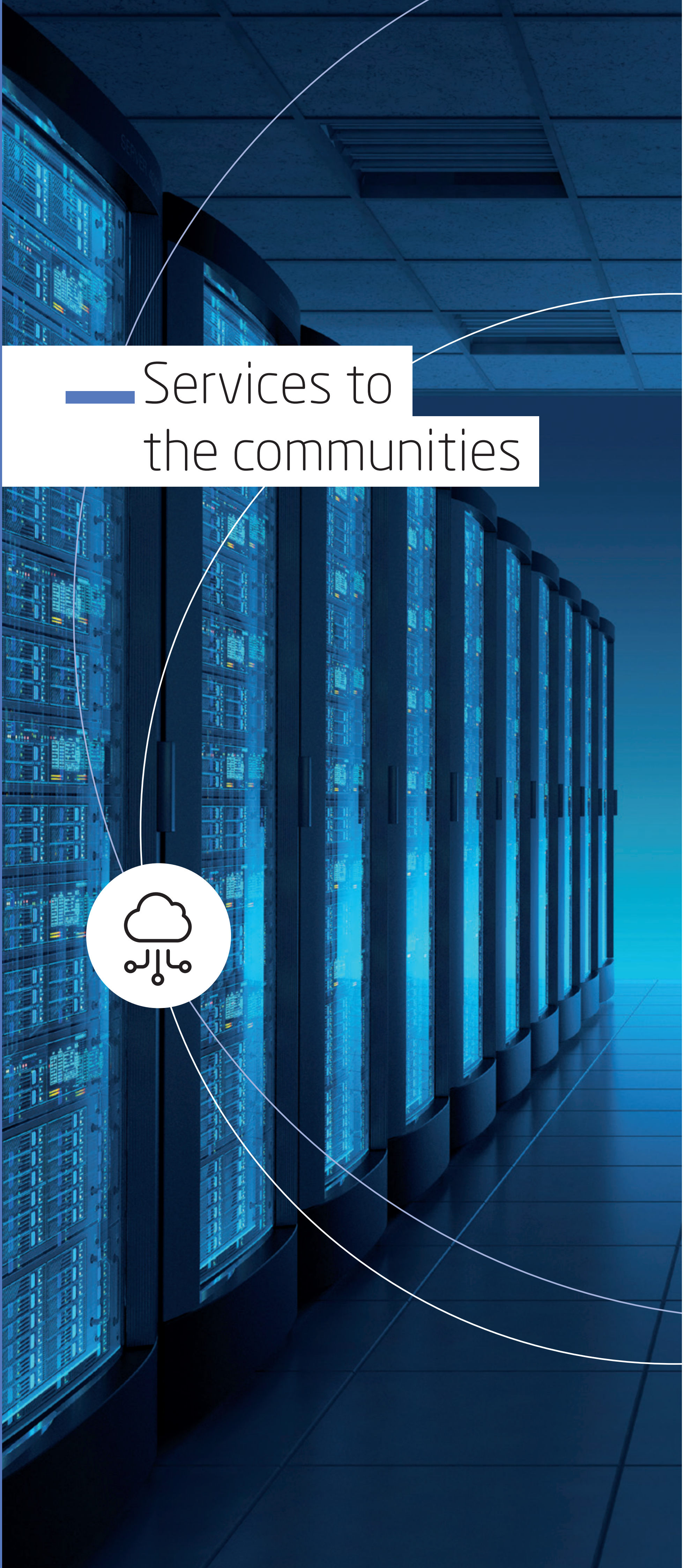


Figure 4 The IFB platform-platform interaction network. The interactions represent ongoing collaborations between IFB platforms on specific areas such as training, tools and other IFB actions. Each platform has a specific colour. The width of the ribbons is proportional to the number of actions where platforms interact.

Services to the communities



A national network of bioinformatics resources

www.france-bioinformatique.fr/en/nncr

THE CHALLENGE OF BIG DATA IN LIFE SCIENCES

In comparison with data treatment in other scientific areas (physics, chemistry, earth sciences etc.), particular features of biological data make them dependent on highly specific hardware architecture, software environment and data collections that are currently not available in the national computing centres.

- Current research projects involve the integration of different data types of heterogeneous nature (genome, transcriptome, epigenome, protein interactions structure etc.).
- These data collections (e.g., sequenced genomes) need to be regularly updated.
- A large variety of programming languages (Bash, Python, Java, R, C etc.) and software (e.g., there are at least 98 software tools for sequence read mapping on a genome) – are used.
- The analyses require the chaining of a diversity of steps into specialised pipelines or workflows.
- Each software tool comes with its many dependencies (libraries, versions), and these dependencies are sometimes incompatible between tools used in the same workflow.
- Biologists express a strong demand for user-friendly interfaces (web, GUI) at different levels of the computing chain: conception of workflows and visualisation of the results.
- Parallelisation strategies rely on segmenting the data rather than the computing tasks.

An important mission of IFB is thus to conceive, implement and deploy highly specialised computing facilities suited to address these challenging requirements.

A network of computing resources: Clusters and clouds

To face these challenges, IFB organised the services through decentralisation of physical, logistics and human resources. The **National Network of Computing Resources (NNCR)** includes all the hardware (IFB-core + regional platforms) contributing to ensure the service offer. NNCR relies, on the one hand, on the separation of IFB-core servers (the cloud in Lyon, the cluster in Orsay) and, on the other hand, on the distribution of the services on eight high-performance computing platforms localised in different regions.

Such a decentralisation of physical resources offers at least four main advantages:

① security of the service in case of local breakdown;
② distribution of data transfer flows;
③ geographic proximity;
④ joint development of good practices and advanced technical expertise for computing, processing and storage of data.

Servers	Location	Compute (#CPU HT*)	Storage (#TB)	RAM (#TB)
ifb-core-cluster	Orsay (IDRIS)	4 000	400	20.01
Abims	Roscoff	1 928	2 000	10.60
Genotoul	Toulouse	6 128	3 000	34.30
GenOuest	Rennes	1 824	2 300	7.50
Bird	Nantes	560	500	4.00
Migale	Jouy en Josas	1 084	350	7.00
TOTAL FEDERATION OF CLUSTERS		**15 524**	**8 550**	**83.41**
ifb-core-cloud	Lyon (CC-IN2P3)	3 936	408	20.41
ifb-genouest-genostack	Rennes	600	350	2.60
ifb-prabi-girofle	Lyon	448	144	1.50
ifb-bird-stack	Nantes	160	50	0.45
ifb-bistro-iphc	Strasbourg	85	50	1.50
TOTAL FEDERATION OF CLOUDS		**5 229**	**1 002**	**26.46**
GRAND TOTAL (clusters + clouds)		**20 753**	**9 552**	**109.87**

Figure 5 The current capacity of IFB National Network of Computing Resources (March 2019).

A FEDERATION OF CLOUDS

www.france-bioinformatique.fr/en/cloud

IFB has set up a federation of clouds named Biosphere[7], which provides distributed services to analyse life sciences data, and relies on interconnected IT infrastructures of the IFB hub and several regional platforms (Figure 5).

Biosphere[7] provides a unified way to deploy a bioinformatics environment in several clouds from the same desk – which helps, for example, to combine large CPU resources or guarantees access to available cloud resources across the different sites. Biosphere is used for scientific production in the life sciences, for methods and software developments and to support events like cloud and scientific training sessions, hackathons or workshops.

---

[7] https://biosphere.france-bioinformatique.fr/

The Biosphere portal (https://biosphere.france-bioinformatique.fr) provides different high-level interfaces to the distributed cloud infrastructure to analyse life sciences data:

- The **RAINBio catalog**[8] offering a catalogue of appliances ready to run in one click and preconfigured machines tailored for specific bioinformatics tasks.
- A **web dashboard**[9] to manage cloud deployments with single virtual machines (VMs) or complex applications with multiple VMs.
- A **data centre**[10] where you will find common public data available in the clouds.

IFB cloud federation was established first at the end of 2016 and now gathers five French academic clouds deployed in the IFB hub and regional IFB platforms (IFB-core, GenOuest, PRABI-LBBE, BiRD, BIstrO). IFB-Biosphere currently offers an infrastructure consisting of over 5,200 cores, summing up about 26,5 TB of RAM and 1 PB of storage (see Figure 5 for details). Other French academic clouds are under deployment and will be added to the federation. The national IFB-core cloud is hosted by the national supercomputing centre of **CNRS IN2P3** department, using benefits from the skilled teams there to take care of system, hardware and cloud administration (with Openstack software) for the IFB. The day-to-day management of life sciences applications and cloud users is devoted to the IFB cloud team.

A FEDERATION OF CLUSTERS

www.france-bioinformatique.fr/en/cluster

The IFB Core Cluster has been in production since November 2018. It consists of 2,000 cores (4,000 cores hyperthreaded) and 400 TB of storage. Its aim is to provide access to general and specialised bioinformatics resources to users of different fields (biologists and bioinformaticians) and levels of expertise (from novice to expert).

To reach all of them, we are setting up different user interfaces such as Secure Shell access and web-interfaced servers (e.g., RStudio, Galaxy). The IFB Core Cluster will also provide a computing environment to deploy bioinformatics resources developed by the French community of bioinformaticians (e.g., Regulatory Sequence Analysis Tools, phylogeny.fr, deliverables of the IFB pilot projects etc.) and will support other resources on request of their developers.

The software collections are mainly accessible via Conda environments and Singularity images. A **Software on demand** service is supported either via our community platform[11] or by directly proposing a merge request on a shared Git repository (Conda[12], RStudio[13]), depending on user skills. Indeed, each tool needs to be submitted, reviewed and installed by our **Continuous Integration (CI)** mechanism.

Finally, shared reference databanks (GenBank, UniProt, UCSC, Ensembl etc.) are available, indexed in different formats (Blast, Bowtie, BWA etc.). In the near future, databank downloads, updates and indexation will be managed using BioMaj (biomaj.genouest.org).

To promote an open dialogue between users and contributors community, we chose to provide user support via a Discourse-based community platform[14], where requests can be treated in either of two modalities:

- using an open community forum where users or external experts can contribute or
- contacting a user helpdesk, with private exchanges if required.

Upstream of each training session, we collect teachers' requirements for tools, banks and student accounts. Shared project folders are created to enable trainers, tutors and trainees to work in same and private workspaces. Finally, for each training session, a customised software environment is created (Conda environment), which ensures the reproducibility of the results, reusability of the teaching material for other training events and portability of the whole software environment (e.g., on the computing facilities of the trainees' institutes).

The administration of the cluster itself is done collaboratively. More than six engineers (two FTEs) from five IFB platforms build and daily contribute to the project. In order to handle the multiple contributions, they are managed by CI (Ansible + GitLab runner) mechanisms connected to a common code repository (GitLab). Besides the IFB Core Cluster, a federation of regional cluster infrastructures, sharing common practices and access usage, also provides storage computing resources.

---

[8] https://biosphere.france-bioinformatique.fr/catalogue/

[9] https://biosphere.france-bioinformatique.fr/cloud/

[10] https://biosphere.france-bioinformatique.fr/data/biomaj

[11] https://community.cluster.france-bioinformatique.fr

[12] https://gitlab.cluster.france-bioinformatique.fr/taskforce/conda-env

[13] https://gitlab.cluster.france-bioinformatique.fr/taskforce/rstudio-packages

[14] community.cluster.france-bioinformatique.fr

Modular software environments to address every user's needs

Tools, containers and workflows will be available in both middleware infrastructure cluster and cloud via various interfaces suited for diverse users: websites, stand-alone applications with graphical user interfaces (via a virtual desktop), VMs, command line tools, application programmatic interfaces and web services.
This environment helps developers to ensure best practices in software development and deployment and support users in finding, accessing and deploying these resources. IFB is also one of the main contributors of the European ELIXIR tools registry bio.tools[15].

THEMATIC CONDA ENVIRONMENTS

To manage its software environments, IFB mainly uses Conda (https://conda.io). This package, dependency and environment manager, has been well adopted and has become a standard in life sciences. It allows to create ready-to-install environments for diverse operating systems (Unix, Mac OS X, Windows). The main actor in the distribution of packages for life sciences – the Bioconda channel (bioconda.github.io) – gathers more than 600 contributors and provides more than 3,000 Conda packages.

Based on this huge software resource, we deployed, on the IFB cluster, a series of standard thematic environments for the mainstream requirements currently expressed by our user and trainer teams: next-generation sequencing (NGS) (RNA-seq, ChIP-seq, DNA-seq), statistical analysis with R, integrative bioinformatics and so forth. In addition, we offer a **service of software on demand**, enabling individual users to deploy a customised environment combining the specific tools required for their own projects. When a software tool is not available under Conda, we package them in order to answer our users' needs and expand the internationally shared resources.

CONTAINERS AND WORKFLOW MANAGEMENT ENVIRONMENTS

🌐 **https://biocontainers.pro**

Software containers are a key technology to enable the rapid deployment of software, facilitate the local installation of bioinformatics software and combine tools into powerful analysis pipelines. In October 2017 IFB, EBI (European Bioinformatics Institute) and the ELIXIR Tools Platform organised a Hackathon[16] in Paris to consolidate a container platform as a service named Biocontainers

The Biocontainers platform provides guidelines to create, manage and distribute bioinformatics packages (e.g., Conda) and containers (e.g., docker, singularity). BioContainers is based on the popular frameworks Conda, Docker and Singularity inside a community platform (Github) that allows anyone to contribute.

> The Biocontainers platform currently proposes 7 600 tools, 17 700 tool versions and 54 900 containers. This platform is co-led by the IFB platform GenOuest (www.genouest.org), the European Bioinformatics Institute (www.ebi.ac.uk) and de.NBI (www.denbi.de).

---

[15] 🌐 https://bio.tools/

[16] 🌐 https://github.com/BioContainers/workshops

WORKFLOW DESIGN AND USAGE

Workflow management systems are available to ensure the reproducibility of complex analyses (e.g., Taverna, Snakemake and Nextflow) with multiple computational steps.

One of the most important workflow management systems for biomedical communities is **Galaxy**: a free and easy-to-use web platform that enables users to connect bioinformatic tools into a workflow without any prior programming expertise, share their analyses with others, and enable others to repeat the same analysis. IFB has developed important thematic frameworks on Galaxy such as Workflow4Metabolomics[17] (developed in collaboration with MetaboHub) dedicated to metabolomics data, ProteoRE for the proteomics community or NGphylogeny for phylogenetic analysis. These thematic instances reach hundreds of users and serve as showcases for the original tools developed by bioinformaticians; they have already achieved a strong national and international reputation. They provide a specialised service, often unique, to user communities in specific fields of life sciences.

IFB should ensure the sustainability of such thematic instances in response to requests from end users and research institutes. The IFB Core Cluster Task Force plans to host a single Galaxy instance with specific subdomains for the different thematic instances mentioned above, following the model implemented on usegalaxy.eu. IFB co-leads the ELIXIR Galaxy community[18].

BIOSPHERE AND RAINBIO

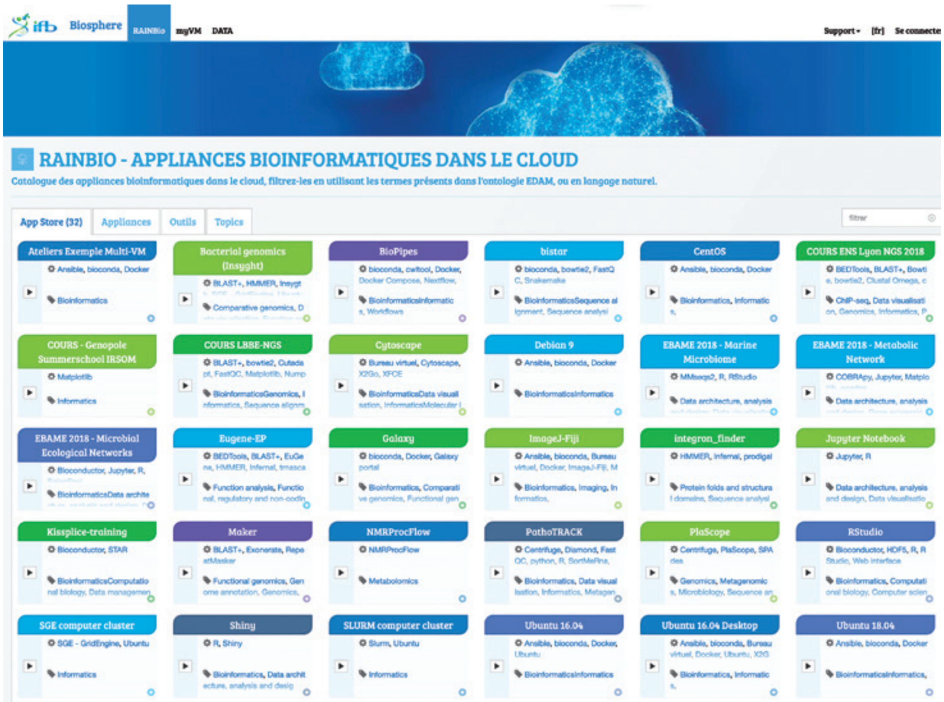
https://biosphere.france-bioinformatique.fr/

The bioinformatics cloud appliances are offered in several flavours, allowing life scientists and bioinformaticians to choose the best configuration suited for their analyses. Currently, there are 27 models of virtual bioinformatics environment developed by some members of the IFB community and recorded in the Biosphere catalogue (RAINBio[19], Figure 6). Some of these virtual environments are providing (i) base environments with IT tools such as Conda – with the Bioconda and R channels, Docker container technology or the Ansible system for software deployment; (ii) high-level scientific gateways with web portals – such as RStudio, Jupyter Notebooks or Galaxy – and GUI interfaces usable through a remote Linux desktop; (iii) complex multicomponents environment such as computing clusters (SGE and Slurm) or workflow engines (Nextflow, Snakemake and CWLtool).

Scientists and engineers can launch their own virtual environments with reserved resources (from 1 CPU with 2 GB of RAM to 128 CPUs with 2,900 GB RAM in a single VM or hundreds or thousands of CPUs and terabytes of RAM in many VMs) that can be adapted to their own needs without interfering with other users. The activity is steadily growing, and the French IFB cloud has already been used for numerous scientific analyses (even intensive analyses with 4,000 cores) and training sessions, workshops, hackathons or several recurring sessions of scientific schools or universities.

Figure 6

Appliances available on RainBio



RAINBIO - APPLIANCES BIOINFORMATIQUES DANS LE CLOUD

Catalogue des appliances bioinformatiques dans le cloud, filtrez-les en utilisant les termes présents dans l'ontologie EDAM, ou en langage naturel.

---

17 http://workflow4metabolomics.org

18 https://www.elixir-europe.org/communities/galaxy

19 https://biosphere.france-bioinformatique.fr/catalogue/

User support

CATALOGUE OF FRENCH RESOURCES IN BIOINFORMATICS

IFB will ensure the findability and visibility of bioinformatics resources (tools, databases, individual expertise, platforms, events and training materials), which will be synchronised with related international catalogues (e.g., ELIXIR bio.tools, Biocontainers and TeSS). This catalogue will also contribute to readily identifying the most adapted resources and their availability to support end-user needs (which tool does what and where is it installed? Who is knowledgeable regarding a given question?).

HELP DESK

https://www.france-bioinformatique.fr/en/helpdesk

The aim of the help desk is to orient users on all the services offered by the IFB. Requests are evaluated in order to identify the most relevant resources (human and computational) available on the French territory, based on the IFB catalogue of bioinformatics resources. The help desk also monitors the activity of the different computing systems, estimates their availability and follows the treatment of their demands.

ACCESS TO THE NNCR

https://www.france-bioinformatique.fr/en/infrastructure

NNCR provides resources (computing, storage) as well as services (support, account opening, software and databases installation), hosting of websites and VMs, development of software/databases/information systems and data bioanalysis). The NNCR is open to public and private French and international research communities in life sciences. Most of the NNCR infrastructures offer a yearly, free-of-charge access to computing and storage with limited resources (1 TB of storage and a few thousand computing hours). Charges may apply for bigger storage options and access to specific computing infrastructure (GPU, big memory etc.)

SUPPORT TO DATABASES

Since the early days of the field, data/knowledge bases are important aspects of bioinformatics. A lot of retrieval systems, collections and websites have been developed since, and this area is characterised by its heterogeneity, both in the biological questions covered and the technologies used. Therefore, IFB engaged a process to label French databases in ELIXIR Service Delivery Plan (SDP), in order to highlight the number and quality of the database resources available.

