French node:

EGI : France Grilles – ELIXIR : IFB

By Tiphaine Martin, CNRS-University of Cambridge
Information provided by Thierry Meinnel, French observer for Elixir; Claudine Médigue, Coordinator of ReNaBi, Jean-Françcois Gibrat, Director of IFB, Genevieve Romier : User Coordinator of France Grilles and Vincent Breton, Director of France Grilles

France Grilles

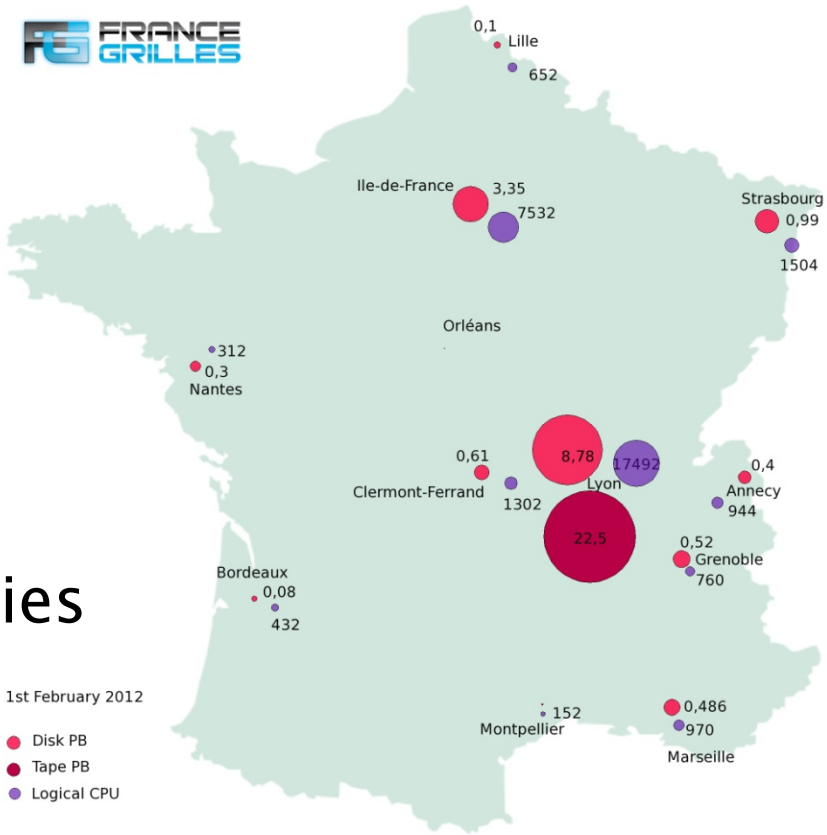
a multidisciplinary national Distributed Computing Infrastructure, http://www.france-grilles.fr

- Share IT resources (grid/cloud-computing) and expertise for science
- Increase joint actions with the users from the networks of supercomputers and research grid
- Support community involvement in e-infrastructures
- Identify, document and publicize the services
  - The Workload Management System (Dirac), data grid software solution (IRODS), cloud-computing (Stratuslab),..
- Develop training and user support

