StratusLab, Bioinformatics and the Cloud

Christophe Blanchet

IDB - Infrastructure Distributing Biology

IBCP CNRS FR3302 - LYON - FRANCE

christophe.blanchet@ibcp.fr

Bioinformatics Today

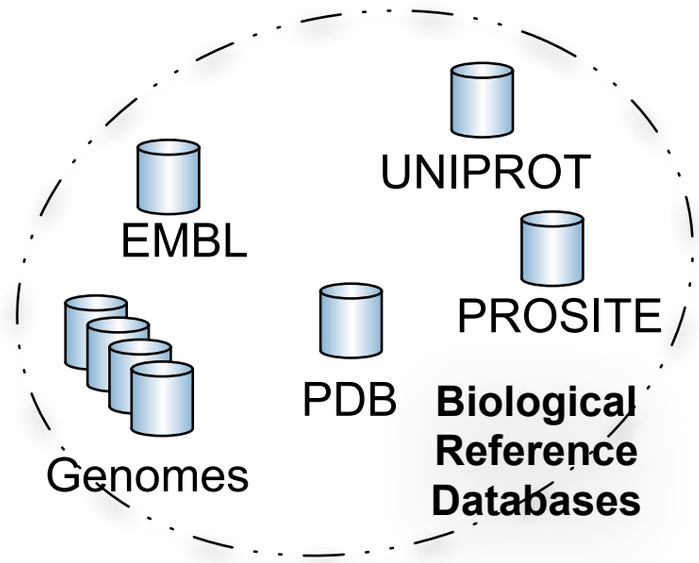
- Size of biological date are tremendous
  - Institut Sanger, UK, 5 PB
  - BGI (Beijing Genome Institute), 4 sites, 10 PB

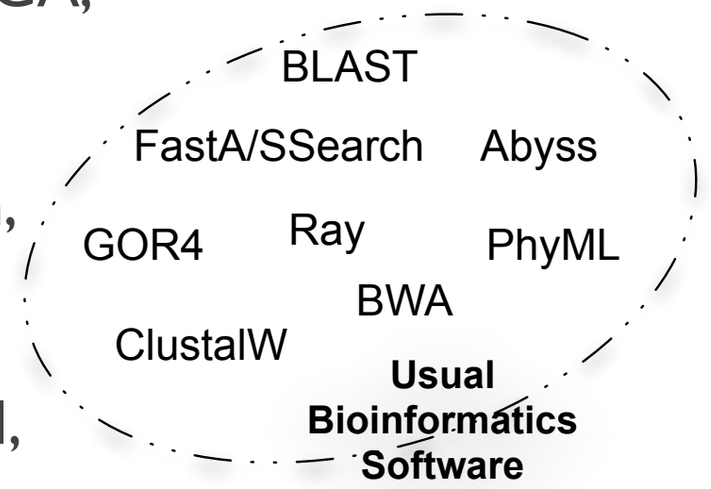
  ➡ **Huge data in lot of places**
- How to analyse these data
  - Scale-up of the analyses : from gene/protein to complete genome/proteome; from on metabolic way to systems biology ...
  - Need of more and more computing resources
  - Usual interfaces: portals, Web services, federation,...