MUTUALISATION OF RESOURCES WITH OTHER NATIONAL INFRASTRUCTURES FOR LIFE SCIENCES AND HEALTH

The landscape of National Infrastructures in Biology and Health (INBS) is extremely complex and unequal. There are strong differences between those infrastructures regarding their means, number of users and own storage and computing resources, if any.

In such a context, IFB is working on the sharing of resources with the other infrastructures. That work is split between several concomitant actions.

- Animating a working group composed of two IFB members and representatives of each infrastructure having bioinformatics needs.
- Interviewing each representative of these infrastructures in order to understand their operating mode, means, needs, potential and will to share on bioinformatics aspects.
- Set up a service of shared storage and computing project space, available in 2019. These spaces will be open during the lifetime of funded projects and will allow storing, sharing and processing such project data. Users will be guided in setting up a data management plan, scaled for each project.
- Sharing good practices on storage bay management and computing means, working closely with NNCR task force.
- Assisting infrastructures in their partnerships on a European scale, especially by guiding them within the ELIXIR infrastructure, for which IFB is the French node.

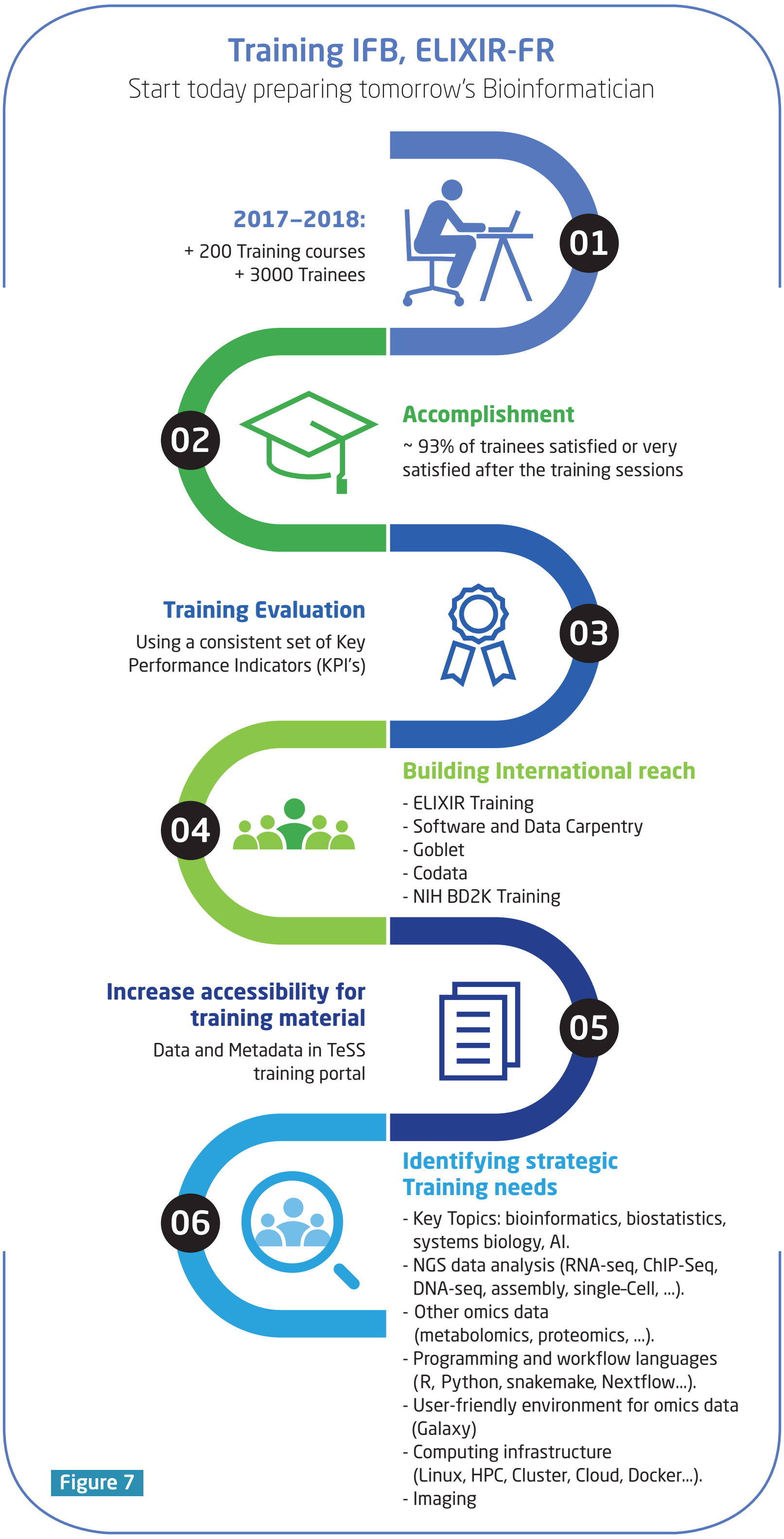
In summary, IFB is a hub of good practices and access to partnerships, as well as a resource available for structures lacking the means or the needs to manage their own hardware infrastructure and software environment.

Training

www.france-bioinformatique.fr/en/training

IFB setup a wide diversity of training events and formats to enhance bioinformatics skills at the national level. The training activities will also be articulated with similar actions developed at the European scale, as in the ELIXIR Training Platform.

Training is an important mission and a substantial activity of the IFB; however, it has challenging aspects.



Training IFB, ELIXIR-FR

Start today preparing tomorrow's Bioinformatician

01 **2017–2018:**
+ 200 Training courses
+ 3000 Trainees

02 **Accomplishment**
~ 93% of trainees satisfied or very satisfied after the training sessions

03 **Training Evaluation**
Using a consistent set of Key Performance Indicators (KPI's)

04 **Building International reach**
- ELIXIR Training
- Software and Data Carpentry
- Goblet
- Codata
- NIH BD2K Training

05 **Increase accessibility for training material**
Data and Metadata in TeSS training portal

06 **Identifying strategic Training needs**
- Key Topics: bioinformatics, biostatistics, systems biology, AI.
- NGS data analysis (RNA-seq, ChIP-Seq, DNA-seq, assembly, single-Cell, ...).
- Other omics data (metabolomics, proteomics, ...).
- Programming and workflow languages (R, Python, snakemake, Nextflow...).
- User-friendly environment for omics data (Galaxy)
- Computing infrastructure (Linux, HPC, Cluster, Cloud, Docker...).
- Imaging

Figure 7

IFB-core and platforms are working together to provide IFB end users (life scientists, biomedical scientists, bioinformaticians etc.) with training formats, taking into account the wide diversity of needs and expectations:

- diversity of the skills for computing (no experience of command line, occasional use and familiarity with a bit of programming);
- specific courses on new technologies or biological applications (e.g., NGS, long reads, metabolomics, single-cell analysis and Hi-C);
- generic methodological and technological skills (statistics, R language, Python programming, Linux environment, best practices for programming etc.);
- advanced use of IT technologies (high-performance computing, virtualisation, cloud computing and interoperability);
- FAIR practices, integrative bioinformatics and, more recently, data management plan and data stewardship;
- best practices in training (pedagogical approaches and reusability of the material).

To this purpose, IFB is developing a seamlessly integrated technical infrastructure that relies on the following components to support training:

- a network of trainers identified in the catalogue of resources (coming soon);
- training events of different formats (short and medium terms);
- diverse training media and resources – static training material, videos, MOOCs;
- diverse training environments – web servers, stand-alone applications with user-friendly interfaces, command lines, specific packages (e.g., R for statistics), VMs, Jupyter notebooks, Docker containers with software for training and Training Infrastructure as a Services (TIaaS) with Galaxy;
- a powerful and flexible infrastructure enabling practical training that sometimes relies on important storage and computing facilities (e.g., NGS analysis, structural bioinformatics etc.);
- feedback mechanisms (satisfaction studies and questionnaires at the end of each training and 6–12 months later to evaluate the impact of the training on their professional practice).

SUMMARY OF THE TRAINING ACTIVITIES

Figure 7 provides some indicators of IFB training activities, highlighting the strong involvement of multiple platforms (number of events) and the important impact in terms of audience (>3,000 trainees in total).

The regional platforms and the IFB-core organised 234 training events in 2017-2018. These training courses cover a wide range of topics, and can be broadly classified into seven main categories.

1. Generalist training in bioinformatics/biostatistics/systems biology.
2. Use and/or administration of the computing infrastructure (high-performance computing, Linux, Cluster, Cloud, Galaxy, Docker etc.).
3. Programming and workflow languages (Perl, Python, Nextflow etc.).
4. Trainings on biostatistics and R language.
5. Trainings on NGS data analysis (assembly, RNA-seq, ChIP-Seq, DNA-seq, variants etc.).
6. Analysis of omics data other than NGS (i.e., metabolomics, proteomics, imaging etc.);
7. Specific bioinformatics topics or software environments (genome annotation, phylogeny etc.).

Several training courses are organised through partnerships between teams from several platforms – such as the Aviesan–IFB School of Bioinformatics (with more than 20 platforms and research teams involved each year), the new University Diploma of Integrative Bioinformatics and the Genopole Summer School (involving 15 IFB platforms for teaching and tutorship).

Diplôme Universitaire en Bioinformatique Intégrative (DUBii)[20]

https://www.france-bioinformatique.fr/en/du-bii

In order to face the demand of biologists to confront the need to get autonomy for the bioinformatics analysis of heterogeneous data resulting from multiple omics technologies, we designed the **University Diploma in Integrative Bioinformatics (DU-Bii)**[21]. This training includes four full weeks of theoretical and practical courses (Unix environment, Python programming, statistics with R, production and analysis of omics data, integrative bioinformatics) followed by a four weeks tutored project led in "full-immersion" on one of the IFB platforms, under the supervision of a dedicated tutor, in order to develop custom integrative pipelines to extract relevant information from data produced by the trainees or their team (**Bring Your Own Data** principle). The DUBii project was launched in partnership with Paris Diderot University and its first session took place in February 2019.

The IFB core cluster was used to run the practical courses as well as the analysis of trainee data during the internship. All the teaching material is freely available (du-bii.github.io/accueil/), the software environment relies on conda-packaged tools, and the practicals have been designed to be reproducible on other computing facilities and on trainees computers.

Trainer team of the Aviesan-IFB bioinformatics school on NGS, Roscoff 2018.

Aviesan–IFB school of bioinformatics – NGS analysis

www.france-bioinformatique.fr/EBAI2018

The yearly **Ecole de Bioinformatique Aviesan–IFB (EBAI)** was launched in 2013 to address an urgent need to develop bioinformatics skills inside the French teams whose projects involves NGS data. This school is addressed to biologists with no prior skills in bioinformatics, but who crucially need to acquire autonomy in handling the NGS data produced by their own research. The school lasts for a full week, and is hosted by the IFB platform ABIMS at Roscoff. The geographic isolation contributes to develop an exceptional climate of intense work and human interactions.

The mainstream courses cover the most popular NGS-based approaches (RNA-seq, ChIP-seq, variant analysis), and their integration (integrative bioinformatics) — complemented by "opening courses" on more specific topics (RAD-seq, small RNA) or ongoing developments (long reads, single-cell). A strength of the school is based on the **Bring Your Own Data** principle, where trainees are led to put in practice the knowledge acquired during the course by conceiving, implementing and starting to run a custom workflow for their research data. To this purpose, they benefit from an exceptional level of support, with ~1 tutor for two participants.

---

[20] University Diploma in Integrative Bioinformatics

[21] www.france-bioinformatique.fr/fr/du-bii

A side-benefit of the school is that it actively contributed to **build a community** among bioinformaticians, since it mobilizes every year ~30 bioinformaticians (teachers, tutors, tech support) from IFB platforms, universities and research teams from all part of France. It also contributes to share good practices in software deployment between several IFB infrastructures (Roscoff, Toulouse, Jouy, Strasbourg, Gustave Roussy and IFB-core).

It also fosters a regular adaptation of the software environments to fast-evolving domain of NGS analysis, as well as the packaging of new tools (or new versions of existing tools) in Galaxy (2013-2016) and Conda (2017-2018). All the teaching material and software packages produced during the school are publicly available (https://www.france-bioinformatique.fr/fr/evenements/EBAI2018), they are updated every year by the teaching teams, in order to cope with the rapid evolution of the domain. This teaching material has been recycled for several other training events.

EBAI represents a considerable involvement of the community, but is considered by its funding institutions (Aviesan and IFB) as a pioneer school to develop new training topics and approaches, and has served as seed for other training events on the whole country (lever effect).

Genopole summer school

www.france-bioinformatique.fr/fr/evenements/genopole-summer-school-2018

Since 2016, the Genopole – with the support of the IFB, France Génomique and the University of Evry – set up an annual residential course tailored for PhD students and researchers who need to apply medicine bioinformatics and genomic techniques to their research projects.

The five-day course is covering statistical methodologies and bioinformatics tools used in human genomics and gut microbial metagenomics. The course includes lectures by international expert speakers, tutorials and hands-on computer sessions. Participants also have the opportunity to give a short presentation of their respective research projects. The summer school accepts 20 participants every year.

Four IFB platforms (IFB-core, Microscope, Migale and Institut Curie) deliver specific trainings according to their competencies. The computing needs are covered by the Cloud Biosphere.

ELIXIR TRAINING

www.france-bioinformatique.fr/en/trainingelixir

ELIXIR training modules is classified into three types according to the target audiences.

1. **Life scientist –** which should follow, as much as possible, the principle of *BringYourOwn Data Workshop*. IFB/ELIXIR have to promote good practices for using software and collating feedback with specific needs and improve training.
2. **Communities of developers –** to enrich the tools and compute platforms based on good practices guidance (European Galaxy Developer Workshop in 2017).
3. **Trainers expert –** including the Train-the-Trainers programme with the best strategies in learning principles and didactic strategies for delivering bioinformatics in the most effective way.

ELIXIR-FR participates (and will continue to participate) in the following ELIXIR events:

- **Software Carpentry and Data Carpentry (SWD/DC) Workshops.** Fundamental skills necessary for software development techniques and data manipulation are taught to help the researchers to improve or speed up their research. ELIXIR and ELIXIR-UK node launched a pilot workshop (SWC/DC) for ELIXIR Nodes. Each node will benefit from two workshops per year.
- **Capacity Building Training.** The objectives are to (i) construct and coordinate ELIXIR-wide 'communities of practices' to support and develop the professionals who deliver advanced data and bioinformatics support and services in ELIXIR Nodes and (ii) develop activities/resources that address the training needs from the new nodes.

Some ELIXIR training events are

- **European Galaxy Developer Workshop (TtD)**
  - 16/01/2017 to 19/01/2017 Strasbourg, France
- **ELIXIR-EXCELERATE Train-the-Trainers (TtT)**
  - 26-03-2018 to 27-03-2018, Institut Pasteur, Paris 75015, France
- **ELIXIR Software Carpentry Workshop**
  - 28-03-2018 to 29-03-2018, Institut Pasteur, Paris 75015, France
- **ELIXIR Training: Plant genome assembly and annotation (TtR)**
  - 24-09-2018 to 28-09-2018, CIRAD, Montpellier, France
- **Workshop Bioschemas, a lightweight approach to enable FAIRer data resources**
  - 24-09-2018
  - NETTAB 2018: Building a FAIR Bioinformatics environment
  - October 22–24, 2018, Genoa, Italy

ELIXIR Genome Assembly and Annotation (G2A) Workshop

ELIXIR-FR codeveloped – with three other ELIXIR Nodes (ELIXIR Sweden, ELIXIR Norway and ELIXIR Belgium) – a training in Genome Assembly and Annotation. This course was developed as a part of the ELIXIR-EXCELERATE[22] efforts in capacity building[23].

The content of the course has been adapted to take into account the needs from the host node regarding organisms or sequencing technology specifications. The training takes place for five days, and students without previous knowledge in bioinformatics or statistics can participate. There have been four sessions in different European countries (Czech Republic, Slovenia, Portugal and France), and a session will be set up in Italy this summer. Four IFB platforms are currently participating in these Workshops: IFB-core, Genotoul, URGI and Southgreen.

More than 80 students around Europe have been trained in this course, and the satisfaction rate exceeds 95%. Moreover, the countries hosting this training workshop have developed IT-level capabilities to run it; currently, four different IT clouds are able to set up the required software environment. The Slovenian e-Learning platform (EeLP) was used to host the presentation and survey. The high impact of these workshops allows them to be included as a pilot project in the ELIXIR implementation study (IS): using clouds and VMs for bioinformatics training (WaaS).

Furthermore, during the preparation phases of these workshops, the teachers set up a wiki-handbook (https://biosphere.france-bioinformatique.fr/wikia2) with the software, pipelines and material used during the training and wrote a best-practice article in F1000 journal[24]. The G2A workshop experience is found in the ELIXIR 2019–2023 programme[25] and will be implemented soon in new areas as single-cell analysis and data management plan.

---

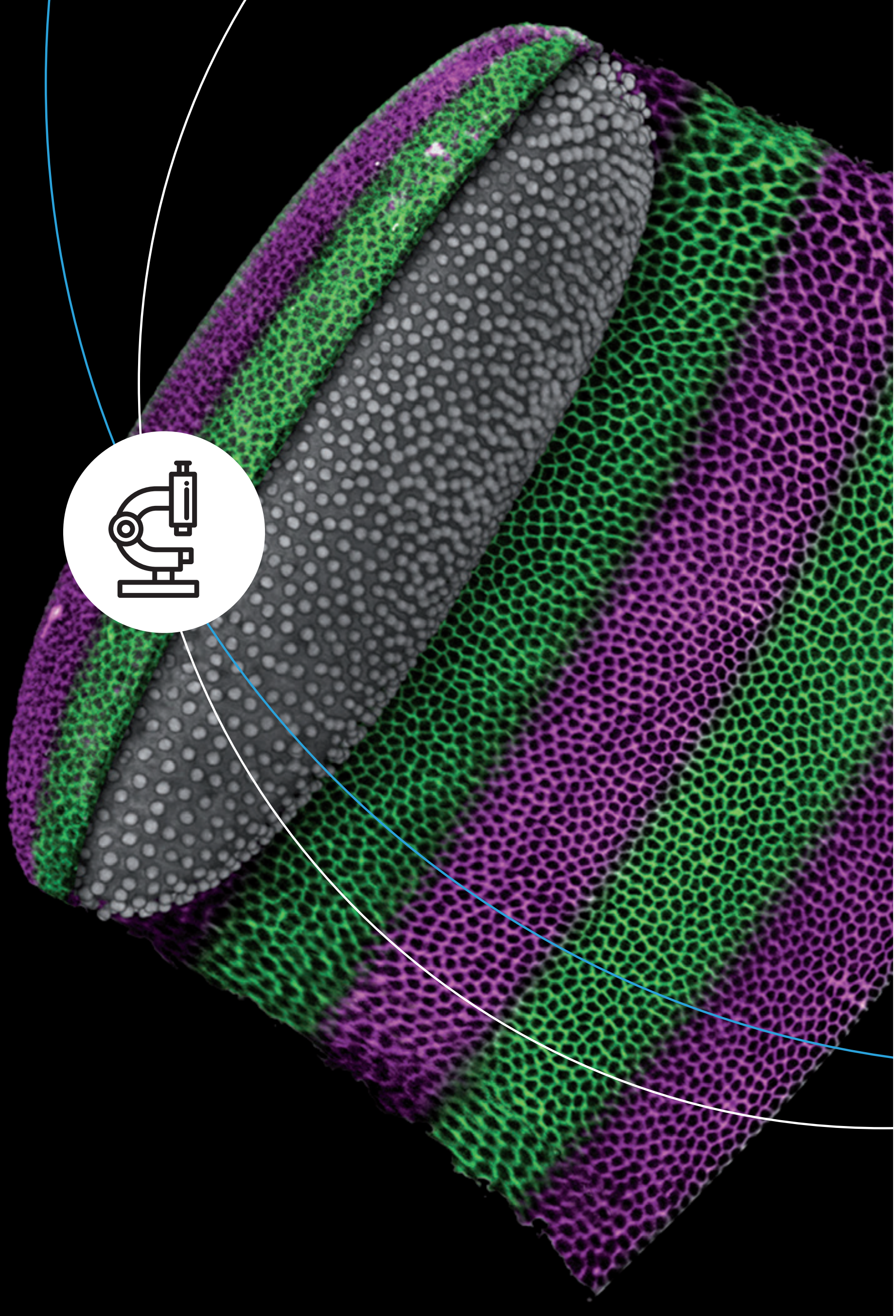
[22] www.elixir-europe.org/about-us/how-funded/eu-projects/excelerate

[23] www.elixir-europe.org/about-us/how-funded/eu-projects/excelerate/wp10

[24] Dominguez Del Angel, V., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C., Vinnere Pettersson, O., Amselem, J., Bouri, L., Bocs, S., Klopp, C., et al. (2018). Ten steps to get started in Genome Assembly and Annotation. F1000Research 7, 148. https://f1000research.com/articles/7-148/v1

[25] www.elixir-europe.org/about-us/what-we-do/elixir-programme

INNOVATION: INTEGRATIVE BIOINFORMATICS



www.france-bioinformatique.fr/en/innovation

Since the turn of the century, every decade has witnessed the development of new high-throughput technologies to monitor diverse layers of biological processes (e.g., genome, transcriptome, proteome, interactome, metabolome). Each of these technologies motivated the development of new bioinformatics approaches and tools to extract relevant information from the raw data. However, an integrated approach is still lagging behind to translate this diversity of complex and heterogeneous data into useful knowledge.