infrastructure: resources

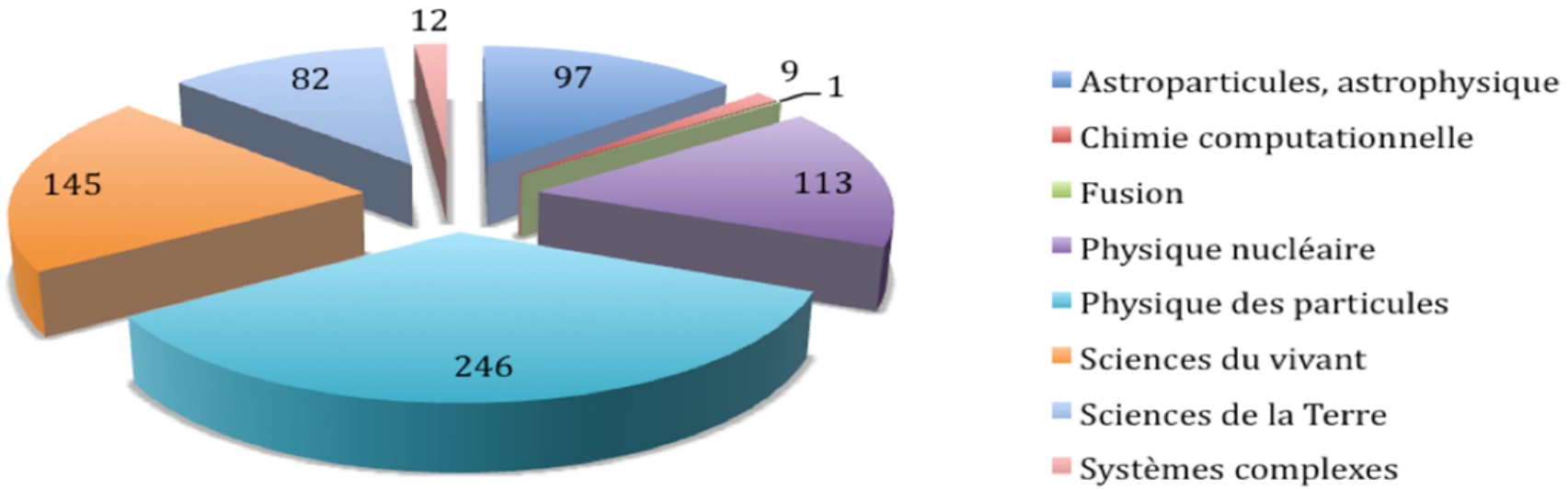
- 18 production sites
- 31 860 logical CPUs
- 280 316 HEPSPEC06
- 15,7 PB disk storage
- 22,5 PB tape storage

Resources shared by communities



France Grilles user communities

- about 700 users from many disciplines
- 89 Virtual Organizations work on France Grilles infrastructure



Users distribution per discipline

French Elixir node : IFB

IFB
www.renabi.fr

- Based on ReNaBi (French Bioinformatics Platforms Network), created in 2004
  - In 2009 : 13 platforms certified by the GIS-IBiSA national organization
  - Now : structured in 6 regional centres
  - ISO 9001: 2000 certification
- Creation in 2013 a legal entity UMS IFB (French National Bioinformatics Institute)
  - Supported by the IBISA consortium (Infrastructures in Biology, Health and Agronomy)
  - Not yet an official member of ELIXIR, should be in 2013

French Elixir node : IFB

- IFB : star network:
  - a **national node** (IFB-core), having its own head, staff and IT infrastructure (10pers., 10M€ +1.5M€/year between 2012-2020)
    - the unique entry point for requests of services from the biological community
    - structuring and following-up projects
    - developing training services
    - promoting European coordination
  - **existing regional PFs,** thematic poles characterised by their international visibility and/or biological data specificity. (about 100 Full-Time Equivalents (FTEs) in terms of permanent staff, and 57 FTEs for people hired on fixed-term contracts)
    - physical infrastructure deployment
    - dissemination of bioinformatics methods and tools
    - support of biologist projects

Main goals of IFB

- **Improve the coordination** of all the components **of the national bioinformatics infrastructure**
  - Ensure know-how transfer between platforms
  - Sharing of knowledge and developments
  - Promotion of interoperability, and standard dissemination
  - Optimize and rationalize usage of computing facilities at the national scale
  - Prevent duplication of efforts
- facilitate the coordination between ELIXIR-ReNaBi and national components of other ESFRI infrastructures
- build up **new components** in collaboration with other European countries
  - resources in metagenomics, chemoinformatics, RNomics,...

Offered services

- Disseminate up-to-date version of db and tools
- Develop and maintain original **data resources**

immunogenetics

carboxydrates

microbes

plants

- Give **Computing and storage capacity**
- Standard access to DBs and **integration of tools** (pipeline available via Web servers).
  - => **MobyleNet** project & **phylogeny.fr** server
- User support for databases, tools & projects
- Teaching and **training** in bioinformatics