  ➡ **Require datacenter with ease of access/use**

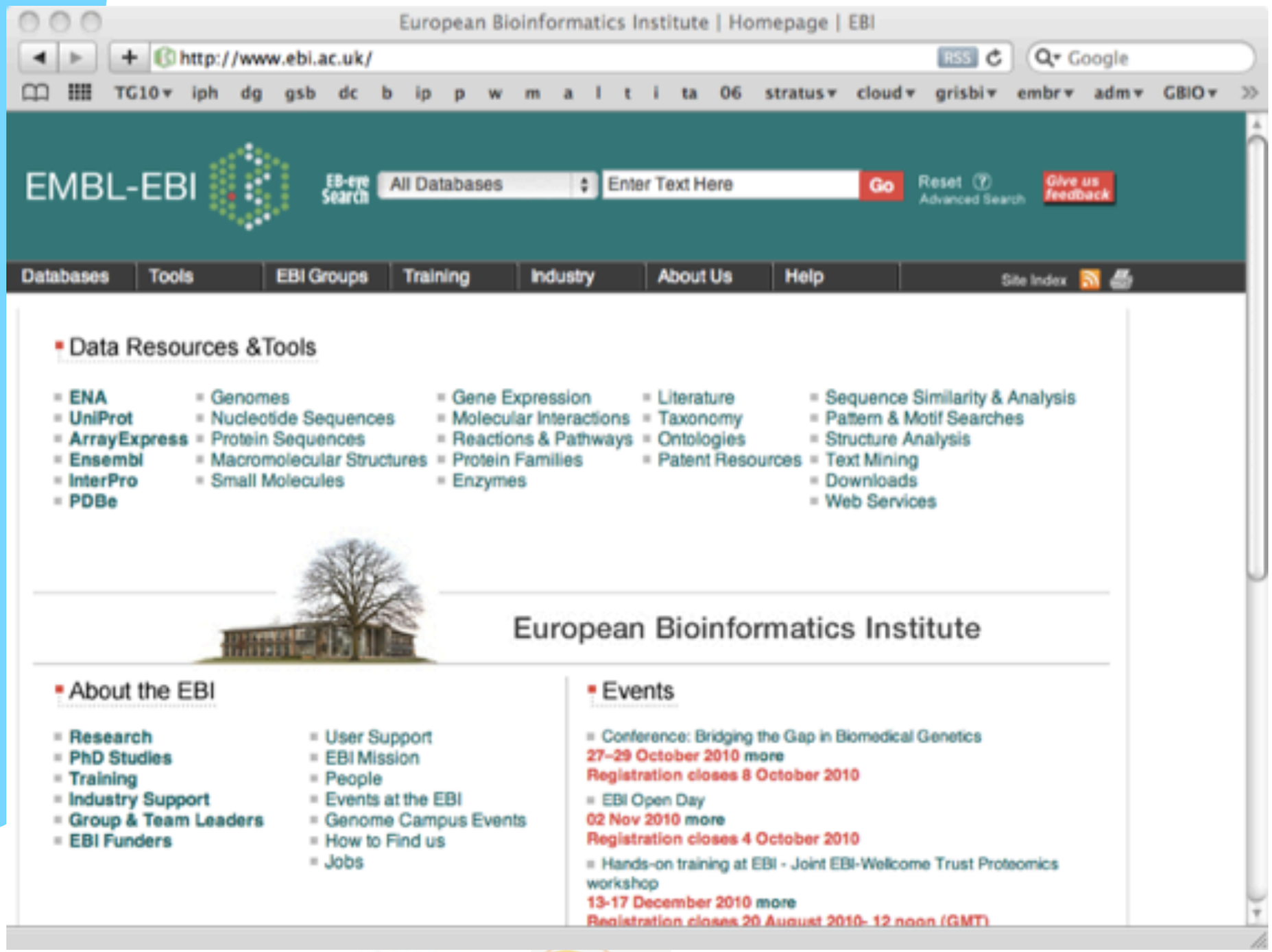
Biological data & Bioinformatics Tools

- 1330 different reference data sources
  - M.Y. Galperin & G.R. Cochrane, NAR 2011
  - UniProt, Génolevures*, Base, AcNuc (**), GenBank, EMBL, PRODOM*, Ensembl, Hogenom*, Homolens*, PDB, Génomes Complets, TransFac, Nr, SRS (**), SUMO(*), PROSITE, ABC, KEGG, ...
- Thousands of different daily-used tools
  - InterPro, pFam, Genmark, Genezilla, Pred. Intron*, Sys. Biology*, Réseaux Méta*, Ancêtres (hiador, MGR), Autodock, Docking@Grid *, Base (stats), Pase* (Base), ASCQ_me*, R, MGA, Mauve, MathLab, Scilab, Show*, R'mes*, EMBOSS, Gromacs, ClustalW, Maft, MAST, MEME, Phred/Phrap, BLAST, FASTA, SSearch, MUSCLE, PhyML, Dialign, multalin, RepeatMasker, Amber, NAMD, JUMNA*, ADAPT*, MaxDo*, Curves*, Prophet*, DALI, SUMO(*), PattInProt*, ...