A number of networks and national projects dedicated to the production, management and interpretation of specific data types (genomics, proteomics, metabolomics, imaging, animal models, cohorts, personalised medicine etc.) have been set up in France. In order to assume a global view of the systems which are studied, this heterogeneous information need to be integrated.

We, therefore, proposed to tackle the challenge of integrative bioinformatics through three actions of the new action plan:

- Organise pilot projects
- Create an interoperability workgroup
- Organise a workshop on integrative bioinformatics

These actions are intimately linked to the IFB NNCR (computing power and storage for developed tools, workflows, databases etc.), ELIXIR and training activities (i.e., DUBii, see the ‘Training’ section).

PILOT PROJECTS

www.france-bioinformatique.fr/en/pilot-projects

We launched a **call for pilot projects in integrative bioinformatics** to foster the development of innovative bioinformatics approaches integrating heterogeneous biological data types. The relevance of the approaches should be demonstrated by applying them to specific use cases relying on data produced by at least two national research infrastructures, animal model facilities, human cohorts, or national plans. The projects should also fit with national priorities – including health, food security and safety, environment, biotechnology and fundamental research –. The pilot projects were conceived to produce results in 18 to 24 months. Successful applications were supported by a fixed-term contract of a maximum of two years.

Five projects were selected:

www.france-bioinformatique.fr/en/selected-projects

1. **INEX-MED** (Integration and Exploration of heterogeneous Biomedical Data)
2. **ProMetIS** (Proteomics and Metabolomics Integrated Software)
3. **IntegrParkinson** (development of an interactive software for the integrative analysis of multimodal datasets in Parkinson's disease and two use cases in radiomics)
4. **MS2MODELS** (Modèles en masse)
5. **PhenoMeta** (plant phenomics and metabolomics integration for studying genetic and physiological controls of drought tolerance in maize)

Three additional projects were reoriented to be supported by other IFB actions 'Support to Databases', 'Catalyzing Interoperability' and 'Sharing of services with national research infrastructures' :

www.france-bioinformatique.fr/en/supported-projects

1. **My EMBRC Image** (the EMBRC centralised imaging database with distributed storage)
2. **B2SH** (Biostatistics–Bioinformatics for Human Health)
3. **BANCCO** (National Bank of Constitutional CNV)

The pilot projects involve ten national research infrastructures (IFB, France Génomique, MetaboHub, ProFI, FRISBI, FBI, FLI, Phenome, EMBRC-Fr, F-CRIN), two animal model facilities and human cohorts (PHENOMIN, BIOBANQUES), and the national plan 'France Médecine Génomique 2025'.

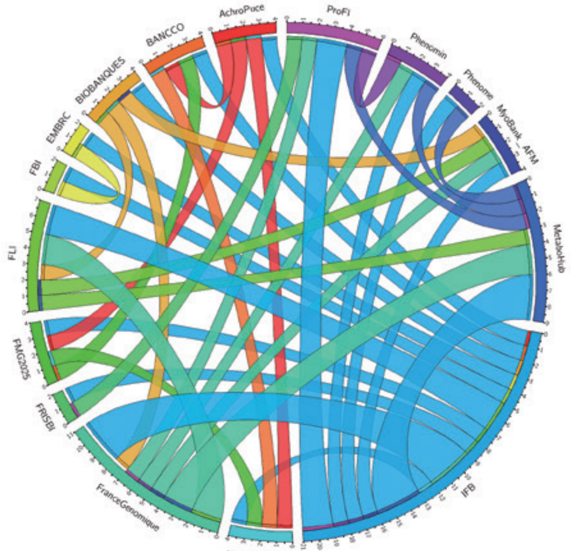


Figure 9

Infrastructures, cohorts, national plans and other structures interacting through IFB pilot projects. Each structure has a specific colour. The width of the ribbons is proportional to the number of projects where structures interact.

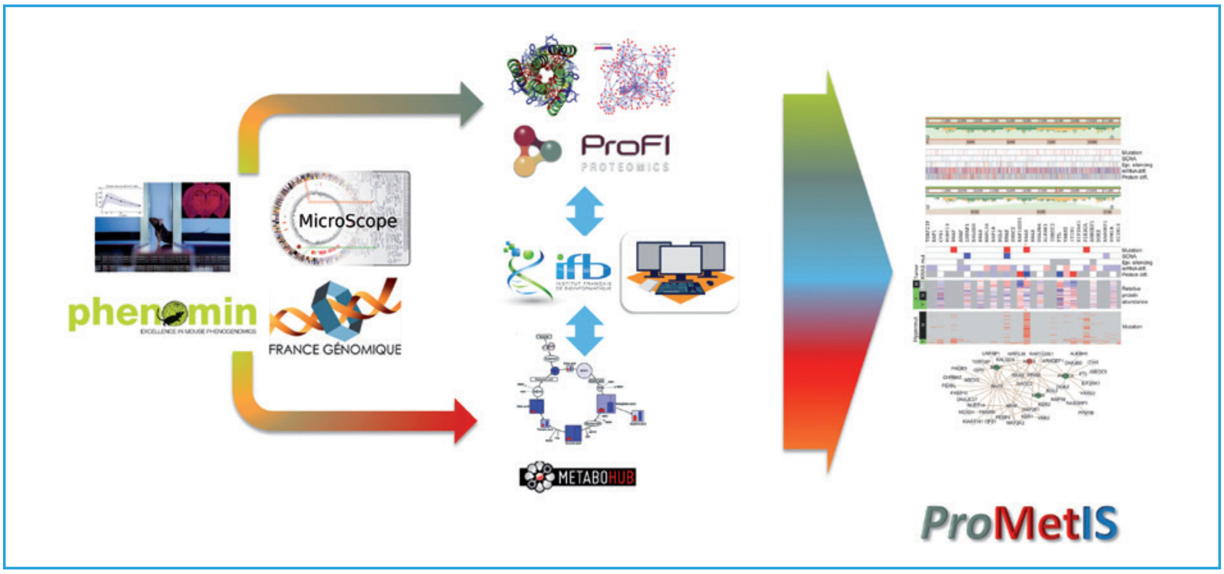


Figure 8 Schematic representation of the ProMetIS project components and their interactions

INTEROPERABILITY

www.france-bioinformatique.fr/en/interoperability

Interoperability solutions aim at providing uniform access to diverse and distributed resources with the purpose of their integrated exploitation. Making biological resources more interoperable is an essential requirement for taking full advantage of their obvious complementarity and gaining new insights in integrative biology.

Interoperability is closely linked to the FAIR principle (Wilkinson et al, 2016). However, several other levels of interoperability and integration exist (e.g., tools, software environments, etc.), each associated to families of solutions and good practices.

In the last year, we attempted to clarify the needs of the communities involved in the pilot projects selected and identify existing standards and interoperability solutions within IFB and at the European level.

A shared concern by three pilot projects deals with semantic interoperability of metadata describing datasets or data resources. Metadata should be expressed with standard terms derived from available ontologies and according to a standard format. Teams involved in the **INEX-MED** and **IntegrParkinson** projects are working on the design of knowledge graphs adapted to facilitate integration between omics data and medical images for machine learning applications. In the **PhenoMeta** project, bioinformaticians already active at the European level on plant standards construction such as MIAPPE and BrAPI, want to develop a plant phenotyping ontology for purpose of plant datasets interoperability and treatments reproducibility.

Another need detected in three other pilot projects is related to tool interoperability. The **ProMetIS** and **MS2MODELS** projects wish to develop software based on flexible combination of interoperable tools in the field of multi-omics analyses (proteomics and metabolomics for the former, proteomics and 3D-interactomics for the latter). Similarly but in the context of sensitive health data, the **B2SH** project aims to combine biostatistics and bioinformatics (related to genome sequence analysis) tools. Interoperability challenges will concern workflow description, provenance and reproducibility, as well as REST APIs and containerisation.

AVIESAN GGB–IFB WORKSHOP ON INTEGRATIVE BIOINFORMATICS

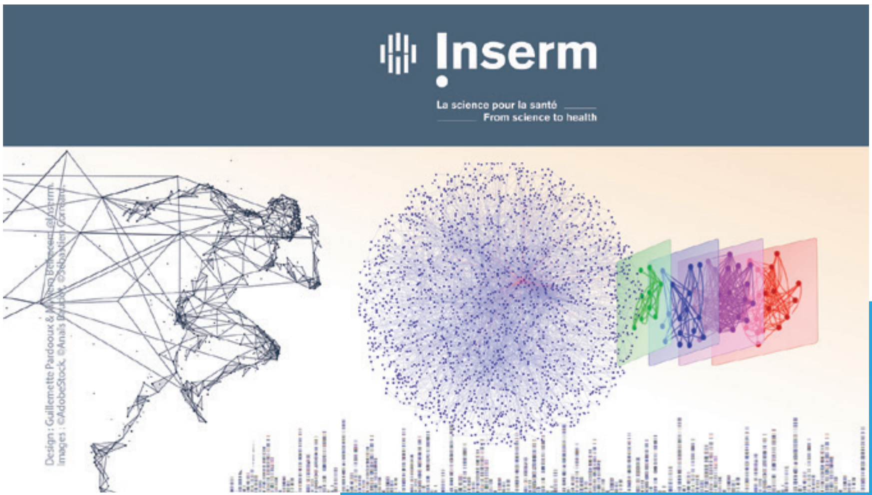
In September 2018 Aviesan ITMO Genetics, Genomics and Bioinformatics (GGB) and IFB coorganised a one-day workshop titled ***Challenges and Perspectives in Integrative Bioinformatics***[26], which gathered 80 attendees.

A series of talks presented the state-of-the-art approaches in integrative bioinformatics (multilevel matrix factorisation methods and multilayer network analysis). Overall, the meeting covered all the steps between data production, computational treatment, statistical analysis, visualisation and interpretation, illustrated by relevant examples from various domains of life sciences and their applications.

The talk sessions were followed by a round-table session where all participants discussed the personal experience, current needs and problems encountered, which led to an animated discussion about current challenges, perspectives and strategies to further develop integrative bioinformatics.

Beyond the scientific interest of the talks and debate, this one-day workshop also enabled us to identify experts in the field, who were then invited, and accepted, to contribute to the Diplôme Universitaire en Bioinformatique Intégrative.

Figure 10



Advertising poster about Aviesan GGB–IFB workshop on integrative bioinformatics

---

[26] https://integrabioinfo.sciencesconf.org/

ELIXIR - EU INTEGRATION

IFB as the French node of ELIXIR

www.france-bioinformatique.fr/en/elixir-fr

ELIXIR aims to empower ELIXIR member states to integrate life science data and services and maximise the member states bioinformatics expertise. For that purpose, ELIXIR consolidates Europe's national centres, services and core bioinformatics resources into a single, coordinated infrastructure and builds a community of computer scientists across Europe. ELIXIR's implementation is supported for the 2015-2019 period by the H2020 ELIXIR-EXCELERATE project coordinated by ELIXIR's hub.

ELIXIR-FR corresponds to IFB and is among the largest nodes, with a large panel of activities and services (France addresses human health, food and agriculture, marine diversity and so forth). An assessment of the services provided by the French community eligible for their inclusion in the ELIXIR catalogue was launched at the end of 2018, and 82 propositions are currently under review.

ELIXIR has been, so far, a powerful environment supporting the European contribution to the development of standards and generic tools for the community in life sciences: Bioschema, EDAM, FAIR suite of standards and tools for plant phenotyping data standardisation, workflows and containers and so forth. It brings an organised European voice in international consortia (e.g., Breeding API, MIAPPE and GA4GH, in which ELIXIR-FR has been very active) and a strong contribution in the development of sustainable environments of data analysis (BioTools, Galaxy.eu etc.) and provides capacity building and training across Europe (e.g., training on Genome Assembly and Annotation in Montpellier [September 2018] and ELIXIR Software Carpentry Workshop, Paris, March 2018).

ELIXIR supports codevelopment activities through internal projects between its nodes. In the 2015–2018 period, ELIXIR-FR was involved in 19 Implementation Studies (IS), staff exchanges, but also, more uncommon, through hackathons. The last one was organised in France (Paris, November 2018) and 150 experts and computer scientists/engineers from ELIXIR's nodes but also, and this was new, from the private sector, were hosted for one week to develop codes together or 'FAIRify' their resources.

ELIXIR enables participation in structuring European projects, such as the H2020 (2019–2023) EUCANCan which aims to align and interconnect existing European and Canadian infrastructures for the analysis and management of genomic oncology data. The H2020 GenRes Bridge coordination and support action will be an opportunity to position the possible role of ELIXIR in the biodiversity domain. ELIXIR is a good opportunity for IFB to align with European Open Science Cloud (EOSC) objectives (H2020 EOSC-Pillar; see also interactions with other ESFRIs).

On the whole, ELIXIR-FR is visible in ELIXIR for

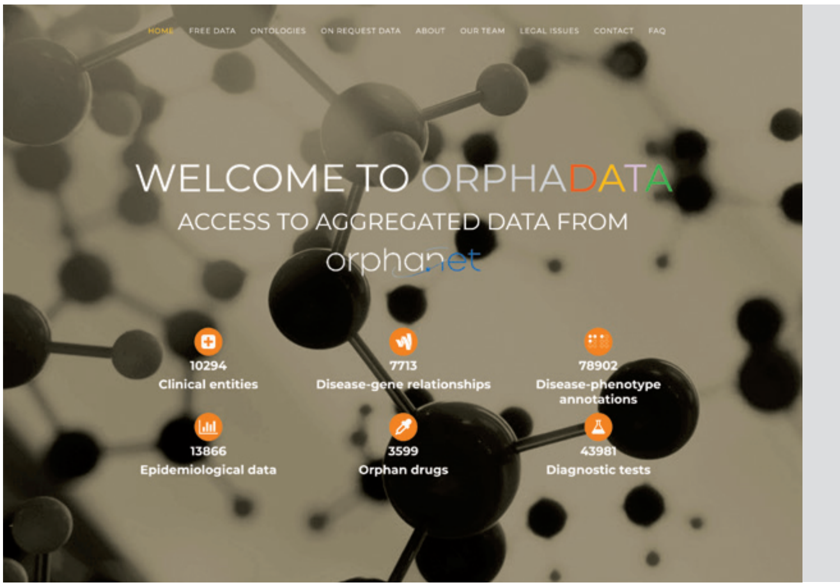
- the nomination of Orphadata (www.orphadata.org) as an ELIXIR Core Data Resource, which recognises its fundamental importance to the wider life-science community
- its activities on workflow development and containerisation and its leading role in the ELIXIR Galaxy community
- its very active or even leading role in several communities like the plant sciences, human CNV, marine metagenomics, proteomics and microbial biotechnology communities that are instrumental to drive priorities for development and capacity building with researchers. ELIXIR-FR brought important contributions to several community-driven projects (e.g., the Beacon project in the human data community; marine microeukaryotic transcriptomic databases and associated assembly and annotation pipelines for marine metagenomics community in the ELIXIR-EXCELERATE project)
- its contribution to training and capacity building activities (e.g., Dominguez Del Angel et al. Ten steps to get started in Genome Assembly and Annotation. F1000Research 2018, 7[ELIXIR]: 148).

ELIXIR is also the largest European research infrastructure in terms of the 23 participating countries. It has built, during the last 5 years, the necessary processes to develop collective roadmaps and make decisions. ELIXIR-FR has been involved in these discussions and is buying some of the processes for its own internal organisation, which also correspond to a federation of platforms. The current work in progress in the domain is about the business model and how to assess impact.

ELIXIR-FR SUMMARIES & HIGHLIGHTS

Orphanet (Orphadata) becomes an ELIXIR Data Core Resource

(decision made on December 12, 2018)



WELCOME TO ORPHADATA

ACCESS TO AGGREGATED DATA FROM

orphanet

> Keeping a strong involvement in the ELIXIR communities

- Launch and co-coordination of new communities
  - Galaxy co-coordinated by ELIXIR-FR (Gildas Le Corguillé)
  - Microbial biotechnology in process for definitive acceptance and co-coordinated by ELIXIR-FR (Vincent Fromion)
  - h-CNV (approved in December 2018) – coordinated by ELIXIR-FR (Christophe Béroud and David Salgado)
- Co-coordination of the plant sciences community (replacement of P. Kersey, EBI, by Cyril Pommier) since March 2018

> Contribution to the development of EU projects

Contribution to the development of the project INFRAEOSC-4 Life (coord. ELIXIR Hub; accepted, see Figure 11) and currently following the development of the project INFRADEV3 (coord ELIXIR Hub).

SME ELIXIR activities

ELIXIR Innovation and SME Forum: Data-Driven Innovation in Rare Diseases and Personalised Medicine, November 2017, Paris (see Industry activities).

Contributions to Implementation Studies (IS)

2017

① A microbial metabolism resource for systems biology
② Architecture for software containers at ELIXIR and its use by EXCELERATE use-case communities
③ Data validation
④ Development of learning paths for users of ELIXIR services
⑤ ELIXIR Beacon
⑥ The reuse, extension, scaling and reproducibility of scientific workflows
⑦ Bioinformatics training (workshops as a service) using clouds and VMs

2018

Contribution to eight IS accepted and to submission of five others on October 30 (to be reviewed)

① Development of architecture for software containers at ELIXIR and its use by EXCELERATE use-case communities
② Bioschemas: Community Adoption and Training
③ Metabolite identification
④ Towards Data Stewardship in ELIXIR: Training & Portal
⑤ Crowdsourcing the annotation of public proteomics datasets to improve data reusability
⑥ Extending open proteomics data analysis pipelines in the cloud
⑦ Galaxy Community Kick-off
⑧ Comparison, Benchmarking and Dissemination of Proteomics Data Analysis Pipelines project, in which Elixir-FR is a partner, submitted on the 30th of October 2018

Lead ELIXIR staff exchanges with EMBL-EBI on Bioschemas

- Creating How-to Guides for Bioschemas at Institut Henri Poincaré, Paris, 4th to the 8th of October 2018.

EU BIOHACKATHON 2018

https://2018.biohackathon-europe.org/

One of the most successful events organised in 2018 by IFB and ELIXIR Hub was the first EU BioHackathon meeting held in Paris on 12 to 16 November 2018. This initiative has been a great opportunity to engage the bioinformatics community in existing ELIXIR activities and also engage ELIXIR with other activities driven by the broader bioinformatics community.

The community and scientific benefits that resulted from the EU BioHackathon are as follows:

- opportunities for technical developers to meet and work together on topics of common interest;
- acceleration and progression of existing projects and upcoming deliverables;
- improvement of community knowledge and interactions of services and technical activities;
- ELIXIR would also benefit from engaging external communities with ELIXIR activities, as well as receiving feedback from external communities on upcoming trends and user requirements for new ideas and relevant work.

The themes proposed at this BioHackathon were aligned to challenges proposed by ELIXIR platforms (computing, data, interoperability, tools and training) – including topics proposed by ELIXIR communities, suggestions from the local organisers (IFB) and the representatives from the Japanese BioHackathon[27]. The BioHackathon activities were driven by practical sessions where people gathered, discussed and implemented ideas/projects during intensive and productive hands-on coding sessions.