4 types of ELIXIR services

Expertise of IFB

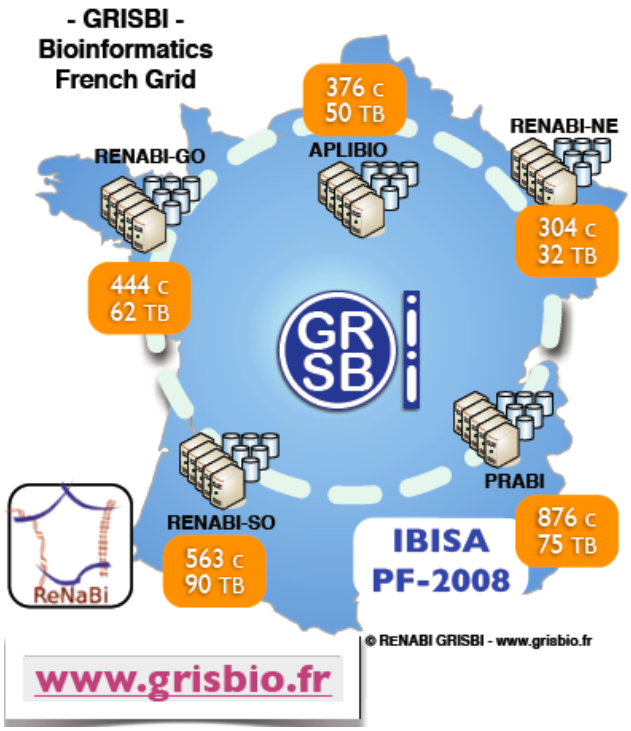
- Domains with available expertise :
  - Plant and Animal genomics, Viral and Microbial genomics, Phylogeny, Non coding RNA, medical genomics
- Resources having a strong international visibility
  - such as IMGT, CAzY.
- Link with national research teams
- Link with data generators (NGS, medical, proteomics, ….)

- Two mains R&D directions :
  - **processing of high-throughput** (« omics ») **data**
  - development of the **computer infrastructure for service**

Computer infrastructure

- Currently, geographically distributed in different platforms
- From 2008, creation of GRISBI infrastructure (6 platforms, http://www.grisbio.fr/)
  - to strengthen the RENABI bioinformatics experts team
  - to federate the computing resources of the French bioinformatics platforms in a distributed Research Infrastructure devoted to serving the Bioinformatics community at the national level
  - supported by the RENABI Network, the IBISA consortium and the France Grilles (French NGI, www.france-grilles.fr)
- 2013-2020 : Creation of a computer/storage infrastructure to answer the needs of IFB

GRISBI infrastructure



- Answer to biology needs
- Pool existing infrastructures
  - Based on EGI technology (glite)
  - Develop procedures for implementation in other centres
  - Renabi (CE gateways), FranceGrille (vo.renabi.fr)
- Adapt uses for biology community
  - Command GR*, biomaj, xtreemfs
- Training/conference
  - Ecole CNRS 2010, tutorial 2011, …
- Study of cloud computing
  - Bring calculations to data : bioinformatic « appliances »
  - Collaboration with StratusLab project
- Using/Use cases :
  - Comparative genomics (S.Penel, http://france-grilles-2011.sciencesconf.org/1133)
  - NGS (T.Martin, http://france-grilles-2011.sciencesconf.org/1132)
  - Computing of NMR structure with ARIA (F, Mareuil)

Debriefing and expectation on the using of distributed infrastructure

- Remarks :
  - Need to have a good network between nodes (bandwidth 1Gbps, or more 10Gbps ; mainly to transfer data)
  - Need to have the computing nodes close to databases and data e.g. for NGS
  - Need to have a lot of memory for some CE and a large capacity for SE
  - Need to have to access easy and simply to different CPU, memory and storage (not to fill in a project in order to access a number of CPU time like for supercomputers)
  - Need to have to access easy and simply to tools and databases
- Expectation :
  - Launch jobs by web service or/and web interface, not by command line for biologists (e.g. galaxie, DIRAC)
  - Have a workflow management system easy to use it
  - Be able to define the size of desired memory (e.g. tools of assembly or mapping)
  - Have a good data management system (e.g. IRODS)
  - Be able to launch a job/workflow on grid, cloud-computing, cluster, supercomputing according the type of data/tools (link with PRACE)

Support required from ELIXIR hub

- **Coordination of**
  - national components of the distributed infrastructure
  - cloud / grid computing facilities
- **Identification, development and dissemination** of resources of interest
  - enhance the mobilization of resources
  - set up top-level European courses, organize workshops
  - promote long term support to databases and services
- **Organization of resources distribution**
  - => Standards to facilitate exchanges and collaborative developments between nodes