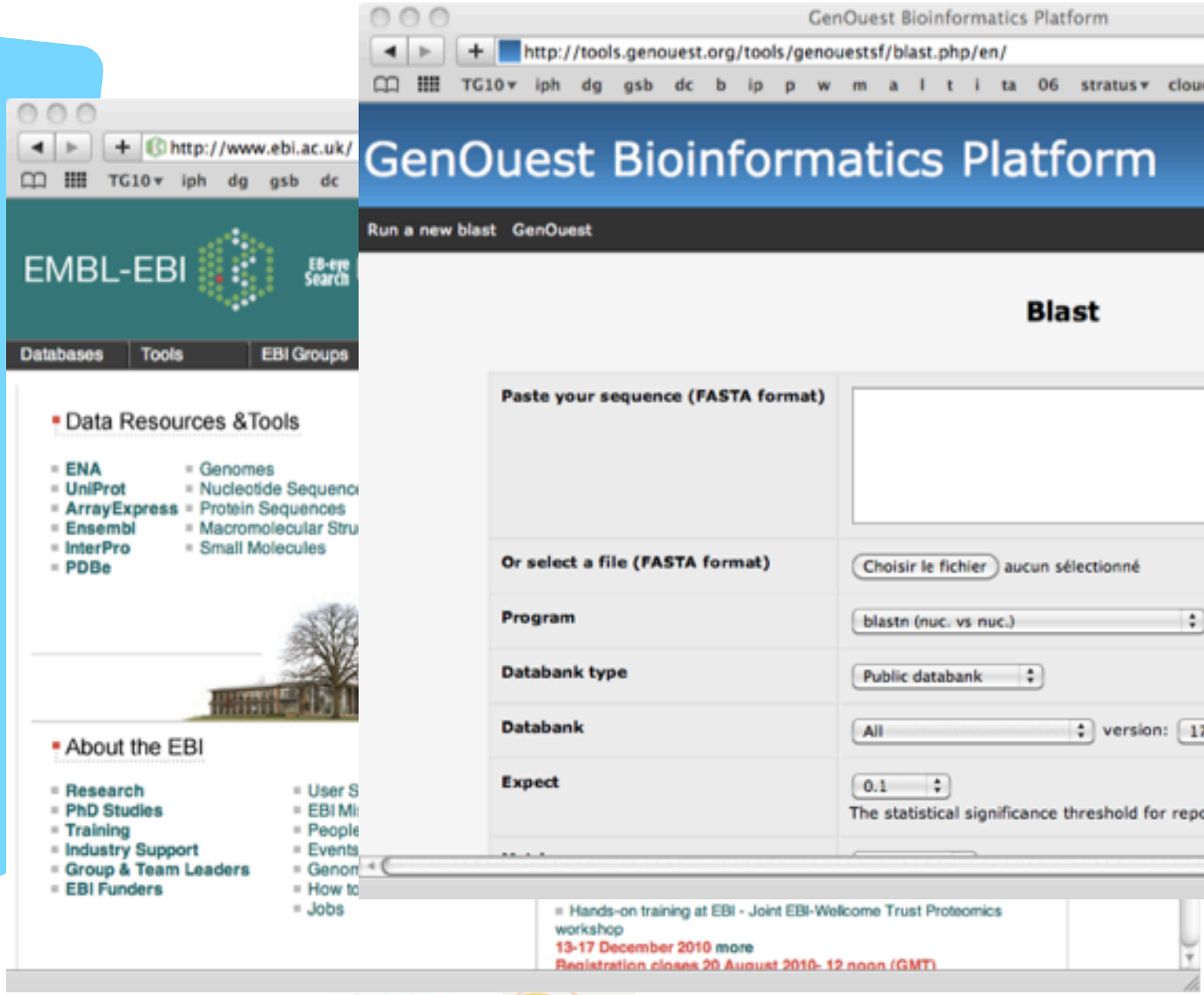




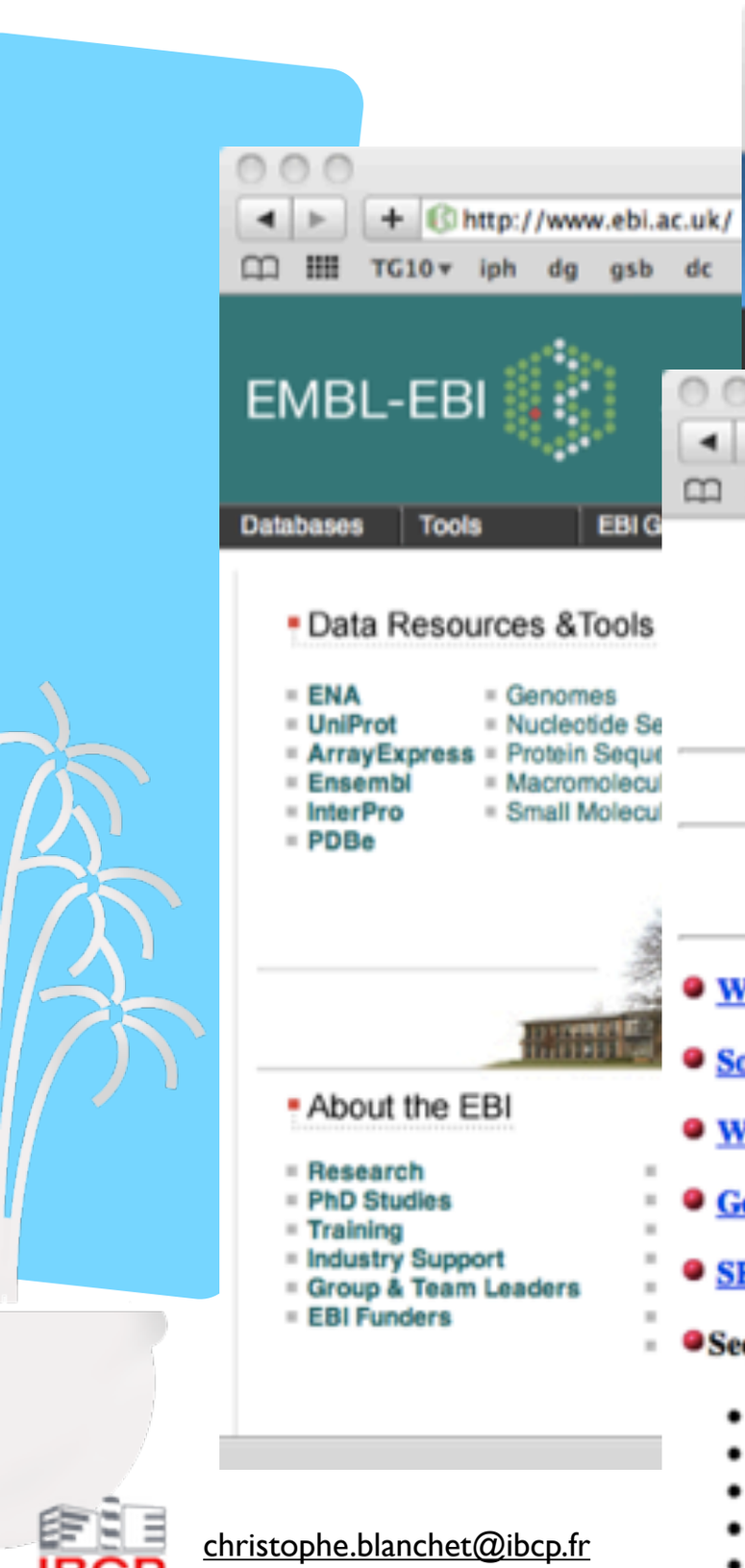
Scientific Gateways

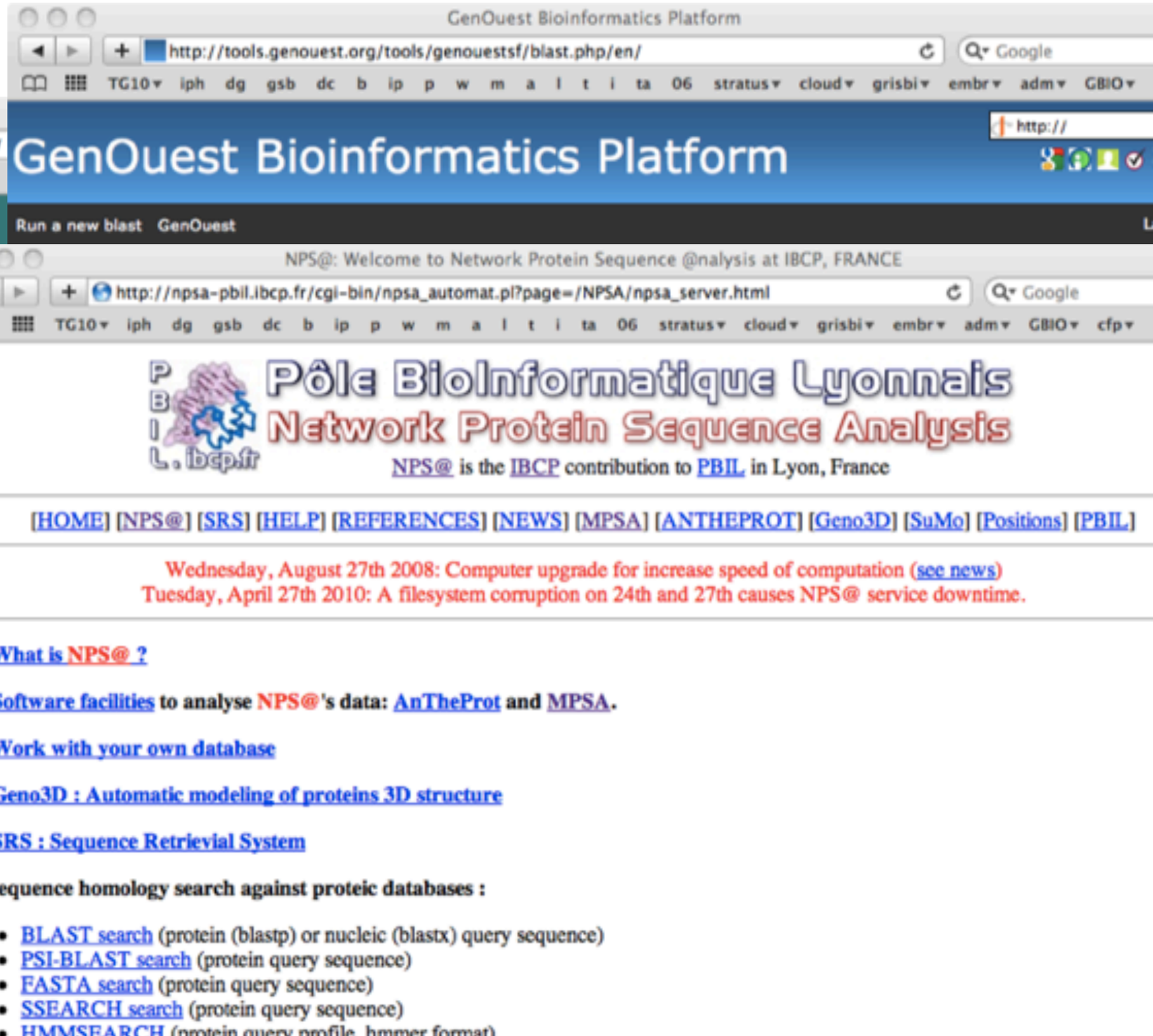