IFB, the Japanese BioHackathon and ELIXIR drove the first EU BioHackathon with the following goals:

- Advance the development of an open-source infrastructure for data integration to accelerate scientific innovation.
- Engage with the technical developers in the bioinformatics community both inside and outside of ELIXIR institutes.
- Strengthen the interactions with ELIXIR platforms and the ELIXIR use-case Communities.
- Promote interoperability through the adoption of data standards, API and tool implementation and development standards, and community-accepted libraries. The EU BioHackathon highlighted Bioschemas (https://www.bioschemas.org/) to drive this goal from the ELIXIR strategic IS approach.

The format of this BioHackathon followed those that were organised and hosted previously in Japan. The EU BioHackathon committee had engaged and communicated with their Japanese BioHackathon's counterpart from the beginning in all activities – for example, planning and reviewing of the topic proposals. For this first EU BioHackathon, the duration of the event was a five-day period. The EU BioHackathon programme committee funded 28 topics[28] to invite key participants to the event. It hosted 150 participants from 14 different countries and the return on investment of this event was of €6.5 for each €1 spent, based on the extra time needed to develop the same activities alone.

IFB platforms (Orphanet, MMG-GBIT, ABIMS, URGI, C3BI and IFB-core) proposed eight hacking projects and were actively involved in the organisation and realisation of this event.

Andrew Smith (ELIXIR hub external relations) and Niklas Blomberg (director of ELIXIR).

INTERACTIONS WITH OTHER ESFRIS

In the last five years, ELIXIR has supported coordinations with other ESFRIs in particular through the ELIXIR communities (e.g., EMPHASIS and ELIXIR, where the French node has been at the initiative), which was also internally helpful to better define the e-infrastructure aspects to be mutualised and the specific ones. It was also useful to ensure that the key objects allow interoperability between phenotyping data produced by EMPHASIS and polymorphism data (stored in ELIXIR's Archives at EMBL-EBI). Another example is the ELIXIR marine metagenomic community, which works closely with the EMBRC Eric in several domains: promotion of the best strategy in terms of e-infrastructure, data management plans and NATO and analysis workflows in relation to genomic observatories projects.

In the coming period, the collaborations with the other ESFRIs in the domain of the life sciences will be supported by the H2020 EOSC-Life project, coordinated by ELIXIR Hub. The French node will be involved in activities with ECRAN, MIRRI, EMBRC and EMPHASIS.

Eric Guittet, French delegate at ELIXIR board.

Anne-Françoise Adam-Blondon (Deputy Head of the ELIXIR French node) and Victoria Dominguez Del Angel (Training coordinator Elixir-FR).

Frédéric Boccard and Eric Guittet, French delegates at ELIXIR board.

ELIXIR consolidates Europe's national centres, services, and core bioinformatics resources into a single, coordinated infrastructure.

---

[27] http://www.biohackathon.org/

[28] http://bh2018paris.info/projects

IFB participation to EU-funded projects

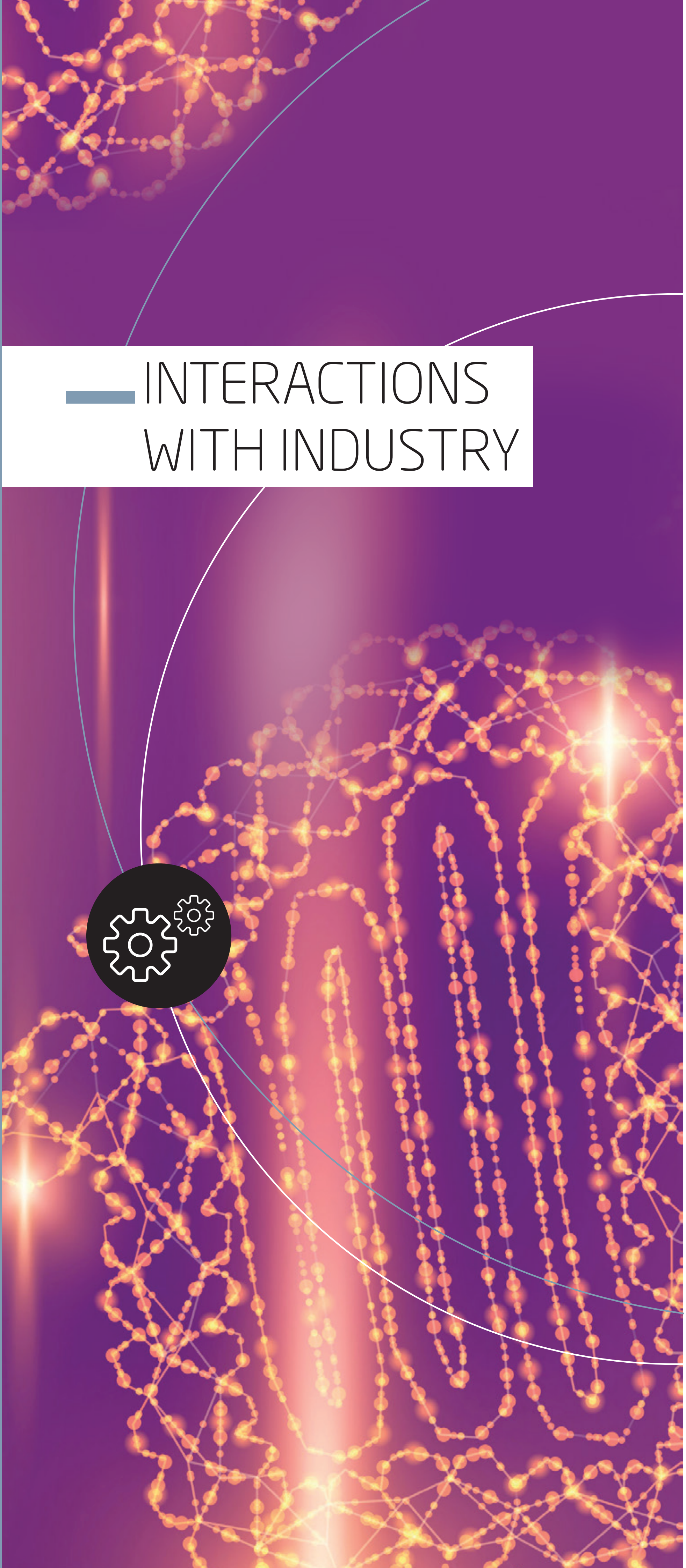
Acronyme	Topics	Start	End	Total budget	FR budget
EXCELERATE (coord ELIXIR Hub)	Bioinformatics infrastructure	2015	2019	22 M€	1 M€
CORBEL					
INFRA-EOSC4 Life (coord ELIXIR Hub)	Providing an open collaborative space for digital biology in Europe. France is mainly involved through its coordination of 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-RARe (https://www6.inra.fr/agrobrc-rare/)	2019	2021	3 M€	80,275 € (out of 388 k€ for FR in total)
COST action Integrape	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

Projects under development where the ELIXIR-FR/IFB 'hat' is an important part of the consortium:

- Two RIA projects in the frame of the SFS28 call with the ELIXIR plant community and EMPHASIS 'hats' (DE/IPK and INRA coord). Developments that will build on/complement the activities of a community IS, submitted in October 2018, as well as mobilisation of the output of the WP7 (plant use case) of the H2020 EXCELERATE project.
- One project in the INFRAEOSC call food cloud (INRA coord).
- One project in the INFRADEV3 call (ELIXIR Hub coord).

INTERACTIONS WITH INDUSTRY



www.france-bioinformatique.fr/en/industry

Europe is a rich breeding ground for biotech start-ups, with 1,799 healthcare biotech companies operating in Europe in 2012. These start-ups and industries are major employers globally, and most of them need to create value and innovation services to pursue their activities, with a strong reliance on bioinformatics resources. Other industrial sectors such as food or agriculture are in the same situation.

One of IFB's missions is to become an effective actor to simplify access to bioinformatic tools, computing, quality control omics data for researchers working in life science industries. In order to fulfil this mission and reinforce its links with industries, IFB set-up an industry working group whose major task consists of:

- Providing support to industry for their needs in bioinformatics
- Being the entry point in bioinformatic innovation services (tools, IT resources and training)
- Increasing industry usage of services deployed by IFB
- Promoting consulting services
- Ensuring effective communication between Industry and IFB (communication channels)
- Helping French SMEs, via ELIXIR Industrial channel, to promote their services across European countries (SME events)
- Organising an IFB-Industry forum
- Ensuring a clear billing policy, built on existing policies in partner institutions or on cloud-based service providers

Patrick Durand, co-head of the IFB-Industry work group

In 2017 IFB hosted and coorganised an ELIXIR Innovation and SME Forum titled *Data-Driven Innovation in Rare Diseases and Personalised Medicine in Paris*[29].

During these two days, attendees were immersed in a world of data-driven innovation illustrated through talks given by innovative special keynote speakers, companies; and presentations of IFB and ELIXIR's open data resources and services. The event allowed attendees to discuss and interact with other companies and IFB partners. Four IFB platforms were actively involved during this event: IFB-core, GenOuest, MMG-GBIT and, more specifically, Orphanet. 110 participants registered for this event, with participants coming from seven European countries, and there was a total of 30 F2F meetings between academic and industrial participants.

Victoria Dominguez Del Angel, co-head of the IFB-Industry work group

---

[29] https://www.elixir-europe.org/events/elixir-innovation-and-sme-forum-data-driven-innovation-rare-diseases-and-personalised

Charles Loomis, SixSq

In 2018 IFB set up the Industrial Advisory Committee (IAC) with the following members:

- Anne-Sophie Coquel, *Head of Genomics, Sanofi, FR*
- Jérôme Wojcik, *Senior Vice President Europe Translational Informatics and Biometrics Europe at Precision for Medicine, CH*
- Charles Loomis, *Chief Technical Officer and co-founder of SixSq, CH*
- Bruno Lacroix, *Senior Director, Technology Research Department at bioMerieux, FR*
- Frédéric Sapet, *Chief of Project at Biogemma, FR*
- Guillaume Boissy, *Head of Data Management and Analysis, Translational Bioinformatics IRT Bioaster, FR*
- Stéphane Bortzmeyer, *Network and Systems architect at AFNIC, FR*

The first IAC meeting was held on March 25, 2019, at Pasteur Institute, Paris.

Ana Rath, Orphanet

On the second semester of 2018, a scientific cooperation between IFB and a French multinational company (Total) aimed at designing an innovative software suite to analyse metagenomics data, was initiated and the collaboration agreement is near to be signed by all parties involved (i.e., Total, CNRS, INRA and CEA). The execution of this project is being conducted as a pilot to specify and assess the process of running an IFB-Industry partnership:

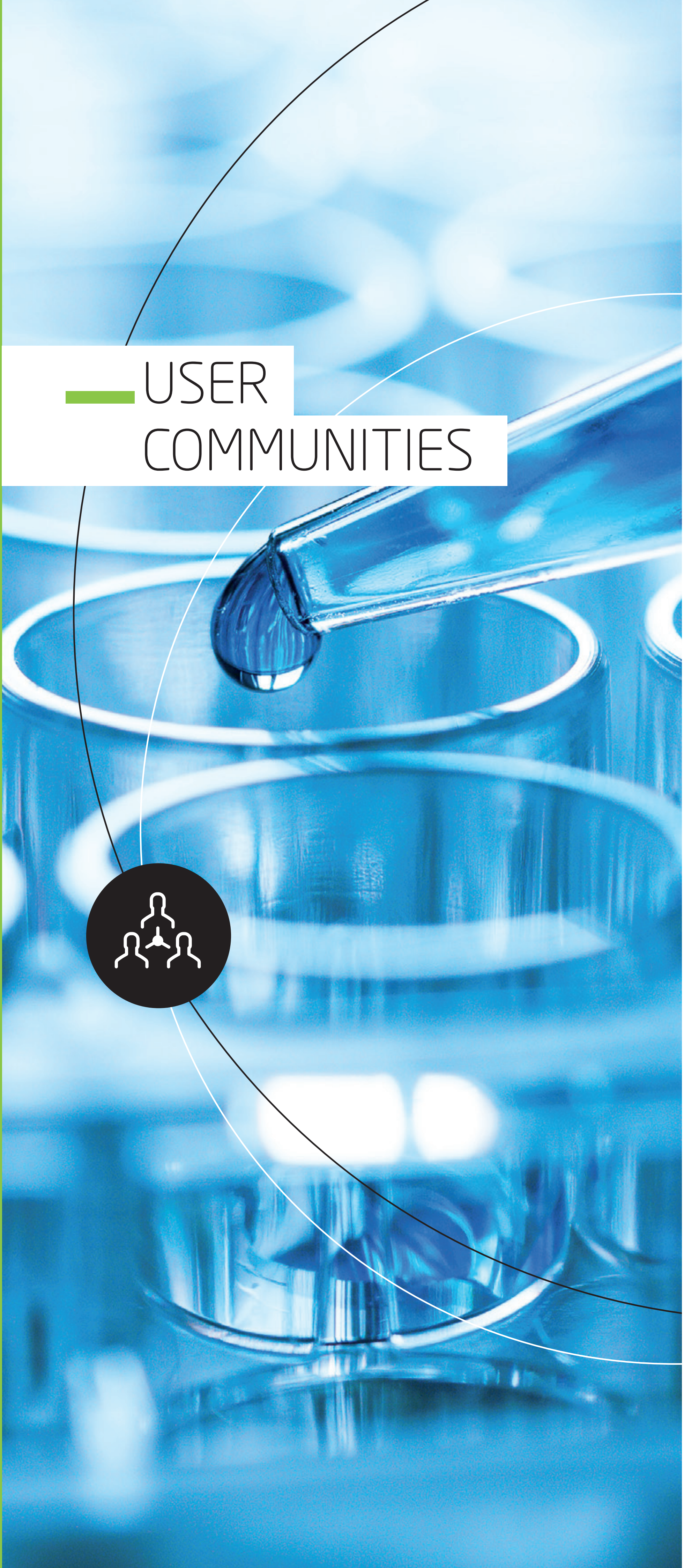
- Defining a working group: IFB members + industry partner
- Defining a document with functional features, technical solutions, tasks & deliverables and a budget proposal
- Writing an agreement between IFB and the company
- Defining the operational management of the project: which IFB platforms are involved and for which competences? what is the role of the IFB-core in the project?
- Defining the minimum duration of the project

Marc Hanauer, Orphanet

Christophe Béroud and David Salgado

ELIXIR Innovation and SME Forum: Data Driven Innovation in Rare Diseases and Personalised Medicine,

Paris, Nov 2017



USER COMMUNITIES

www.france-bioinformatique.fr/en/innovation

IFB user communities are defined by the 'customers' from the academic as well as industrial research world who make use of the IFB resources (services, training, achievements of the innovation axes etc.).

Our user communities cover different disciplines (biologists, physicians, bioinformaticians, computer scientists, mathematicians etc.), fields of application (fundamental research, health, agriculture, environment, biotechnologies etc.) and institutional levels. They also include

- **Research teams of the IFB supporting institutions** (CNRS, INRA, INSERM, CEA, Institut Pasteur, Institut Curie, CIRAD, IRD and universities) working in life sciences and their applications to different domains: health, agriculture, environment, biotechnologies and so forth.
- Other **national infrastructures** specialised in various data-production technologies for life sciences: genomics, proteomics, metabolomics, phenotypic data and so forth. given that IFB should be seen as a provider of bioinformatics services allowing to address biological questions requiring the integration of different biological data types.
- **Industrial partners**, the application fields of which cover both agronomy and health. Their needs are reported by the IAC, see IFB Governance, page 53).

Bioinformatics for Health

In September 2018, a working group dedicated to bioinformatics for health data has been created. This working group is led by Ivan Moszer and David Salgado and includes 23 members from various IFB platforms.

This working group has several objectives:

- Report to IFB and to our institutions about activities done in this perimeter.
- Respond to the various solicitations received by IFB with a specific focus on health data, such as participation in meetings, workshops or projects.
- Interact with the national sequencing programme PFMG2025 (Plan France Médecine Génomique 2025).
- Follow up on the various activities carried out in ELIXIR, particularly in the Human Data, Rare Diseases and Human Copy Number Variation communities.
- Work on a white paper to get an overview of the skills and expertise available at the national level.

MAIN ACHIEVEMENTS SO FAR

Since its creation, the group has been required to contribute to various activities such as

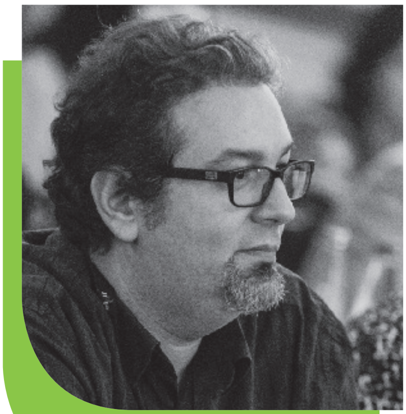
- ELIXIR activities, especially a grant writing for an INFRADEV 3 proposal, and the identification of possible partners for participating in a human federated data IS.
- Representing IFB in a meeting called *Journée recherche et santé* on the topic *Clinical phenotyping and Systems Biology* organised by INSERM and Aviesan ITMO Physiopathology, Metabolism and Nutrition (Christophe Béroud talk on *Data Science for Health and Medicine: The French Landscape* on November 22, 2018).
- Participation in the INSERM transversal project called *Human Development Cell Atlas (HuDeCa).*
- Participation in the organisation of an INSERM workshop dedicated to health infrastructures. The goal of this workshop was to make a presentation on the **French ecosystem of bioinformatics platforms for human data and health** on January 25, 2019 (Claudine Médigue and Jacques van Helden).
- Coorganisation of a wide **user survey** of INSERM laboratories about their needs in bioinformatics (software tools, external bioinformatics services, internal capacity building, training programmes, recruitment policies), as detailed in the next section.

User surveys about needs in bioinformatics

SURVEY OF THE NEEDS OF INSERM UNITS

In December 2018, the Pôle Infrastructure INSERM invited IFB to co-organise a one-day national workshop on **the French ecosystem of bioinformatics platforms for human data and health.** The workshop was organised on January 25, 2019. It took place simultaneously in Paris (~80 persons) and 12 other INSERM regional delegations, with a videoconference system enabling all participants to intervene at distance. In total, the workshop gathered ~150 participants.

The initial goal of the workshop was to evaluate the bioinformatics resources currently available to INSERM research teams. We took benefit of this event to organise a survey of the needs expressed by INSERM research units on the different types of services: hardware (survey previously led by INSERM DSI), software tools and environment, training, conception and follow up of bioinformatics analysis projects, recruitment policy and so forth.


Marc Hanauer, Orphanet


David Salgado, co-head of the IFB work group «Bioinformatics for health»


Ivan Moszer, co-head of the IFB work group «Bioinformatics for health»

This survey (Figure 12) confirmed the strategic choices for the IFB action plan 2018–2019, in terms of services (bottom panel) and training (top and middle panel).

For services, the first priority expressed by researchers is a support for the treatment and analysis of their data, followed by the development of bioinformatics resources (databases, tools and workflows) and, to a lesser extent, to provide upstream advice at the conception of research projects.

The survey of requirements for training confirms the trends described in the Training section of this document, and addressed by our training programmes.

- The data-production technologies are strongly oriented towards multiomics (NGS, metabolome, proteome) and imaging.
- Life scientists express a strong demand to acquire skills for advanced statistics and to master the R statistical package. User-friendly interfaces (Galaxy, Cytoscape) are still popular, but we also observe a clear demand of skills in Unix.

A NATIONAL SURVEY OF USER NEEDS IN BIOINFORMATICS

Based on the results of the survey of INSERM user needs, we will generalise the approach and lead a general survey of needs expressed by academic users at the national level, in collaboration with our supporting and partner institutions.

The IAC will also start a similar survey for the needs of companies in terms of bioinformatics.

OTHER THEMATIC COMMUNITIES

The community 'Bioinformatics for Health' was initially created to answer demands from one of our supporting institutions (INSERM), and it proves particularly active, thanks to repeated interactions on different activities.

In 2019 several new thematic communities will be launched in collaboration with our supported research institutions: bioinformatics for agriculture, biotechnologies and so forth.

IFB will offer support to animate scientific events between these user communities in order to define their needs in bioinformatics and evaluate the ways to address these needs in the context of future actions (e.g., call to challenges for integrative bioinformatics, scheduled to start in the second semester of 2019). Other communities may emerge on a user-driven mode.