Scientific Gateways



Scientific Gateways





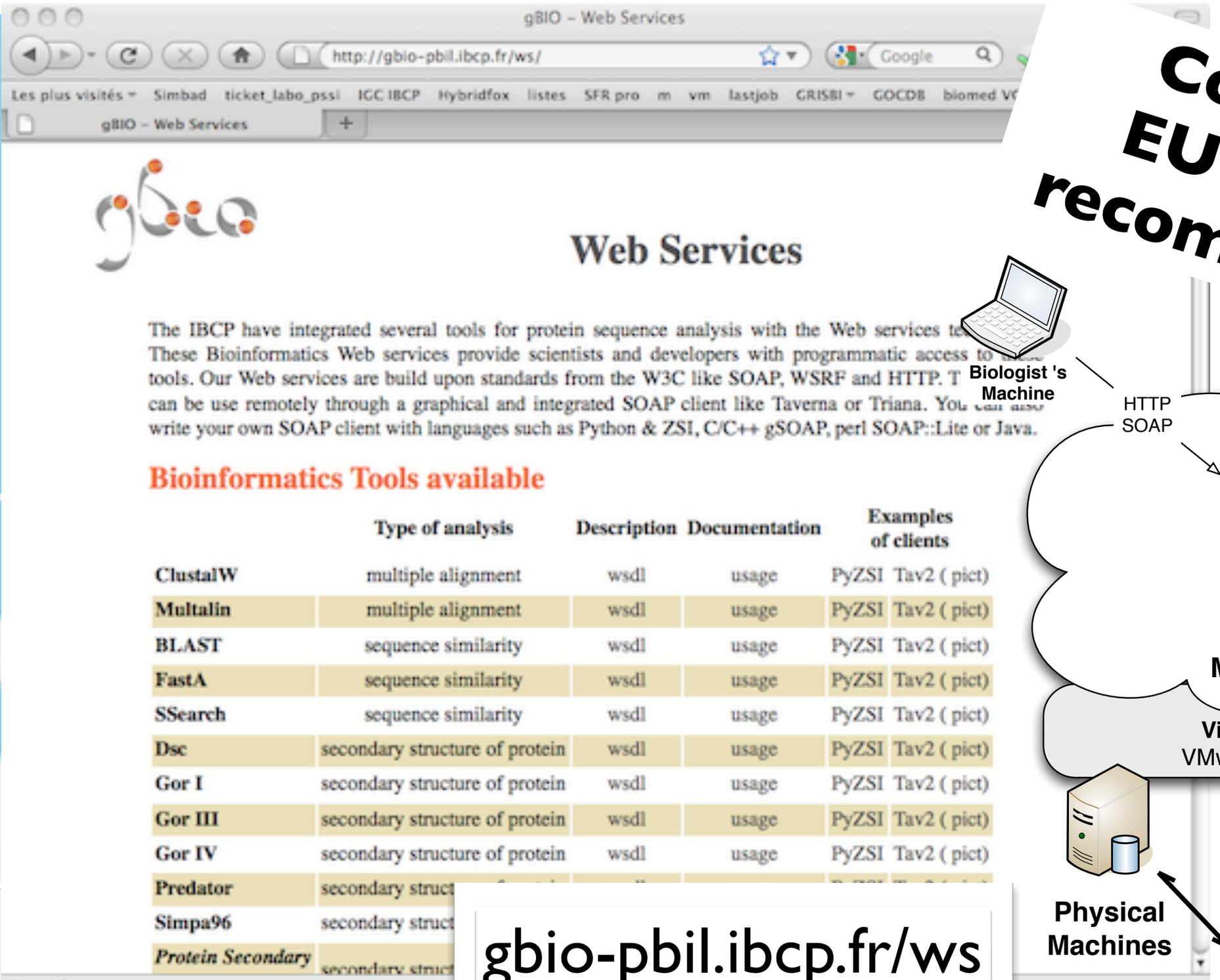
christophe.blanchet@ibcp.fr

Scientific Gateways

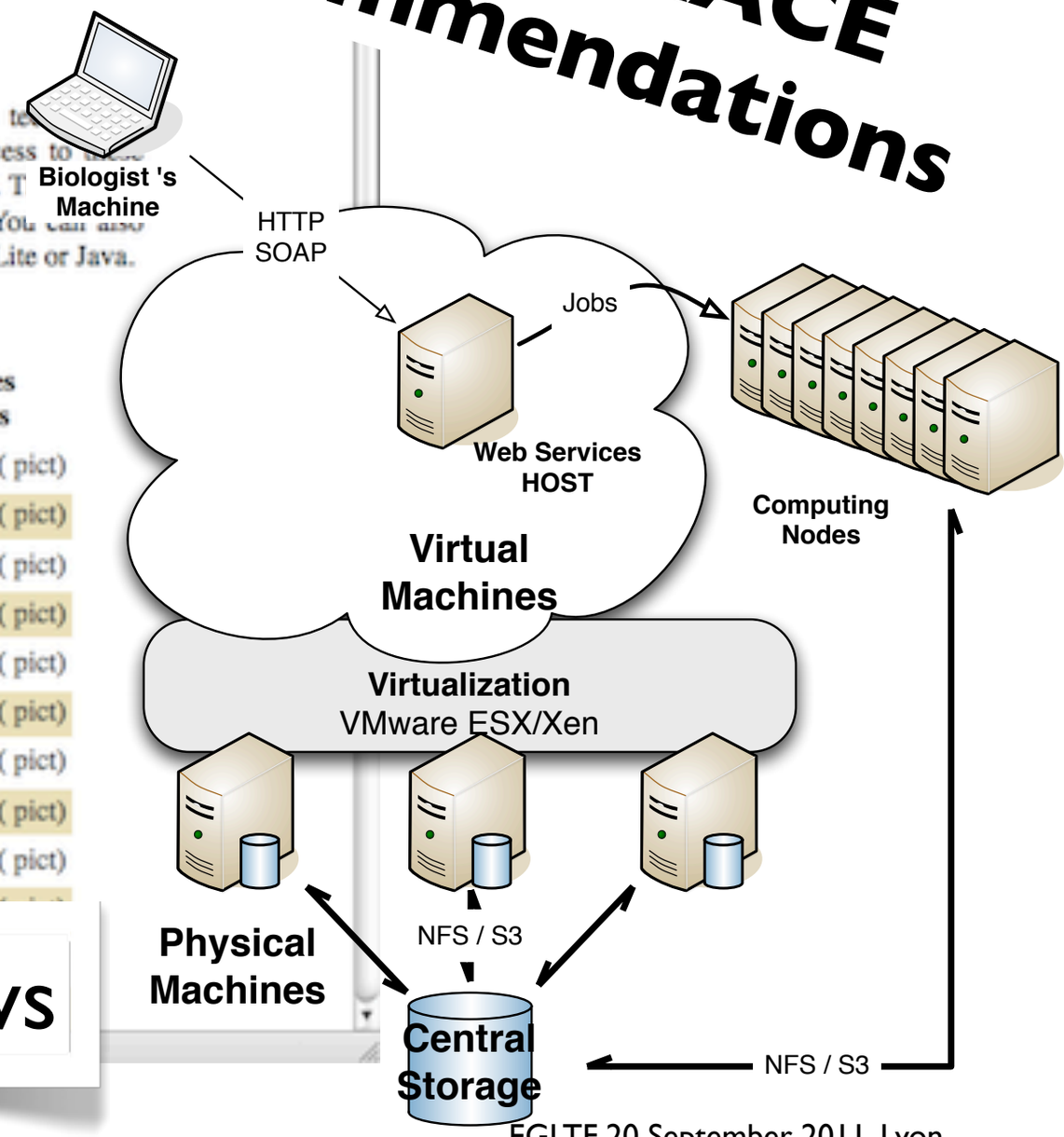
Lot of **bioinformatics portals** providing access to data and tools **without** authentication

christophe.blanchet@ibcp.fr

Scientific Gateways (2)