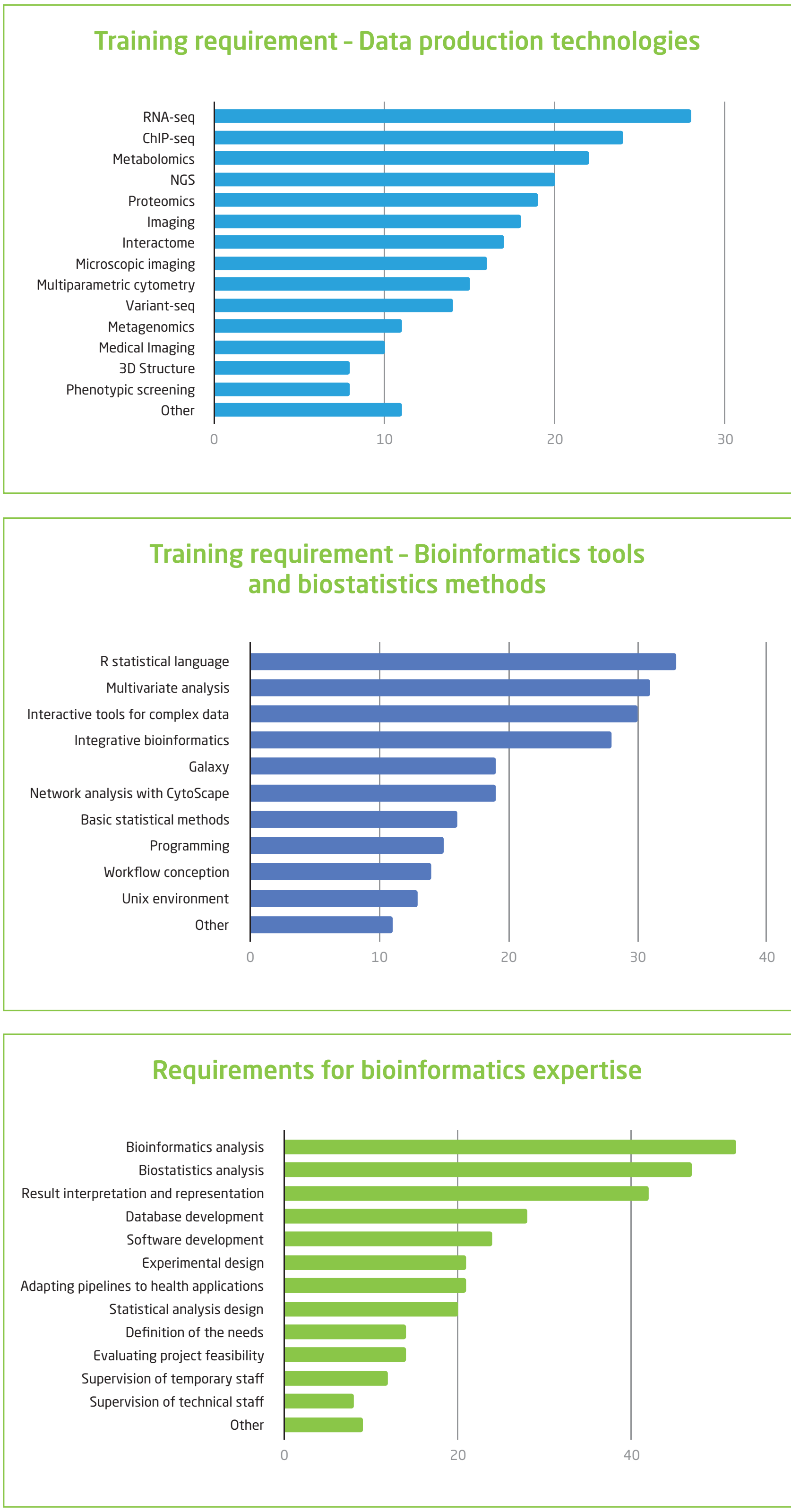


Figure 12
Synthesis of the user needs expressed by INSERM teams for bioinformatics services.

IFB PLATFORMS

www.france-bioinformatique.fr/en/platforms

IFB roots its power of action in the regional platforms who maintain daily interactions with end-users, organise the services on their local facilities, share their expertises and efforts in the mutualised task forces, take responsibilities in the organisation of the IFB work plan and contribute to the strategic decisions about the organisation and future of the IFB network.

The following pages are dedicated to a brief presentation of each platform with a synthetic description of its activities, contribution to the IFB services and main achievements in 2017–2018.

ABiMS

(Analysis and Bioinformatic for Marine Science)

HEAD:

Corre Erwan

🌐 http://abims.sb-roscoff.fr/

The mission of the ABiMS platform is to assist researchers of the marine community and, more broadly, of the life sciences, in the bioinformatic analysis of their data as well as in the development of software and databases. It is one of the national platforms of the French Institute of Bioinformatics (IFB). It is also connected to the EMBRC infrastructure, is part of the IBiSA network via the regional BioGenOuest project and is ISO 9001: 2015 certified. Through its numerous interactions with research units, ABiMS is involved in several projects, with national and European impacts involving bioanalysis activities, software, and e-Infrastructures Development.

SERVICES

ABiMS provides a computing and storage infrastructure (2000 cores and 2 PB) associated with more than 500 softwares dedicated to data analysis. It provides expertise in VREs deployment for data analysis (Galaxy) and genome annotation (Gmod / Apollo), software engineering, bioanalysis and training.

MAIN 2017-2018 ACHIEVEMENTS

- Implication in the establishment of the national IFB-core cluster
- Participation in the ELIXIR Galaxy Community
- Development of the marine micro eukaryotic transcriptomic database
- Involvement in the FG Phaeoexplorer project
- Development and supply of marine observation databases
- Hosting several genome projects of marine cyanobacteria, fungi and macroalgae.
- Further developments within the W4M project

SELECTED PUBLICATIONS

❶
Lebreton,A., Meslet-Cladière,L., Morin-Sardin,S., Coton,E., Jany,J.-L., Barbier,G. and Corre,E. (2018) Comparative analysis of five Mucor species transcriptomes. Genomics, 10.1016/j.ygeno.2018.09.003.

❷
Batut,B., Hiltemann,S., Bagnacani,A., Baker,D., Bhardwaj,V., Blank,C., Bretaudeau,A., Brillet-Guéguen,L., Čech,M., Chilton,J., et al. (2018) Community-Driven Data Analysis Training for Biology. Cell Systems, 6, 752-758.e1.

❸
Ten Hoopen,P., Finn,R.D., Bongo,L.A., Corre,E., Fosso,B., Meyer,F., Mitchell,A., Pelletier,E., Pesole,G., Santamaria,M., et al. (2017) The metagenomic data life-cycle: standards and best practices. GigaScience, 6.

ATGC

HEAD:

Lefort Vincent, Rivals Eric, Guindon Stéphane

🌐 http://www.atgc-montpellier.fr/

The ATGC bioinformatics platform provides services for evolutionary, comparative and functional genomics. It highlights the methodological developments carried out by the Methods and Algorithms for Bioinformatics (MAB) team at LIRMM. These developments are disseminated to the scientific community as freely accessible services from the platform's website: http://www.atgc-montpellier.fr. Most of the tools can be run directly online, via a dedicated web interface from which each user can load his own data. These interfaces have been designed to facilitate the presentation of tools, analysis and interpretation of results.

SERVICES

More than 20 original bioinformatic tools are available on the ATGC platform website. We also organize 3 courses every year:

- Molecular phylogeny
- Bioinformatics for the processing of high throughput sequencing data
- Linux and script for bioinformatics

MAIN 2017-2018 ACHIEVEMENTS

According to Web of Science, the article describing the SMS software (DOI: 10.1093/molbev/msx149) is a 'Hot Paper'.

SELECTED PUBLICATIONS

❶
Lefort, V., Longueville, J.-E. & Gascuel, O. SMS: Smart Model Selection in PhyML. Molecular Biology and Evolution 34, 2422-2424 (2017).

❷
Paulet, D., David, A. & Rivals, E. Ribo-seq enlightens codon usage bias. DNA Research 24, 303-210 (2017).

❸
Chakiachvili, M., Milanesi, S., Arigon Chifolleau, A.-M. & Lefort, V. WAVES: a web application for versatile enhanced bioinformatic services. Bioinformatics 35, 140-142 (2018)

AUBI

HEAD:

Mahul Antoine, Peyret Pierre

https://mesocentre.uca.fr/projets-associes/plateforme-aubi/

AuBi is the Clermont bioinformatics platform for life sciences (fundamental biology, microbiology, agronomy, environment, health and epidemiology). AuBi leans on UCA Mesocentre to promote access to computing facilities for data analysis, storage, training in bioinformatics and hosting web services for research. The platform is committed to serving UCA teams and Associates (public sector, companies ...). Important skills are developed to support scientific projects in the field of large scale sequence analysis in genomics (assembling, annotation, variant analysis, DNAseq, ...), transcriptomics (RNAseq, ChIPseq, ...) and epigenomic variations (BSseq), metagenomics, metatranscriptomics, metabolomics, molecular dynamics but also statistics and imaging.

SERVICES

- Data analysis with the addition of a SMP computing, data storage,
- Bioinformatic training and web services for bioinformatic analysis for examples :
  http://pangenehome.lmge.uca.fr/
  http://phrogs.lmge.uca.fr
  http://g2im.u-clermont1.fr/kaspod/index.php
  http://g2im.u-clermont1.fr/hispod/index.php
  http://g2im.u-clermont1.fr/serimour/phylarray/
  http://g2im.u-clermont1.fr/serimour/goarrays.html
  http://metavir-meb.univ-bpclermont.fr/
  http://panam-meb.univ-bpclermont.fr/

MAIN 2017-2018 ACHIEVEMENTS

We focused on opening a Galaxy bioanalysis platform in late 2018 for research teams. We also organized a training in Galaxy administration after the Oslo Galaxy workshop in 2018.

SELECTED PUBLICATIONS

❶

Batut, B. et al. ASaiM: a Galaxy-based framework to analyze microbiota data. GigaScience 7, (2018).

❷

Amato, P. et al. Active microorganisms thrive among extremely diverse communities in cloud water. PLOS ONE 12, e0182869 (2017).

BILILLE

HEAD:

Touzet Hélène, Marot Guillemette

https://wikis.univ-lille.fr/bilille/

Bilille is the Lille bioinformatics platform, which was founded in 2016 under the joint supervision incentive of the University of Lille, CNRS, INSERM, Institut Pasteur de Lille, CHU Lille and Inria. It has expertise in bioinformatics, biostatistics and bioanalysis and offers a complete service to Lille research labs in biology and health. This includes: support for scientific projects, HPC resources, training etc. The application fields cover analyses of omics data (genomics, transcriptomics, proteomics,...), genome annotation and evolution, integrative biology, cytometry, phenotypic screening, glycobiology...

SERVICES

- Training for biologists and bioinformaticians
- Scientific bioinformatics colloquium
- Participation to biosphere

MAIN 2017-2018 ACHIEVEMENTS

- More than 20 projects related to the analysis of omics data (genomics and transcriptomics mainly)
- Development of new projects in cytometry and high throughput screening

SELECTED PUBLICATIONS

❶

Pericard, P., Dufresne, Y., Couderc, L., Blanquart, S. & Touzet, H. Matam: reconstruction of phylogenetic marker genes from short sequencing reads in metagenomes. Bioinformatics 34, 585-591 (2017).

❷

Couvin, D. et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Research 46, W246-W251 (2018).

BIRD

HEAD:

Redon Richard, Jérémie Bourdon, Bihouée Audrey

http://pf-bird.univ-nantes.fr

BiRD is co-managed by the ITX and the LS2N, employs six computational biologists.
Thanks to the skills of this dedicated and highly qualified staff, BiRD advises, proposes and develops bioinformatic services based on high-throughput sequencing data.
BiRD has expertise in large-scale data analysis and developed dedicated bioinformatics workflows that standardize processing from raw data to biological significance.
Based on these expertises we propose training to our users on data analysis or programming languages. These services are supported by a dedicated computing and storage infrastructure which is remotely accessible through several services and open to all scientists, regardless of their host institution.

SERVICES

BiRD provides computing and storage infrastructure for bioinformatics representing 720 cores and 450TB (Cluster, OpenStack Cloud integrated into the Biosphere project). This infrastructure provides appropriate bioinformatics resources for the processing and analysis of omics data. Based on our expertises we propose support on NGS data analysis and trainings.

MAIN 2017-2018 ACHIEVEMENTS

- BiRD in partnership with the genomics core facility is laureate of the «Integrative Research Clusters» call for proposals of the I-SITE NExT, the objective of which is to bring together an interdisciplinary research community around a common objective: to anticipate the emergence of systemic medicine by co-developing 3 large-scale genomic screening approaches: genome sequencing of patient populations, genomic profiling on single cells and metagenomics applied to microbiota(s).
- Implementation of a new 2.5-days training course on RNAseq data analysis under GALAXY. This training is intended for biologists wishing to use data from RNAseq experiments.

SELECTED PUBLICATIONS

❶ Kilens, S. et al. Parallel derivation of isogenic human primed and naive induced pluripotent stem cells. Nature Communications 9, (2018).

❷ Mandakovic, D. et al. Structure and co-occurrence patterns in microbial communities under acute environmental stress reveal ecological factors fostering resilience. Scientific Reports 8, (2018).

❸ Bézie, S. et al. Ex Vivo Expanded Human Non-Cytotoxic CD8+CD45RClow/− Tregs Efficiently Delay Skin Graft Rejection and GVHD in Humanized Mice. Frontiers in Immunology 8, (2018).

BISTRO-PF

HEAD:

Cognat Valérie, Thompson Julie

http://bioinfo-bistro.fr/bioinfo-bistro/

The BISTRO platform brings together bioinformatics teams and services from the Strasbourg site: GMGM, IBMC, IBMP, ICube, IGBMC, IPHC and LGM. The platform offers expertise, bioinformatics tools and resources, as well as data mining algorithms, focused on evolutionary and functional analyses in various application fields, including biomedical, plant, yeast and bacterial studies.

SERVICES

BISTRO maintains databases and analytical software in integrative plant biology, genetic diseases, comparative genomics, proteomics, non-coding RNA, and microorganisms.
We are involved in the development of the IFB Cloud, training on the Galaxy framework, and the pilot project for integrative bioinformatics.

MAIN 2017-2018 ACHIEVEMENTS

- Implementation and deployment of OrthoInspector, the database of orthology relationships with the largest worldwide coverage in terms of number of species.
- Integration of cutting-edge proteomics analysis tools in the GalaxEast Galaxy platform.
- Creation of a number of integrated knowledge bases dedicated to human rare genetic diseases.

SELECTED PUBLICATIONS

❶ Carapito, C. et al. Validating Missing Proteins in Human Sperm Cells by Targeted Mass-Spectrometry- and Antibody-based Methods. J Proteome Res. 16, (2017).

❸ Cognat, V. et al. The nuclear and organellar tRNA-derived RNA fragment population in Arabidopsis thaliana is highly dynamic. Nuc Acids Res 45, (2017).

BORDEAUX BIOINFORMATICS CENTER

(CBIB)

HEAD:

Groppi Alexis, Nikolski Macha

http://www.cbib.u-bordeaux.fr/

The CBiB is a bioinformatics core facility that provides access to high-performance computing resources, biological data analysis and programming expertise. The resources serve scientists and private labs to fulfill the bioinformatics needs of their research in an efficient and cost-effective manner. We offer state-of-the-art technologies for working with clinical, translational, and basic science data – from acquisition and storage to analysis and sharing. Our resources are secure and standards-compliant. From a few samples to several tens of thousands, the Bioinformatics Centre provides complete DNA, RNA, metabolomics, proteomics as well as image data analysis and integration services.

SERVICES

The CBiB provides access to its computing and secure data storage infrastructure, standard "omics" data analysis pipelines, as well as tailored services such as database development and deployment, bioinformatics methods development in particular for "big data" and data integration approaches.

MAIN 2017-2018 ACHIEVEMENTS

The CBiB has been part of the Human Cell Atlas Pilot projects funded by the Chan-Zuckerberg Initiative. In this project we have developed DypFISH, an approach to investigate the relationship between mRNA and corresponding protein distributions over time by combining micropatterning of cells at high resolution with innovative downstream image and statistical analysis. We introduced a range of analytical techniques for quantitatively interrogating single molecule RNA FISH data in combination with protein immunolabeling. The preprint is available on bioarxiv (https://doi.org/ 10.1101/536383).

SELECTED PUBLICATIONS

❶

Henriet, E. et al. Argininosuccinate synthase 1 (ASS1): A marker of unclassified hepatocellular adenoma and high bleeding risk. Hepatology 66, 2016-2028 (2017).

❷

Hooks, K. B., Konsman, J. P. & O'Malley, M. A. Microbiota-gut-brain research: a critical analysis. Behavioral and Brain Sciences 1-40 (2018). doi:10.1017/

❸

Hemadou, A. et al. Pacific Biosciences Sequencing and IMGT/HighV-QUEST Analysis of Full-Length Single Chain Fragment Variable from an In Vivo Selected Phage-Display Combinatorial Library. Frontiers in Immunology 8, (2017)

CENTRE DE BIOINFORMATIQUE, BIOSTATISTIQUE ET BIOLOGIE INTÉGRATIVE

(C3BI)

HEAD:

Gascuel Olivier, Malabat Christophe, Dillies Marie Agnès

https://c3bi.pasteur.fr/

Created in 2015, the C3BI includes 5 research units in Computational Biology and the Hub of Bioinformatics and Biostatistics.
The mission of the center is to develop methodological research in bioinformatics and biostatistics, give visibility of the Institut Pasteur in this field on an international level, offer support to experimental research units, and develop the computational and analysis skills of the campus. The platform activities comprise participation in research and analysis projects, on-site assignments within campus units and platforms, training and teaching sessions open to our partners, and provide a number of resources to the national and international community.

SERVICES

The C3BI platform provides a number of publicly available Resources :

- Web services like the Galaxy web portal (543 tools, 1448 CPUs, 24 To RAM), NGPhylogeny.fr, Booster, JASS, AriaWeb
- Open databases like macsyDB, CONJDB, iPPI-DB, listeriomics
- bioanalysis tools like Goalign/Gotree, MacSyFinder, SARTOOLS, GRAAL

Complete list on c3bi.pasteur.fr/hub/tools/

MAIN 2017-2018 ACHIEVEMENTS

- Creation of a new research unit on sequence algorithms
- Inauguration of a new building to host 130 scientists in Computational Biology
- 300h of teaching and training in bioinformatics and biostatistics including for instance Hands on NGS, molecular phylogeny, and a « Single Cell » course co-organized with multiple IFB partners

SELECTED PUBLICATIONS

❶

Lemoine, F. et al. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature 556, 452-456 (2018).

❷

Ruppé, E. et al. Prediction of the intestinal resistome by a three-dimensional structure-based method. Nature Microbiology 4, 112-123 (2018).

eBIO

HEAD:
Toffano Nioche Claire, Gautheret Daniel
http://ebio.u-psud.fr/

The eBio Bioinformatics platform (http://ebio.u-psud.fr/), created in December 2009, is at the heart of the University of Paris-Sud's bioinformatics system. eBio is certified by IBISA/Reseau National des Plateformes Bioinformatiques (RENABI) and a member of the Alliance des Plateformes Bioinformatiques d'Ile de France (APLIBIO). The main site of eBio is located at the University of Paris-Sud's 400th building and integrated into the I2BC (UMR9198, Gif sur Yvette).

eBio's missions include: bioinformatics support to research projects and development of bioinformatics web services and genomic databases. The platform has supported or hosted more than 50 research projects since early 2010.

MAIN 2017-2018 ACHIEVEMENTS

- Redesign of the microbial CRISPR loci prediction software with the addition of a decision criterion, a confidence score, and the prediction of the associated CAS proteins.
- DE-kupl software: exhaustive capture of biological variation in RNA-seq data through k-mer decomposition.

SELECTED PUBLICATIONS

❶
Couvin, D. et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Research 46, W246-W251 (2018).

❷
Audoux, J. et al. DE-kupl: exhaustive capture of biological variation in RNA-seq data through k-mer decomposition. Genome Biology 18, (2017).