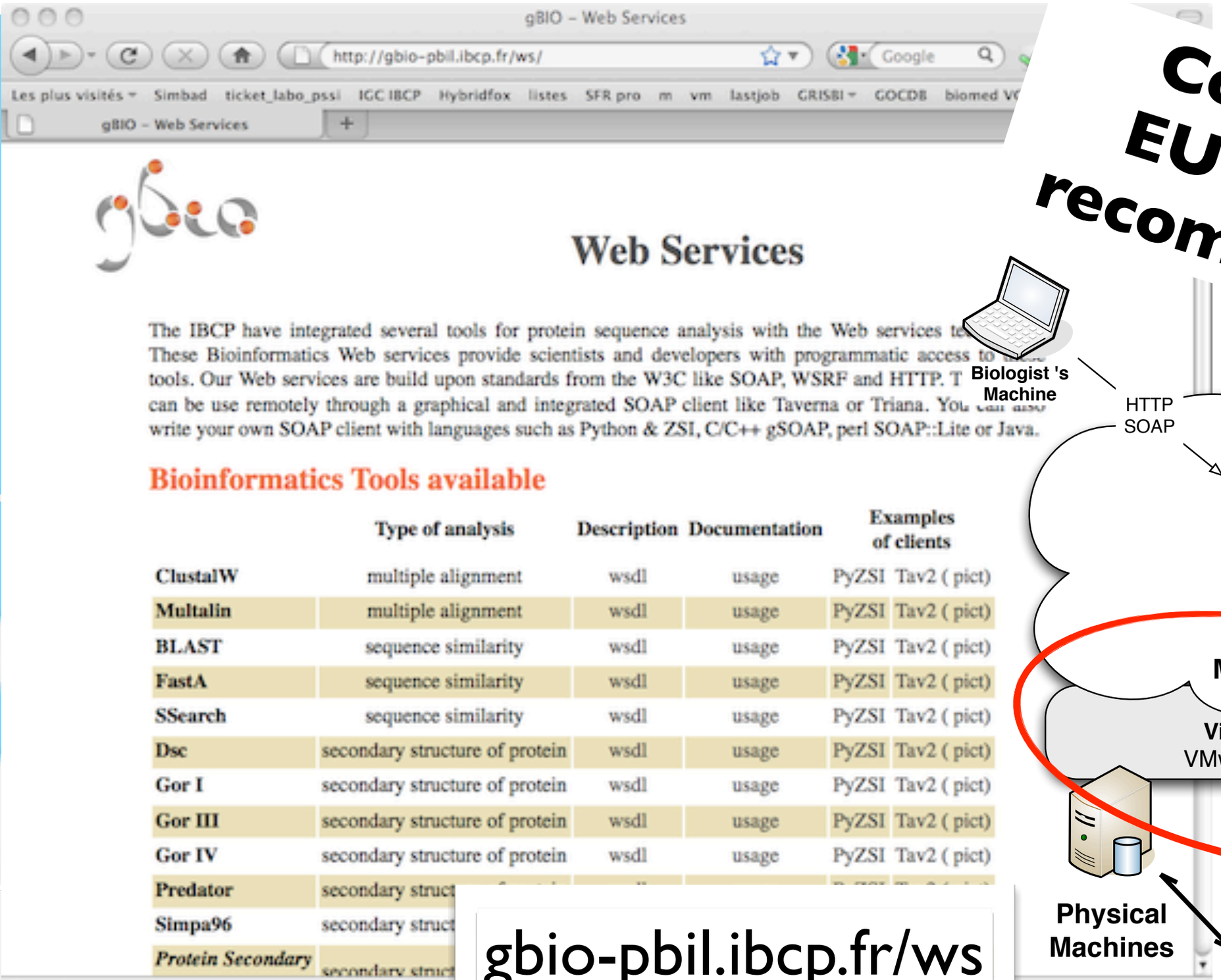


Compliant to EU EMBRACE recommendations

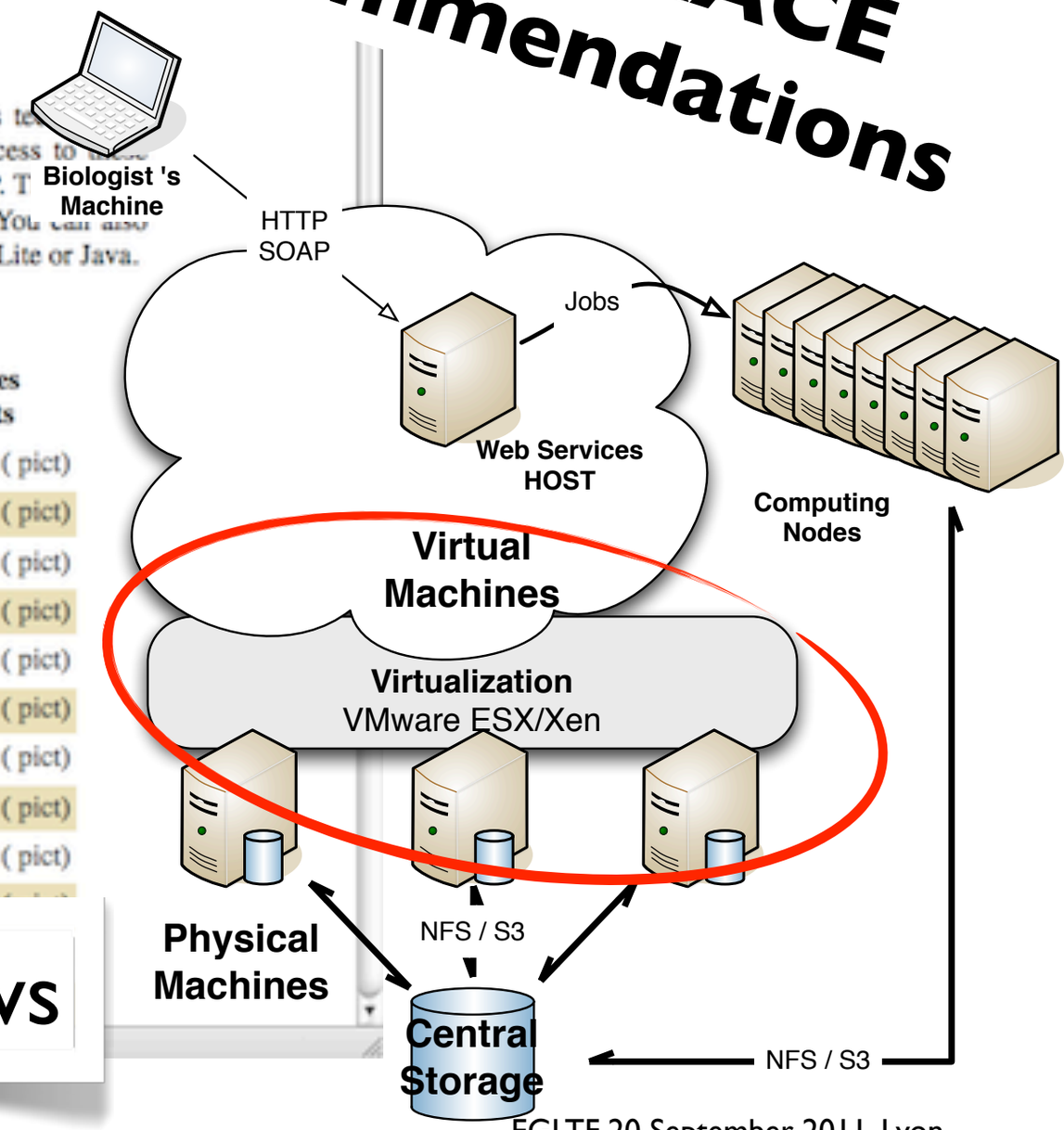


gbio-pbil.ibcp.fr/ws

Scientific Gateways (2)