GENOTOUL BIOINFO

HEAD:
Klopp Christophe, Hoede Claire, Gaspin Christine
http://bioinfo.genotoul.fr/

The platform is located in Toulouse, south of France. Its goal is to bring together equipment resources and human skills that offer to life science research programmes an access to state of the art know-how and top level technologies. During past years the team has build up an expertise in diverse applications of sequence analysis. This knowledge has been used in software and pipeline developments as well as in data analysis projects including genome assembling (short/long reads), annotation (coding/non coding), (s)RNA-seq, methyl-seq data and variant analyses, metagenomics (metabarcoding/whole genome) and, more recently, data integration. Supported communitities are agriculture, alimentation, human health, ecology and bioinformatics.

SERVICES

- Access to high-performance computing and storage resources, updated international databanks and software,
- Environment for the development and deployment of novel resources (databases, software)
- Support for data analysis (mainly data produced by HTS technologies),
- Development of novel bioinformatics methods/tools,
- Organization of training sessions and knowledge sharing.

MAIN 2017-2018 ACHIEVEMENTS

Main 2017-2018 achievements mainly include the evolution of the infrastructure architecture, our contribution to non coding RNA identification in microsporidia organisms and the publication of MixKernel and D-GENIES tools. mixKernel aims at providing methods to combine kernel for unsupervised exploratory analysis. Different solutions are provided to compute a meta-kernel, in a consensus way or in a way that best preserves the original topology of the data. D-GENIES is a standalone and WEB application performing large genome alignments using minimap2 software package and generating interactive dot plots.

SELECTED PUBLICATIONS

❶
Mariette, J. & Villa-Vialaneix, N. Unsupervised multiple kernel learning for heterogeneous data integration. Bioinformatics 34, 1009-1015 (2017).

❷
Cabanettes, F. & Klopp, C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ 6, e4958 (2018).

❸
Belkorchia, A. et al. Comparative genomics of microsporidian genomes reveals a minimal non-coding RNA set and new insights for transcription in minimal eukaryotic genomes. DNA Research 24, 251-260 (2017).

GENOUEST

HEAD:
Collin Olivier, Nicolas Jacques
http://www.genouest.org/

Hosted at INRIA-IRISA, the GenOuest core facility offers a complete bioinformatics environment with hardware and software infrastructure, public databases, software and workflows, all with associated support. The technological portfolio rely on several computing resources (cluster, cloud, docker, galaxy portal), data management solutions (BioMAJ). During the years GenOuest has invested in the development of data-centered tools. Thanks to the CeSGO project, GenOuest offers a set of collaborative tools to manage projects and scientific data in the best possible way. GenOuest offers development of bioinformatics applications as well as training and technological transfer of new tools developed by research teams of the Institute.

SERVICES

- Access to its computational infrastructure: cluster, cloud, docker environment.
- IFB-core cloud support system
- Monitoring of the Biosphere cloud federation.
- Hosting more than 30 scientific services and databases freely accessible to the community

MAIN 2017-2018 ACHIEVEMENTS

- The main achievement of these last years is the development of the CeSGO environment that allows scientists to manage their projects and data in tight conjunction with the computing environment. CeSGO is operational since 2017 and will help to foster good practices for FAIR data management.
- GenOuest is involved in Biosphere, the french bioinformatic cloud federation with other core facilities.
- With ABiMS and BiRD, GenOuest has set up BioDataRing, a data replication infrastructure based on iRODS.

SELECTED PUBLICATIONS

❶
Belkorchia, A. et al. Comparative genomics of microsporidian genomes reveals a minimal non-coding RNA set and new insights for transcription in minimal eukaryotic genomes. DNA Research 24, 251-260 (2017)

❷
Darde, T. A. et al. TOXsIgN: a cross-species repository for toxicogenomic signatures. Bioinformatics 34, 2116-2122 (2018).

❸
Gruening, B. et al. Recommendations for the packaging and containerizing of bioinformatics software. F1000Research 7, 742 (2018).

iCONICS

HEAD:
Durrleman Stanley, Moszer Ivan
https://icm-institute.org/fr/iconics-the-bioinformatics-and-biostatistics-core-facility/
https://neuroinformatics.icm-institute.org/

The iCONICS platform (Brain and Spine Institute – ICM) develops and makes available software solutions and methodological expertise to meet three important needs for biomedical research: data curation, standardization, annotation, structuration, integration and visualization (data managers, software engineers); high-throughput omics data processing, including NGS data such as whole-exome, RNA-seq (bioinformaticians); basic and advanced statistical analysis, especially integration of multimodal and high-dimensional data (biostatisticians). The platform thus supports scientific and clinical teams at every step of their research projects: from study design to data management, processing, analysis and interpretation, considering a wide variety of data acquisition approaches (clinical, imaging, genomics, etc.).

SERVICES

The iCONICS platform mainly contributes to project support activities of IFB: it provides biologists and clinicians with an expert assistance for dealing with clinical, omics and imaging data in biomedical studies. The operational head of the platform is a member of the IFB Directorate and coordinator of two specific actions.

MAIN 2017-2018 ACHIEVEMENTS

- Creation of the Center for Neuroinformatics of the ICM, designed as a distributed structure that aims at harmonizing and sharing best practices in data management and analytics across the Institute (FAIR principles).
- Start of a new activity in data management (clinical, imaging and molecular data).
- Conception of a software architecture for building the data lake and data warehouse of the ICM.

SELECTED PUBLICATIONS

❶
Garali, I. et al. A strategy for multimodal data integration: application to biomarkers identification in spinocerebellar ataxia. Briefings in Bioinformatics 19, 1356-1369 (2017).

❷
Marie, C. et al. Oligodendrocyte precursor survival and differentiation requires chromatin remodeling by Chd7 and Chd8. Proceedings of the National Academy of Sciences U S A 115, E8246-E8255 (2018).

IGR

HEAD:

Meurice Guillaume, Gautheret Daniel

http://gustaveroussy.fr/fr/content/plateforme-de-bioinformatique-activites

The platform is organised around 3 major areas: data management (structuring of data from from robots), tool development (web applications, databases, pipelines) and support for research projects in an oncogenomic context. We now have a set of conventional pipelines to analyze WES, WGS, ChIP-Seq, RNA-Seq, TGS, microarray oncoscan/cytoscan.

SERVICES

Training in exome data analysis and RNA-seq (Aviesan, Cancéropole Ile de France)

MAIN 2017-2018 ACHIEVEMENTS

Package Eacon: consistent copy number analysis from microarray and WES sources. https://github.com/gustaveroussy/EaCoN

SELECTED PUBLICATIONS

❶

Selimoglu-Buet, D., Rivière, J., Ghamlouch, H., Bencheikh, L., Lacout, C., Morabito, M., ... Solary, E. (2018). A miR-150/TET3 pathway regulates the generation of mouse and human non-classical monocyte subset. Nature communications, 9(1), 5455. doi:10.1038/s41467-018-07801-x

❷

Damien D., Daniel G., Stefan M., A benchmark study of scoring methods for non-coding mutations, Bioinformatics, Volume 34, Issue 10, 15 May 2018, Pages 1635-1641

INSTITUT CURIE

HEAD:

Hupe Philippe, Barillot Emmanuel

http://u900.curie.fr/

The expertise of the platform is versatile on many aspects of high-throughput data management, processing, integration and statistical and functional analysis in biology and in clinics.
We cover fundamental, translational and clinical research, including clinical trials.
In all these domain, the platform also develop appropriate methods, implement tools and automatic pipelines, and package and release them publicly or grant on-line access to the community.
Software optimisation for high-performance computing is also part of our know-how.
Finally the platform has developed an experience in training biologists, clinicians and bioinformaticians in all above-mentioned fields.

SERVICES

- Collaboration for high-throughput data analysis
- Training in bioinformatics

MAIN 2017-2018 ACHIEVEMENTS

- Pipeline to analyze CRISPR screens
- Pipelines for single-cell data analysis
- Pipeline for proteogenomics
- Contribution to the activity of the SeqOIA platform (France Médecine Génomique 2025)
- Certification by the COFRAC of our activities to support the routine oncologic diagnostic for the analysis NGS
- Information system to support real-time analysis for clinical trials

SELECTED PUBLICATIONS

❶

Basse, C. et al. Relevance of a molecular tumour board (MTB) for patients' enrolment in clinical trials: experience of the Institut Curie. ESMO Open 3, e000339 (2018).

❷

Cacheux, W. et al. Exome sequencing reveals aberrant signalling pathways as hallmark of treatment-naive anal squamous cell carcinoma. Oncotarget 9, (2017).

❸

Girard, E. et al. Familial breast cancer and DNA repair genes: Insights into known and novel susceptibility genes from the GENESIS study, and implications for multigene panel testing. International Journal of Cancer 144, 1962-1974 (2018).

MBI

HEAD:

Ritchie Dave, Devignes Marie Dominique

https://mbi.loria.fr/

The MBI platform stands for "Modelling of Biomolecules and their Interactions". It is essentially a research platform offering services in the frame of scientific collaborative projects. The main activities of the platform concern 3D modelling of proteins in interaction with ligands, proteins, ss RNA or DNA, as well as functional annotation of biomolecules. Available programs include molecular dynamic simulation (NAMD software), very fast rigid docking and shape comparison programs, which are run on hybrid clusters of CPUs and GPUs.

Data science expertise (a.k.a. AI), including symbolic data mining, machine learning, complex graph analysis, is also available for bioinformatics and biomedical applications.

SERVICES

None. A Galaxy instance for structural bioinformatics is under construction in 2019.

MAIN 2017-2018 ACHIEVEMENTS

The MBI platform was successfully used during a hands-on tutorial presented by I. Chauvot de Beauchene about RNA-protein docking, at the Winter School AlgoSB, 14-18 janvier 2019, in Marseille.

SELECTED PUBLICATIONS

❶ Alborzi, S. Z., Ritchie, D. W. & Devignes, M.-D. Computational discovery of direct associations between GO terms and protein domains. BMC Bioinformatics 19, (2018).

❷ Couvineau, P., de Almeida, H., Maigret, B., Llorens-Cortes, C. & Iturrioz, X. Involvement of arginine 878 together with Ca2+ in mouse aminopeptidase A substrate specificity for N-terminal acidic amino-acid residues. PLOS ONE 12, e0184237 (2017).

❸ Dey, S., Ritchie, D. W. & Levy, E. D. PDB-wide identification of biological assemblies from conserved quaternary structure geometry. Nature Methods 15, 67-72 (2017).

MICROSCOPE

HEAD:

Dubois Mathieu, Médigue Claudine, Vallenet David

www.genoscope.cns.fr/agc/microscope/

The LABGeM team from the CEA/Genoscope has developed MicroScope, a web-based platform for prokaryotic genome analysis and expert functional annotation.

MicroScope combines tools and graphical interfaces to analyze genomes in a comparative and metabolic context.

This platform provides data from thousands of genome projects together with post-genomic experiments allowing users to improve the understanding of gene functions. Expert annotations are continuously gathered in the MicroScope database contributing to the improvement of the quality of microbial genome annotations.

MicroScope can be used as a community resource for comparative analysis and annotation of publicly available genomes but also as a private resource with restricted access rights on genomic data.

SERVICES

- Microbial genome annotation (complete, WGS or metagenome-assembled genomes)
- Comparative genomics and pangenomics
- Function and biological process predictions
- Metabolic network reconstruction
- Analysis of transcriptomics data
- Trainings on prokaryotic genome annotation and on the curation of metabolic networks

MAIN 2017-2018 ACHIEVEMENTS

- Integration of >200 genomes per month
- More than 10K genomes analyzed so far
- >4,200 user accounts +1,100 accounts since 2017
- 3,000 human-expertized annotations a month among a total of 370K gene functions manually reviewed
- Organization of the first MicroScope open days
- >200 citations since 2017
- ISO 9001:2015 and NF X50-900:2016 certifications

SELECTED PUBLICATIONS

❶ Médigue, C. et al. MicroScope-an integrated resource for community expertise of gene functions and comparative analysis of microbial genomic and metabolic data. Briefings In Bioinformatics 2017 Sep 12 (2017)

❷ Vallenet, D. et al. MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. Nucleic Acids Research 45, D517-D528 (2017).

❸ Borriss, R. et al. Bacillus subtilis, the model Gram-positive bacterium: 20 years of annotation refinement. Microbial Biotechnology 11, 3-17 (2017).

MIGALE

HEAD:
Loux Valentin, Schbath Sophie
http://migale.jouy.inra.fr/

The Migale bioinformatics platform is a team of the INRA MaIAGE research unit (Applied Mathematics and Computer Science, from Genome to the Environment).
Since 2003, it is provides four type of services to the life sciences community an open infrastructure dedicated to life sciences data analysis (500 Tb, 1000 equivalent CPUs) dissemination of expertise in bioinformatics (annual "Bioinformatic by practice" training session)design and development of bioinformatics applications (genome browser, databases)data analysis in genomics, metagenomics and metatrascriptomics

SERVICES

- Access to infrastructure
- Galaxy server
- Annual training session
- Analysis services genomics, comparative genomics, meta genomics and metatranscriptomics
- Bioinformatics development : Galaxy (proteomics, metagenomics), databases

MAIN 2017-2018 ACHIEVEMENTS

- ProteoRE (Proteomics Research Environment) aims at centrally providing the proteomics community with an online research service enabling biologists/clinicians without programming expertise to perform the functional analysis and to interpret their proteomics data through the Web
- More than 100 person trained each year
- Florilege web application combines the information computed by advanced text-mining of the scientific literature and information from Biological Resource Centers.
- Easy16s, a convivial web interface to explore metabarcoding results
- Bioinfomatcis analysis of France Génomique project: analysis diversity of 44 AOP French cheese

SELECTED PUBLICATIONS

❶
Delforno, T. P. et al. Comparative metatranscriptomic analysis of anaerobic digesters treating anionic surfactant contaminated wastewater. Science of The Total Environment. Volume 649, 482-494

❷
Poirier, S. et al. Deciphering intra-species bacterial diversity of meat and seafood spoilage microbiota using gyrB amplicon sequencing: A comparative analysis with 16S rDNA V3-V4 amplicon sequencing. PLOS ONE 13, e0204629 (2018).

❸
Christo-Foroux, E. et al. Manual and expert annotation of the nearly complete genome sequence of Staphylococcus sciuri strain ATCC 29059: A reference for the oxidase-positive staphylococci that supports the atypical phenotypic features of the species group. Systematic and Applied Microbiology 40, 401-410 (2017).

MMG-GBIT

HEAD:
Salgado David, Béroud Christophe
https://geneticsandbioinformatics.eu/

The MMG-GBIT IFB platform is linked to the Bioinformatics & Genetics research team of Inserm U1251. It benefits from this strong interaction and provide users with the team expertise in human genetics, rare diseases and oncogenetics as well as NGS data analysis.
It provides international reference systems to collect, annotate, filter and interpret human genetics data in relation to diseases. These tools, databases, registries and observatories have already received millions of worldwide queries. In addition, we provide trainings and can assist researchers for any bioinformatics project or tools development related to human genetics from design to analysis.

SERVICES

- Development and deployment of bioinformatics systems and tools
- Development and curation of patients' databases; national mutations databases (SNV and CNV); and international Locus Specific Databases
- Project support
- Organization of training in bioinformatics

MAIN 2017-2018 ACHIEVEMENTS

- Representative of the French ELIXIR node in various communities
- Leader of the new ELIXIR Human CNV community
- Development and deployment of new bioinformatics systems and tools (VarAFT; Crawfish; SKIP-e)
- Continuous development and curation of patients' databases; national mutations databases (SNV and CNV); and international Locus Specific Databases
- Participation to the JOBIM 2018 organization
- Co-organization of the ELIXIR 2018 biohackathon
- Participation as an IFB platform representative to the ELIXIR-SME forum
- Participation to international trainings to promote databases and tools (3Gb-test; RD-Connect; HVP-VEP and various local trainings)
- Various project supports leading to publications

SELECTED PUBLICATIONS

❶
Desvignes, J.-P. et al. VarAFT: a variant annotation and filtration system for human next generation sequencing data. Nucleic Acids Research 46, W545-W553 (2018).

❷
Zaïda Koeks et al. Clinical Outcomes in Duchenne Muscular Dystrophy: A Study of 5345 Patients from the TREAT-NMD DMD Global Database. JND 4, 293-306 (2017).

❸
Jallades, L. et al. Exome sequencing identifies recurrent BCOR alterations and the absence of KLF2 , TNFAIP3 and MYD88 mutations in splenic diffuse red pulp small B-cell lymphoma. Haematologica 102, 1758-1766 (2017).

ORPHANET

HEAD:

Rath Ana, Hanauer Marc

https://www.orpha.net/

Orphanet is a knowledge base on rare diseases and orphan drugs. It aims to curate and generate data on rare diseases so as to improve the diagnosis, care and treatment of patients. The Orphanet website (www.orpha.net) aims to provide high-quality information on rare diseases, and ensure equal access to knowledge for all stakeholders. Data from the knowledge base is available for researchers on www.orphadata.org and in the Orphanet Rare Disease Ontology. Orphanet also maintains the Orphanet rare disease nomenclature (ORPHA number), essential in improving the visibility of rare diseases in health and research information systems, acting as interoperability vector.

SERVICES

Orphanet provides a number of tools to improve the interoperability of data concerning rare diseases across the fields of health care and research: a standard nomenclature (ORPHA numbers), an ontology (Orphanet Rare Disease Ontology), and datasets for research purposes (cross-referencing of terminologies, epidemiological data, diseases-gene relationships, disease-phenotype annotations).

MAIN 2017-2018 ACHIEVEMENTS

In 2017, Orphanet marked its 20th anniversary. Orphanet's data download platform www.orphadata.org was designated as an ELIXIR Core Data Resource at the start of 2019 after adopting at the end of 2018 a CC BY licence and making the Orphanet epidemiological datasets open to all users via this licence. Orphanet launched HOOM, the Human Phenotype Ontology-Orphanet Rare Disease Ontology Ontological Module, in order to provide users with the possibility explore annotations between a clinical entity and phenotypic abnormalities according to a frequency and by integrating the notion of diagnostic criterion.

SELECTED PUBLICATIONS

❶ Maiella, S. et al. Harmonising phenomics information for a better interoperability in the rare disease field. European Journal of Medical Genetics 61, 706–714 (2018).

❷ Rath, A. et al. A systematic literature review of evidence-based clinical practice for rare diseases: what are the perceived and real barriers for improving the evidence and how can they be overcome? Trials 18, (2017).

❸ Pavan, S. et al. Clinical Practice Guidelines for Rare Diseases: The Orphanet Database. PLOS ONE 12, e0170365 (2017).

INCA-SLC

HEAD:

Viari Alain

http://synergielyoncancer.fr

The Synergie-Lyon-Cancer (SLC) Bioinformatics Facility was initiated in 2009 by Pr. Gilles Thomas to promote the exploitation of massive amounts of data being generated from large scale genomic projects in the field of Cancer Research. It is hosted at the Centre Léon Bérard in Lyon. The staff currently includes ten people under the supervision of Alain Viari. The facility has focused its efforts on the management and analysis of large volumes of NGS data, mostly whole genome sequencing (WGS). It also developed a strong expertise in transcriptomic analysis using various NGS techniques.

SERVICES

Software available : wginr, criscross and participation to the Osiris INCa/Siric initiative to define the minimum data set for the sharing of clinical-biological data in oncology.

MAIN 2017-2018 ACHIEVEMENTS

Completion of WGS analysis of prostate cancers from african caribbean men versus European ancestry men. Work done in the context of International Cancer Genome Consortium.

SELECTED PUBLICATIONS

❶ Huet S., Tesson B., Salles G. A gene-expression profiling score for prediction of outcome in patients with follicular lymphoma: a retrospective training and validation analysis in three international cohorts. Lancet Oncol. 2018 Apr;19(4):549-561.

❷ Davies H., Glodzik D., Nik-Zainal S. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures.Nat Med. 2017 Apr;23(4):517-525.

PACA-BIOINFO

HEAD:

Santini Sébastien, Claverie Jean Michel

https://www.igs.cnrs-mrs.fr/paca-bioinfo/

The PACA Bioinfo services focus on microbial genomics, metagenomics, comparative genomics and molecular phylogeny. It provides the community with various free access online tools proposed without prior registration. The platform also collaborates to genomic or metagenomic projects for which it constitutes the main bioinformatic support, in particular for teams of the Mediterranean Institute of Microbiology in Marseille. Significant progresses have resulted from the genomic analyses of giant viruses, two sponges, and metagenomes from ancient frozen soils (permafrost). The platform also provides generic statistical tools for comparing large datasets of counts (ACD-tools) as encountered in classical or metagenome-based ecological studies.