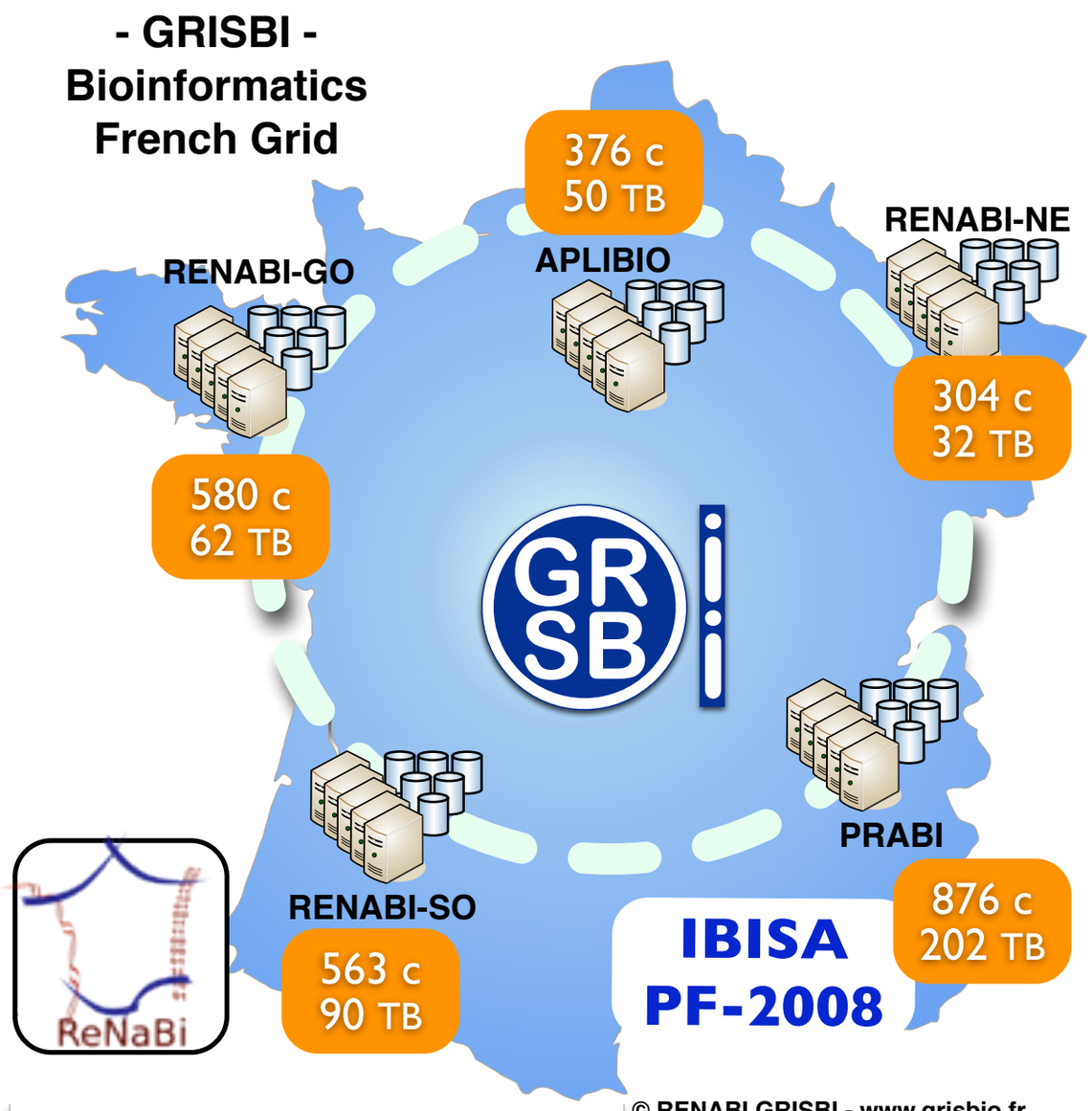
Compliant to EU EMBRACE recommendations



gbio-pbil.ibcp.fr/ws

French RENABI GRISBI

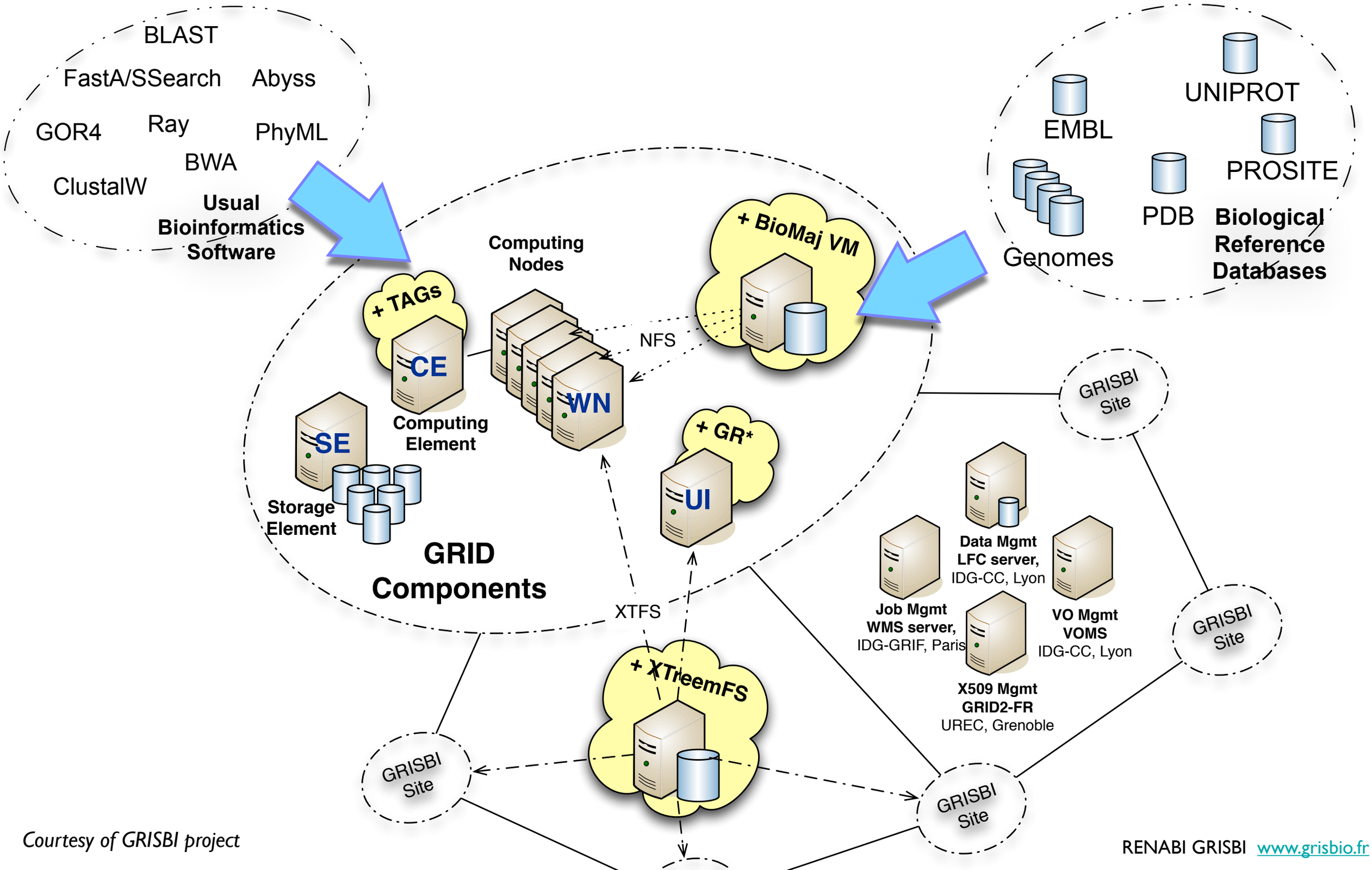
Federate Life Science community and provide with IT answers to challenging bioinformatics applications