SERVICES

Phylogeny.fr was first developed for, and used by, non-specialists and is now a reference throughout the world. ACDTool, the latest production of the platform, allows a unique type of statistical comparison of large datasets of counts.

MAIN 2017-2018 ACHIEVEMENTS

- First microbiome analysis of a diatom
- Detail comparative genomics of various families of giant viruses
- Implementation of an extended online version of the highly cited and popular AC-test
- Design of a basic genomic analysis self-serve workflow for the whole Mediterranean Institute of Microbiology (under Galaxy).

SELECTED PUBLICATIONS

❶ Claverie, J.-M. & Ta, T. N. ACDtool: a web-server for the generic analysis of large data sets of counts. Bioinformatics 35, 170-171 (2018).

❷ Zhang, H.-H. et al. Unexpected invasion of miniature inverted-repeat transposable elements in viral genomes. Mobile DNA 9, (2018).

❸ Belahbib, H. et al. New genomic data and analyses challenge the traditional vision of animal epithelium evolution. BMC Genomics 19, (2018).

PRABI-AMSB

HEAD:

Navratil Vincent, Perrière Guy

http://www.prabi.fr/

The PRABI-AMSB (for Analysis and Modelling of Biological Systems) platform proposes bioinformatics services for biologists who need assistance with specific tools or in-depth expertise for more important projects. The expertise available at PRABI-AMSB covers the following areas:

- expression data (RNA-seq).
- interaction data (ChIP-seq, Tap-tag/MS, Yeast Two Hybrid screens).
- metagenomics and metatrans-criptomics.
- comparative genomics and phylogeny.
- genome assembly and annotation.
- systems biology (metabolic, protein interaction and regulatory networks).

SERVICES

- Training (R language, NGS analysis with Galaxy, phylogeny).
- Galaxy service (http://galaxy.prabi.fr/)

MAIN 2017-2018 ACHIEVEMENTS

Ouverture du cluster de calcul aux cinq unités de la FR BioEnviS.

SELECTED PUBLICATIONS

❶ Muyle, A. et al. Genomic imprinting mediates dosage compensation in a young plant XY system. Nature Plants 4, 677-680 (2018).

❷ Galia, W. et al. Strand-specific transcriptomes of Enterohemorrhagic Escherichia coli in response to interactions with ground beef microbiota: interactions between microorganisms in raw meat. BMC Genomics 18, (2017).

PRABI-Lyon-Grenoble

HEAD:
Perrière Guy
http://doua.prabi.fr/main/index

The PRABI-Lyon-Grenoble is a consortium of eight research teams involved in bioinformatics in Lyon and Grenoble. Therefore, the expertise available at PRABI-Lyon-Grenoble covers a broad range of domains:

- molecular phylogeny and evolution.
- metabolic networks.
- biostatistics for medicine.
- multivariate statistics in ecology.
- plants population genetics.
- expression data.
- high-dimension statistics.
- proteomics.
- comparative genomics.
- transposable elements.
- systems biology (metabolic, protein interaction and regulatory networks).

As the PRABI-Lyon-Grenoble is a consortium of research teams, it is involved in many training and teaching activities (from first year students to professionals).

SERVICES

Web service (http://www.prabi.fr/) with an access to sequence databases (EMBL,GenBank, UniProtKB, etc.) and general bioinformatics software. This service also allows to download the programs developed by the different teams (e.g. SeaView, Bio++, TPMS)

SELECTED PUBLICATIONS

❶ Laurin-Lemay, S., Rodrigue, N., Lartillot, N. & Philippe, H. Conditional Approximate Bayesian Computation: A New Approach for Across-Site Dependency in High-Dimensional Mutation-Selection Models. Molecular Biology and Evolution 35, 2819-2834 (2018).

❷ Guéguen, L. & Duret, L. Unbiased Estimate of Synonymous and Nonsynonymous Substitution Rates with Nonstationary Base Composition. Molecular Biology and Evolution 35, 734-742 (2017). Davín, A. A. et al. Gene transfers can date the tree of life. Nature Ecology & Evolution 2, 904-909 (2018).

RPBS

HEAD:
Rey Julien, Tufféry Pierre
http://bioserv.rpbs.univ-paris-diderot.fr/events.html

RPBS is a platform dedicated to structural bioinformatics. It proposes:

- the development of structural bioinformatics methods/protocols.
- service hosting and on-line deployment of services of the field
- training and consulting in the field.
- calculation hosting through PAAS.

SERVICES

- On-line RPBS services are freely available at bioserv.rpbs.univ-paris-diderot.fr and mobyle. rpbs.univ-paris-diderot.fr.
- Services cover protein structure and function analysis and modeling.

MAIN 2017-2018 ACHIEVEMENTS

Recent achievements include: on-line prediction of protein-protein complex structure (InterEvDock2), protein-peptide docking (pepATTRACT, protein loop modeling (DaReUs-Loop).

SELECTED PUBLICATIONS

❶ De Vries, S. J., Rey, J., Schindler, C. E. M., Zacharias, M. & Tuffery, P. The pepATTRACT web server for blind, large-scale peptide-protein docking. Nucleic Acids Research 45, W361-W364 (2017).

❷ Quignot, C. et al. InterEvDock2: an expanded server for protein docking using evolutionary and biological information from homology models and multimeric inputs. Nucleic Acids Research 46, W408-W416 (2018).

SOUTH GREEN

HEAD:
Tando Ndomassi, Pitollat Bertrand, Rouard Mathieu, Ruiz Manuel, Christine Tranchant

http://www.southgreen.fr/

South Green is a bioinformatics platform dedicated to the genomics of tropical and Mediterranean plants and related pathogens. It federates bioinformaticians from different units and institutes of Montpellier (Bioversity, CIRAD, INRA and IRD) with a multidisciplinary expertise in data integration, software development, sequencing data analyses and high-performance computing. South Green ensures the development of original information systems such as GreenPhyl, SNiPlay, Gigwa or AgroLD, and offers bioinformatic pipelines through Galaxy and TOGGLe workflow managers. The platform has built a strong expertise in the development of Genome Hubs, integrated information systems, deployed on several plants and currently being extended to pathogens.

SERVICES

A significant part of activities comprises hands-on trainings that are regularly offered in the local community as well as with partners in Africa and Asia on the following topics: Galaxy, NGS analyses, R, Perl, Linux, HPC administration. Besides, South Green provides access to computing facilities for both users and developers engaged in this scientific area. Finally, South Green provides free access to a wide range of tools and information systems for thousand of users.

MAIN 2017-2018 ACHIEVEMENTS

- Co-organisation of an ELIXIR Training: Plant Genome Assembly and Annotation
- Training course on HPC deployment and NGS data analysis in West Africa
- Co-organisation of the Conference Galaxy 2017 in Montpellier
- Release of new Genome Hub in Cacao and Sugarcane
- Release of GIGWA version 2
- Release of TOGGLe version 0.3.7 and deployed in several national and international clusters

SELECTED PUBLICATIONS

❶
Venkatesan, A. et al. Agronomic Linked Data (AgroLD): A knowledge-based system to enable integrative biology in agronomy. PLOS ONE 13, e0198270 (2018).

❷
Argout, X. et al. The cacao Criollo genome v2.0: an improved version of the genome for genetic and functional genomic studies. BMC Genomics 18, (2017).

❸
Rouard, M. et al. Three new genome assemblies support a rapid radiation in Musa acuminata (wild banana). Genome Biology and Evolution (2018). doi:10.1093/gbe/evy227

TAGC-BCF

HEAD:
Pascal Rihet, Bianca Haberman, Christine Brun

https://tagc.univ-amu.fr/

TAGC is a mixed research & services Inserm units specialised in the development and application of omics approaches to various biological systems. The lab includes a sequencing facility (TGML, member of France Génomique) and a strong team of bioinformaticians who develop and give public access to bioinformatics resources (software tools and databases) in domains ranging from the analysis of single protein and DNA sequences to large-scale analysis and integration of omics data, with a specific focus on genome variations, genetic and epigenetic regulation, and network analysis.

SERVICES

The Web sites and tools developed by our lab include the RSAT software suite, ReMap, MoonDB, Clust&See, hh-motif or mitoXplorer. We also organize hands-on workshops to train users to bioinformatics approaches, including data analysis techniques and code development.

MAIN 2017-2018 ACHIEVEMENTS

- Publication of a major update of the ReMap resource, a high-quality catalogue of regulatory elements, in collaboration with JASPAR team.
- Update of the MoonDB database, which holds information on extreme multifunctional and moonlighting proteins.
- Celebration of the 20th birthday of the RSAT-suite.
- Organisation of JOBIM, the yearly conference regrouping the whole community of French bioinformaticians.

SELECTED PUBLICATIONS

❶
Ribeiro, D. M., Briere, G., Bely, B., Spinelli, L. & Brun, C. MoonDB 2.0: an updated database of extreme multifunctional and moonlighting proteins. Nucleic Acids Research 47, D398-D402 (2018).

❷
Nguyen, N. T. T. et al. RSAT 2018: regulatory sequence analysis tools 20th anniversary. Nucleic Acids Research 46, W209-W214 (2018).

❸
Chèneby, J., Gheorghe, M., Artufel, M., Mathelier, A. & Ballester, B. ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. Nucleic Acids Research 46, D267-D275 (2017).

URGI PLATFORM

HEAD:

Amselem Joelle, Adam Blondon
Anne Françoise
urgi.versailles.inra.fr/Platform

The main missions of the platform are to contribute to an open science compatible management of patrimonial data produced by INRA and its partners and to propose tools and suitable environments (including training) for data analysis to researchers.

The main activities of the platform during the last 5 years were:

- Plant and fungi data integration in the GnpIS information System
- Giving access to computing resources and analysis tools, analysis support

Users training (use of GnpIS, annotation, SNP calling ...)

SERVICES

- Access to data in GnpIS (https://urgi.versailles.inra.fr/gnpis/)
- Data integration of genetic, genomic and phenotypic data for INRA and international partners included the official repository of the International Wheat Genome Sequence.
- Access to computing and associated storage resources (associated with fees depending of volumes and partners).
- Training sessions (Mainly GnpIS navigation and Transposable Elements annotation)

MAIN 2017-2018 ACHIEVEMENTS

- GnpIS is the repository of the IWGSC (the wheat genome sequence consortium) and demonstrate new pilots aiming at improving access to FAIR data at international level
- URGI platform gives access to the developments carried out in the framework of ELIXIR-Excelerate: (i) an ontology for forest trees phenotyping, (ii) an improved version of the MIAPPE standard of metadata for phenotyping experiments, (iii) implementations of the international standard web service BrAPI to give computational access to GnpIS data.
- Development of a VRE (Virtual Research Environment) to annotate Transposable Elements in genomes successfully transferred to several cloud-based e-infrastructures and used in several trainings on Transposable Elements annotation.

SELECTED PUBLICATIONS

❶ Alaux, M. et al. Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. Genome Biology 19, (2018).

❷ Dominguez Del Angel, V. et al. Ten steps to get started in Genome Assembly and Annotation. F1000Research 7, 148 (2018).

ISFINDER

HEAD:

Siguier Patricia
https://www-is.biotoul.fr/

ISfinder is a public database and international reference center for prokaryotic IS and a research tool.

The project is divided into two major types of activity, intimately connected to the ISfinder database. The first involves embellishment of ISfinder itself and the development of tools necessary to exploit the information included in the database. The second lead directly from this and involve detailed analysis of the stored information.

ISfinder activity involves database analysis, facilitating IS classification, identification of new families and provides information concerning their genomic impact (e.g. effects linked to target site choice).

SERVICES

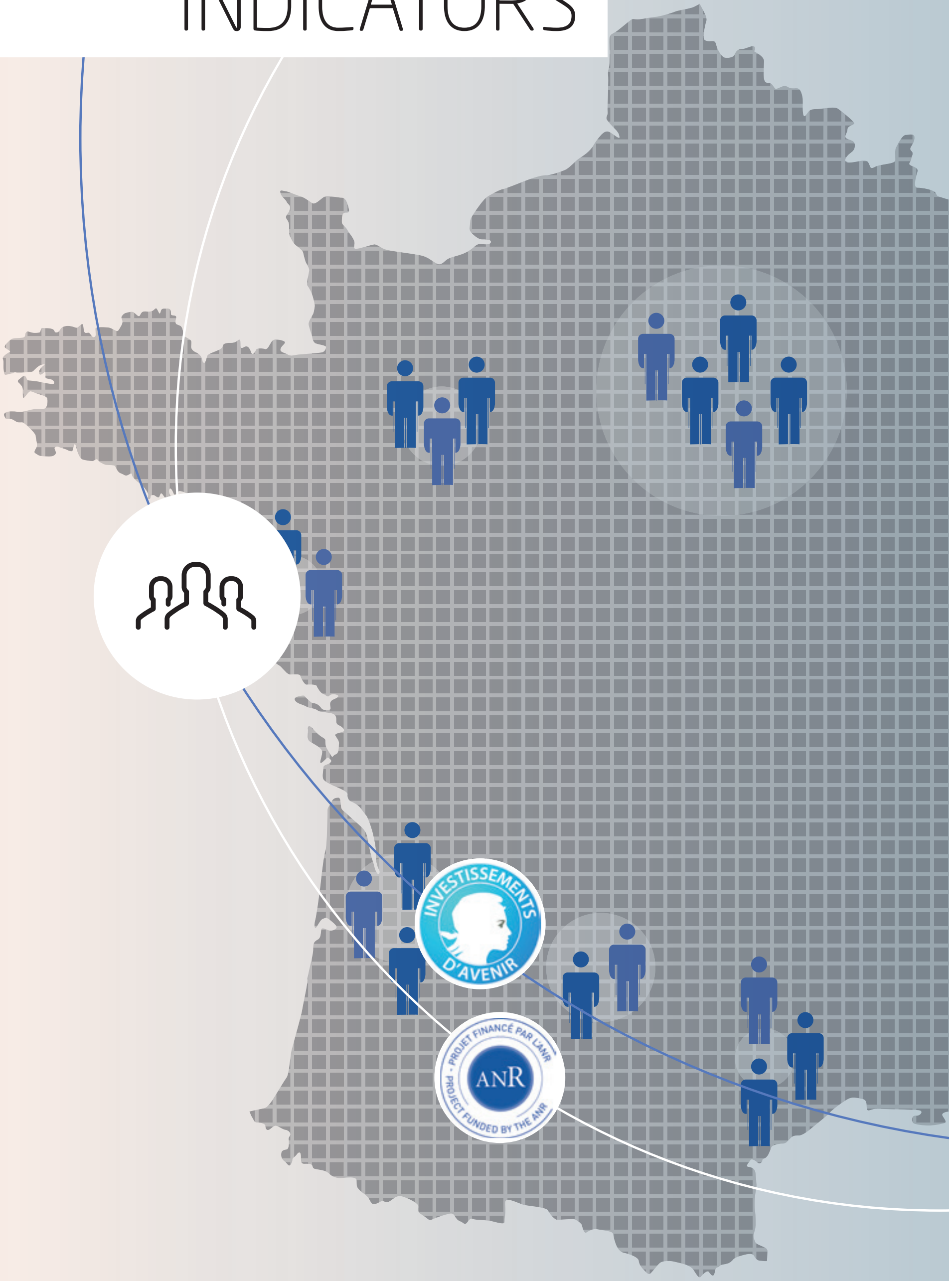
ISfinder provides a service either by assistance with the annotation (advice, sequence checking) or by the annotation of genomes for French or foreign units/teams.
In addition, the site provides specific tools : Issaga, a semi-automatic tool for genomic IS annotation, it also allows the visualization of complete and partial IS in expertly annotated prokaryotic genomes.

SELECTED PUBLICATIONS

❶ Quentin Y., Siguier P., Chandler M., Fichant G. Single-strand DNA processing: phylogenomics and sequence diversity of a superfamily of potential prokaryotic HuH endonucleases. BMC Genomics. 2018 Jun 19;19(1):475

❷ Siguier P., Gourbeyre E., Chandler M. Known knowns, known unknowns and unknown unknowns in prokaryotic transposition. Curr Opin Microbiol. 2017 Aug;38:171-180.

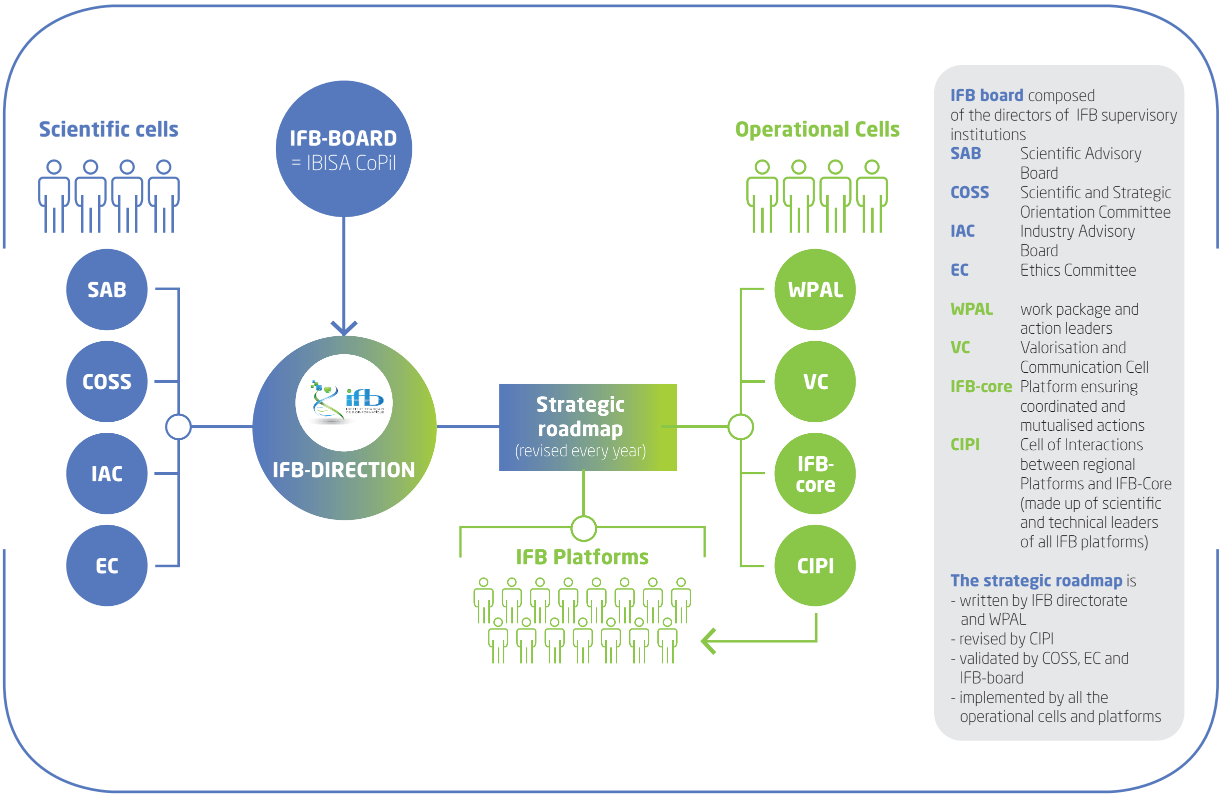
GOVERNANCE AND GENERAL INDICATORS



IFB governance

🌐 **https://www.france-bioinformatique.fr/en/about-us/governance**

The legal status of the IFB is a mixt service unit (UMS) depending on five supervisory authorities (CNRS, INRA, INSERM, CEA and INRIA). The organisation is based on a continuous interaction with the supervisory authorities, the other '-omics' national infrastructures and the major national plans.
It separates the scientific strategy from the operational tasks.



IFB board composed of the directors of IFB supervisory institutions

SAB	Scientific Advisory Board
COSS	Scientific and Strategic Orientation Committee
IAC	Industry Advisory Board
EC	Ethics Committee
WPAL	work package and action leaders
VC	Valorisation and Communication Cell
IFB-core	Platform ensuring coordinated and mutualised actions
CIPI	Cell of Interactions between regional Platforms and IFB-Core (made up of scientific and technical leaders of all IFB platforms)

The strategic roadmap is
- written by IFB directorate and WPAL
- revised by CIPI
- validated by COSS, EC and IFB-board
- implemented by all the operational cells and platforms

Figure 13 Governance and operational organisation of the IFB. Update of the structure proposed in the 2018-2021 workplan.

IFB GOVERNANCE AND OPERATIONAL ORGANISATION

> IFB directorate (collège de direction)

The **IFB directorate (collège de direction)** is composed of the IFB director, the deputy director and three other IFB members; it ensures the management of the overall project and proposes the scientific orientations. Members are nominated by the IFB directors, approved by the CIPI and appointed by the board of directors, for a duration of three years. Meetings of this directorate are organised every two weeks. The directorate is guided by the IFB scientific cells, which consists of four advisory bodies (SAB, COSS, IAC, EC), and is helped by four operational cells in charge of managing the actions of the strategic roadmap (IFB-core, WPAL, CIPI and VS).

> IFB board of directors

The **IFB board of directors** is a subset of the IBISA board of direction, restricted to the representatives of the IFB supervisory authorities (CNRS, INRA, Inserm, CEA and INRIA), the Ministry of Research and the Agence Nationale de la Recherche. It approves IFB's scientific programme and budget, appoints and dismisses the IFB Directors, and appoints the members of the IFB Scientific Advisory Board (SAB). This Board can establish additional advisory bodies when necessary.

> Scientific Advisory Board (SAB)

The **SAB** reviews IFB actions and advises the Board of Directors on IFB's scientific strategy. It meets once a year. Its members are appointed by the Board of Directors.

> Committee in charge of the Strategy and Science Orientation (COSS)

Committee in charge of the Strategy and Science Orientation (COSS), made up of Scientific representatives from the supervisory authorities, a representative of the ministry, the heads of national infrastructures (France Génomique, ProFi, MetaboHUB, Frisbi, France BioImaging, France Life Imaging, Biobanks, EMBRC-FR, Constance) and national scientific programmes (e.g. France Medicine Genomic 2025 Plan). This committee defines the main strategic orientations with regard to the Scientific Advisory Board report and the associated required budget. It interacts regularly with the heads of the operational cells, supervises the different actions and discusses new orientations as often as necessary.

> Ethics Committee

The **Ethics Committee (EC)** assists the IFB Directorate on legal, ethical and societal issues related to the processing of any activity in the missions of IFB (e.g. protection of sensitive data, ethical questions related to personal genomics, ...). It is made up of scientists, ethicists, jurists and people from any discipline whose expertise may contribute its missions. Its activities include the evaluation of ethical issues, the organisation of intra-IFB and public debates, as well as communication on ethical questions.

> Industry Advisory Committee

The **Industry Advisory Committee (IAC)** provides IFB with high-level, strategic advice and input from industry. It includes representatives of industries, of the IFB directorate and of user communities, as well as the co-heads of the workgroup "IFB – Industry relationships". Members of this committee have skills and experience in the diverse areas covered by IFB including health, agronomy, environment, biotechnology and so forth. They also have experience in provision of bioinformatics services and software publishing.

> Work Package and Action Leaders committee (WPAL)

The **Work Package & Action Leaders (WPAL)** is composed of the Directorate and the heads of work packages and actions defined in the IFB strategic roadmap. It regularly takes stock of the progresses of the actions and proposes adapted measures to meet the recommendations of the advisory bodies (COSS, SAB, IAC and EC). It contributes to the IFB reports by providing key elements to the Directorate. It meets at least once a month.

> CIPI : Cell of Interactions between regional Platforms and IFB-Core

The **CIPI** regroups the heads of bioinformatics platforms and the IFB Directorate. Its function is to promote communication, ensure follow-up of the services, and foster collaboration between the platforms involved in IFB actions (NNCR, task force, services to life science communities, help and orientation desks, tool developments, interoperability, compliance with standards, etc.). Meetings are organised every two months (face-to-face or teleconferences).

> IFB-Core

IFB-core is the IFB platform ensuring the coordination of the national infrastructure IFB, and the management of the mutualised resources. Its personnel includes the IFB directors, permanent staff assigned to IFB by the partner research organisms, as well as the temporary staff recruited to carry out the actions of the work plan. IFB core personnel is hosted by the IFB platform that best suits their missions. As the French node of ELIXIR, IFB core ensures the link between all the IFB platforms, the ELIXIR hub and the other nodes.

> The Valorisation and Communication cell (VC)

The **Valorisation and Communication cell (VC)** is an operational cell ensures the dissemination of IFB activities, realisations, projects and services towards users and stakeholders, via diverse media (Internet, printed reports, press). It also facilitates the engagement of new partners and new user communities. In addition, it catalyses the transfer of knowledge and the contractualisation of partnerships between IFB and industry. It is made up of the heads of the concerned actions (in particular, communication, valorisation and IFB-industry relationships), two members of the directorate, and delegates from the knowledge transfer departments of IFB supervisory institutions. The list of members of the VC cell will be established in 2019.

LIST OF GOVERNING BODIES

IFB directorate

Claudine Médigue *(CNRS, Evry)*
Jacques van Helden *(AMU+CNRS, Marseille)*
Christine Gaspin *(INRA, Toulouse)*
Ivan Moszer *(ICM, Paris)*
Vincent Lefort *(CNRS, Montpellier)*

Board of directors

Gilles Aumont *(INRA)*
Vincent Berger *(CEA)*
Hugues Berry *(INRIA)*
Gilles Bloch *(Inserm)*
Daniel Boujard *(CNRS)*
Christine Cherbut *(INRA)*
Alix De La Coste *(CEA)*
Claire Giry *(Inserm)*
André Le Bivic *(CNRS)*
Laure Sabatier *(CEA)*
Bruno Sportisse *(INRIA)*

Committee in charge of Scientific and Strategic Orientation (COSS)

Hugues Berry *(INRIA)*
Jacques BITTOUN *(France Life Imaging)*
Frédéric Boccard *(CNRS)*
Christophe Calvin *(CEA)*
Jérôme Garin *(ProFi)*
Yann Hérault *(Phenomin/Celphedia)*
Bruno Klaholz *(Frisbi)*
Pierre Le Ber *(France Génomique)*
Jacques Le Gouis *(Phenome/Emphasis)*
Franck Lethimonnier *(FMG2025)*
Hervé Monod *(INRA)*
Catherine Nguyen *(INSERM)*
Nicolas Pade *(EMBRC-FR)*
Dominique Rolin *(MetaboHUB)*
Jean Salamero *(France BioImaging)*
Marie Zins *(Constance)*

Scientific Advisory Board (SAB)

Amos Bairoch *(Swiss Institute of Bioinformatics, Switzerland)*
Lodewyk Wessels *(Cancer Systems Biology Center, The Netherlands)*
Christine Orengo *(University College London, UK)*
Anton Nekrutenko *(PennState University, USA)*
Soren Brunak *(University of Copenhagen, Denmark)*
Ana Conesa Cegarra *(University of Florida, USA)*

Ethical Committee

Gaëlle Bujan *(CNRS)*
Anne Cambon-Thomsen *(CNRS)*
François Eisinger *(Inserm)*
Nathalie Gandon *(INRA)*
Christine Gaspin *(IFB + INRA)*
Marc Hanauer *(Gustave Roussy)*
Claudine Médigue *(IFB)*
Catherine Nguyen *(Aviesan)*
Jacques Nicolas *(INRIA)*
Ana Rath *(Orphanet)*
Alice René *(CNRS / INSB)*
Jacques van Helden *(IFB)*
Alain Viari *(INRIA)*

Industry Advisory Committee (IAC)

Guillaume Boussy *(BIOASTER)*
Frédéric Sapet *(BIOGEMMA)*
Bruno Lacroix *(Biomérieux)*
Jérôme Wojcik *(QuartzBio)*
Anne Sophie Coquel *(SANOFI)*
Charles Loomis *(Sixsq)*
Patrick Durand *(IFB, Ifremer)*
Christine Gaspin *(IFB, INRA)*
Claudine Médigue *(IFB, CNRS)*
Jacques van Helden *(IFB, AMU, CNRS)*
Yves Vandenbrouck *(IFB, CEA)*

Work Package and Action Leaders committee (WPAL)

The names of the WPAL members (Work Package and Action leaders) are provided page 55.

Cell of Interactions between regional Platforms and IFB-core (CIPI)

The names of the CIPI members (heads of IFB platforms) is provided in the section “IFB platforms” (p37-50).

WORK PACKAGE LEADERS AND IFB-CORE STAFF IN 2018

2017-2018 workplan

WP1. National network of services in bioinformatics (NNCR)
Christophe Blanchet[1] & Jacques van Helden[2]

Physical infrastructure: compute and storage
Julien Seiler[3] & Christophe Blanchet[1]

Software and data environment
Gildas Le Corguillé[4] & Olivier Collin[5]

Support to Databases
Claudine Médigue[6] & Guy Perrière[7]

Catalogue of French resources in bioinformatics
Jacques van Helden[2] & Hervé Ménager[8]

Access to end-users
Christine Gaspin[9] & Erwan Corre[10]

Shared services with other national infra
Christophe Bruley[11] & Jean-François Dufayard[12]

Consulting and orientation desk
Vincent Lefort[13] & Ivan Moszer[14]

WP2. Innovation: towards integrative bioinformatics
Claudine Médigue[6] & Olivier Sand[15]

Pilot projects in integrative bioinformatics
Claudine Médigue[6] & Etienne Thévenot[16]
& Olivier Sand[15]

Call for challenges and unlocking of technological and scientific barriers
Ivan Moszer[14] & Jacques van Helden[2]

Catalysing interoperability and integration between resources
Alban Gaignard[17] & Marie-Dominique Devignes[18]

WP3. Links with international networks and industry
Anne-Françoise Adam-Blondon[19]
& Claudine Médigue[6]

IFB as the French node of the European ELIXIR network
Anne-Françoise Adam-Blondon[19]
& Claudine Médigue[6]

Partnership with industries
Patrick Durand[20] & Yves Vandenbrouck[21]
& Victoria Dominguez Del Angel[22]

WP4. Training and outreach
Jacques van Helden[2] & Morgane Thomas-Chollier[23]

Training
Jacques van Helden[2] & Hélène Chiapello[24]

Join actions with SFBI and GDR BIM
Morgane Thomas-Chollier[23] & Hélène Touzet[25]

Communication & Valorisation
Claudine Médigue[6] & Victoria Dominguez Del Angel[22]

WP5. Management and governance
Claudine Médigue[6] & Jacques van Helden[2]

Governance and coordination bodies
Claudine Médigue[6] & Jacques van Helden[2]

Quality management system
Claudine Médigue[6]

IFB Economic model
Christine Gaspin[9]

Communities

Bioinformatics for health
David Salgado[26] & Ivan Moszer[14]

IFB-core staff 2018
Laurent Bouri (Catalogue + Web)[27]
Mélanie Buy (Pilot-projects)[28]
Etienne Camenen (Pilot-projects)[29]
Nicole Charrière (NNCR)[30]
Victoria Dominguez Del Angel
(ELIXIR, Industry, Comm.)[22]
Maxime Folschette (Pilot-projects)[31]
Jonathan Lorenzo (NNCR)[32]
Olivier MERZEAU (Communication)[33]
Sylvain Milanesi (Consulting and orientation desk)[34]
Guillaume Postic (Pilot-projects)[35]
Valentin Saint-Léger (Admin & management)[36]

_INDICATORS OF ACTIVITIES FOR 2017-2018

Answer to user needs

	2017	2018
Number of supported research projects	517	872
Average time to open new accounts days	1.5	1.8
Number of new user accounts created in the year	3 043	2 793
Number of hosted Unix accounts (storage and compute)	11 523	-
Number of unique users for Orphanet	9 645 238	12 057 375
Number of unique users for IFB Web tools	-	331 230
Downloads of tools developed by IFB platforms	303 552	183 740
Downloads of tools developed by IFB platforms without Orphanet	-	19 616
Training events organised by IFB platforms	81	153
Participants to IFB training events	1100	2073

User satisfaction surveys

88%

98%

Percentage of satisfied
+ very satisfied users for
delivered service on projects

-

95.3%

Percentage of satisfied
+ very satisfied users for
followed training

Collaborative projects and service delivery

245

388

Number of internal projects
(projects for the host laboratory
of the IFB platform)

261

229

Number of international
academic projects (involving
foreign academic labs)

520 K€

1 810 K€

Amount of money collected
by IFB platforms on EU
and international projects

2017

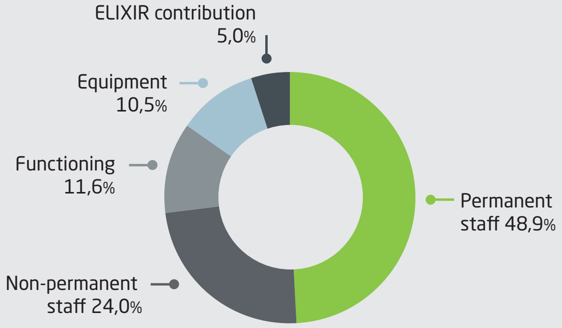
2018

FINANCIAL DATA FROM THE IFB NETWORK (31 PLATFORMS)

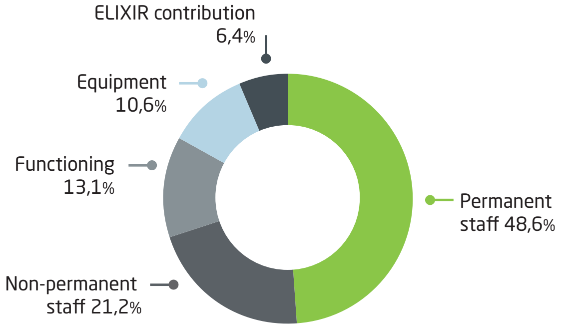
The 2017 numbers result from a full-cost analysis enabling to estimate the total cost of the national infrastructure as a whole, by collecting all types of financial income and expenses from all the IFB platforms, including direct as well as indirect costs (environment offered by the hosting institutions).

Full-cost analysis enables to estimate the real cost of the service offered to the community. In 2018 the full-cost analysis has not been led, and the income has been gathered from platform-provided indicators. Note that the income show yearly fluctuations according to new contracts (e.g. a 3 M€ EU project for Orphanet in 2017) but that the corresponding expenses are smoothed over several years.

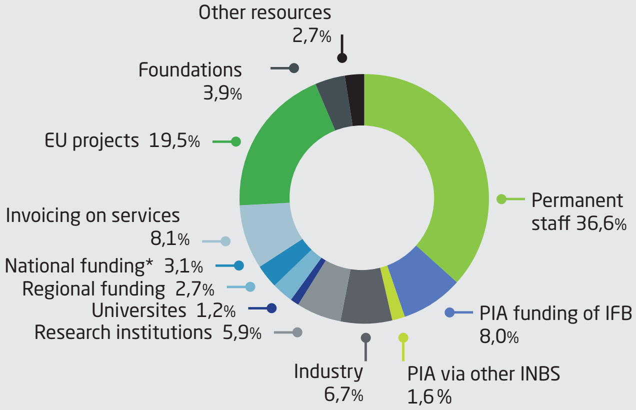
2017 expenses for all IFB platforms (total : 16,8 M€)



2018 expenses for all IFB platforms (total : 17,8 M€)

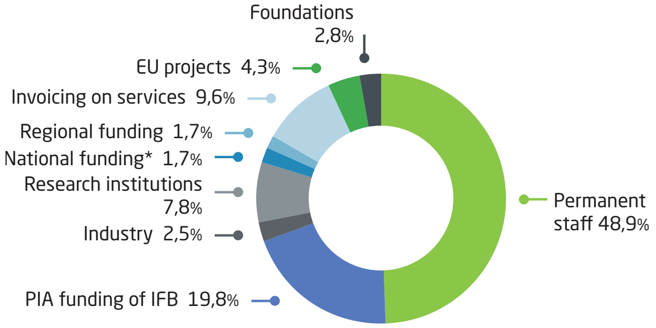


2017 income for all IFB platforms (total : 22,4 M€)



* *Except PIA*

2018 income for all IFB platforms (total : 17,8 M€)

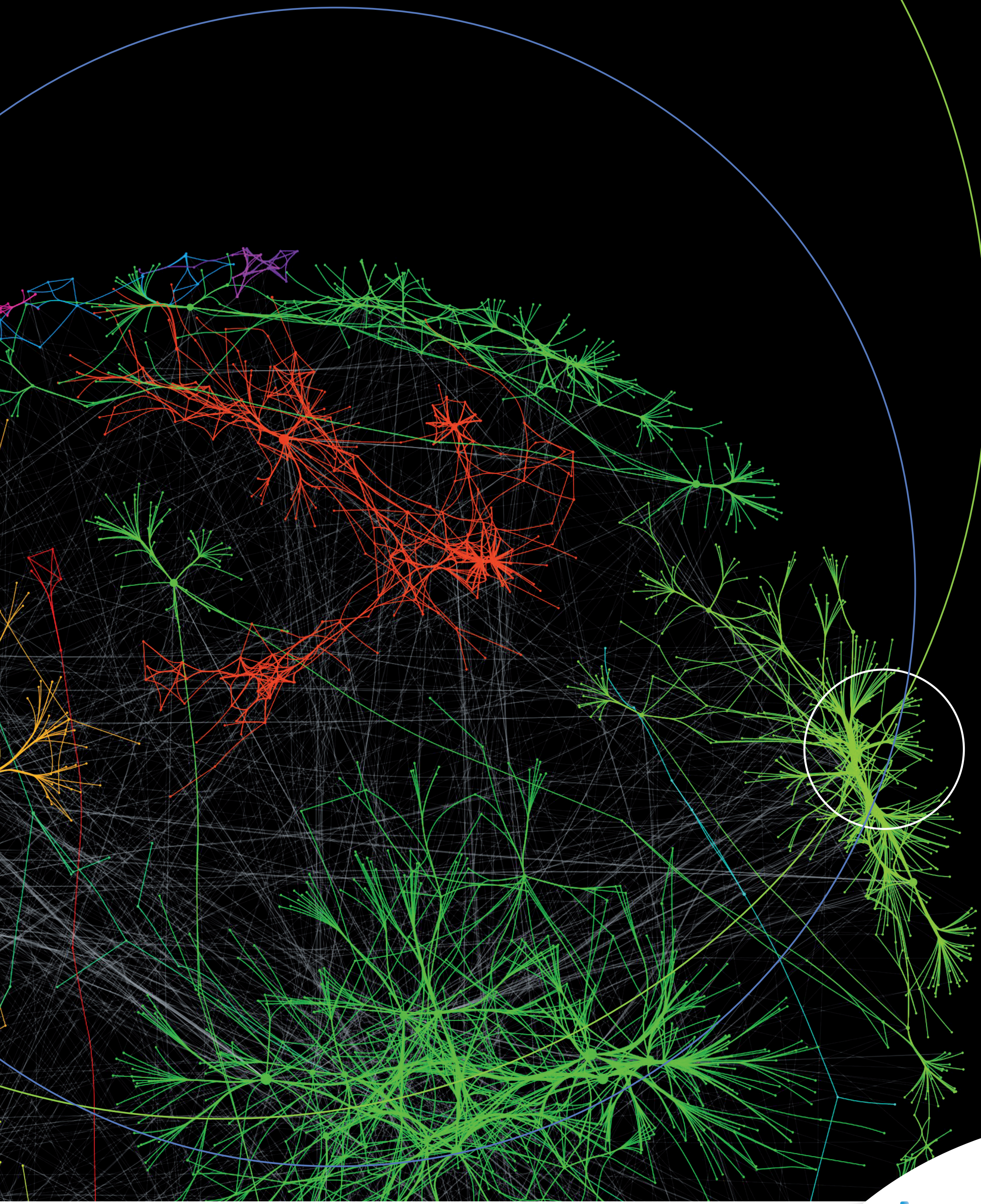


TRIBUTE TO CHRISTOPHE CARON

At the time of finalizing this IFB activity report, we would like to address special thoughts to our late colleague and friend Christophe Caron, whose investment has deeply marked the Institut Français de Bioinformatique.

Since the beginning of the IFB, Christophe was a pioneer of several projects (Galaxy community, development of resources in connection with MetaboHub, MOOC for metabolomics) and became a driving force of the 2017-2018 restructuration. He was deeply involved in the conception of the 2018-2021 work plan, and proposed the collaborative decentralisation model underlying the National Network of Computing Resources for bioinformatics. He conceived and established the mutualised task forces, took the co-leadership of the WP on bioinformatics services, and became member of the IFB directorate. Without him, the work presented in this report would not have been possible.

CNRS UMS 3601
IFB-Core, Génoscope
2 rue Gaston Crémieux,
91057 Evry Cedex
contact@groupes.france-bioinformatique.fr



www.france-bioinformatique.fr