- Working group on technological aspects:
  e.g. gLite, DIET, GridWay, BioMaj, ActiveCircle, Caringo, HDFS, XtreemFS, dCache, …
- Build a distributed infrastruture
  - among the regional centers of RENABI (5)
  - based on the bioinformatics platforms (9)
    - approved as national RIO / IBISA
    - ~70 registered members
  - Computing resources
    - in PFs **2700 cores, 440 TB** storage
    - in grid **860 cores, 30 TB** storage
- Financial support by **RENABI , IBISA 2008-2011, Institut des Grilles 2009-2010**
- In collaboration with the national IT infrastructures: Institut des Grilles, GENCI, Grid5000, regional centers

Courtesy of GRISBI project

Bioinformatics Infrastructure



Courtesy of GRISBI project

RENABI GRISBI www.grisbio.fr

StratusLab Project

Goal

- Create comprehensive, open-source, IaaS cloud distribution
- Support a wide range of use cases

Information

- 1 June 2010—31 May 2012 (2 years)
- 6 partners from 5 countries
- Budget : 3.3 M€ (2.3 M€ EC)

Contacts

- Site web: www.stratuslab.eu
- Twitter: @StratusLab
- Support: support@stratuslab.eu

CNRS (FR)

UCM (ES)

GRNET (GR)

SIXSQ (CH)

TID (ES)

TCD (IE)

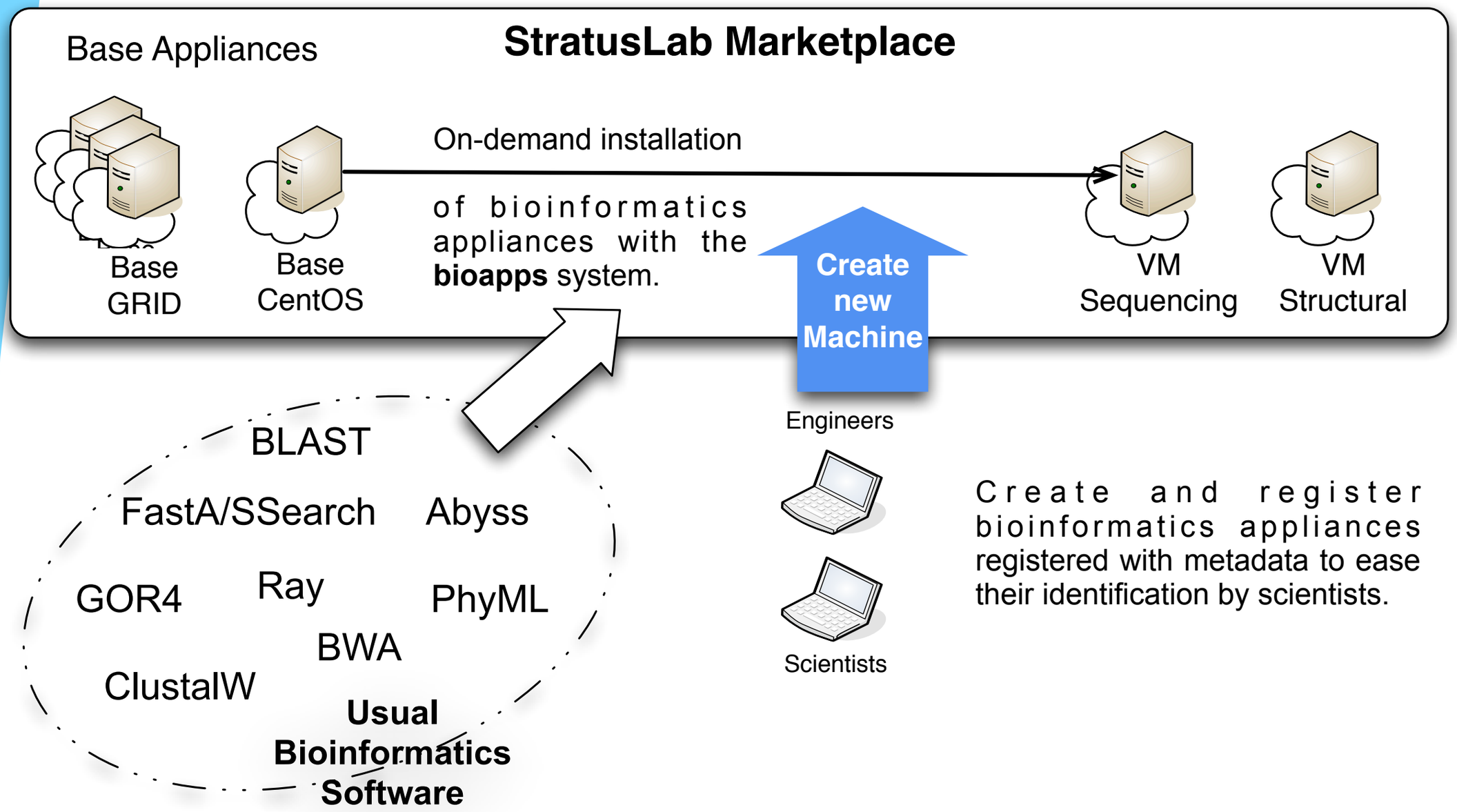
Bioinformatics Usecase

- Create **bioinformatics appliances** containing usefull applications that scientists and engineers can **deploy on demand**.
- Provide **POSIX access** (like NFS) to the storage volumes in the cloud repository containing the **biological databases**.
- Biologists and bioinformaticians are regularly **combining multiple software packages** to study their data via analysis pipelines
- Provide them with composable services via **web service** technologies with a **standard** programmatic, public, web service interface (**EMBRACE EU NoE**)

Expectations from Cloud

- Provide scientists with bioinformatics appliances to deploy on academic or commercial datacenters, or on their own computer or private cloud.
- Make the cloud infrastructure tightly connected to the storage of the the biological data.
- Ease the procedure of access by using the community's existing authentication methods
  - for example, single sign-on across portals and web services with Shibboleth technology.
- Help bioinformaticians to build and to deploy single machines, clusters, or web service infrastructures to run a complete analysis pipeline.

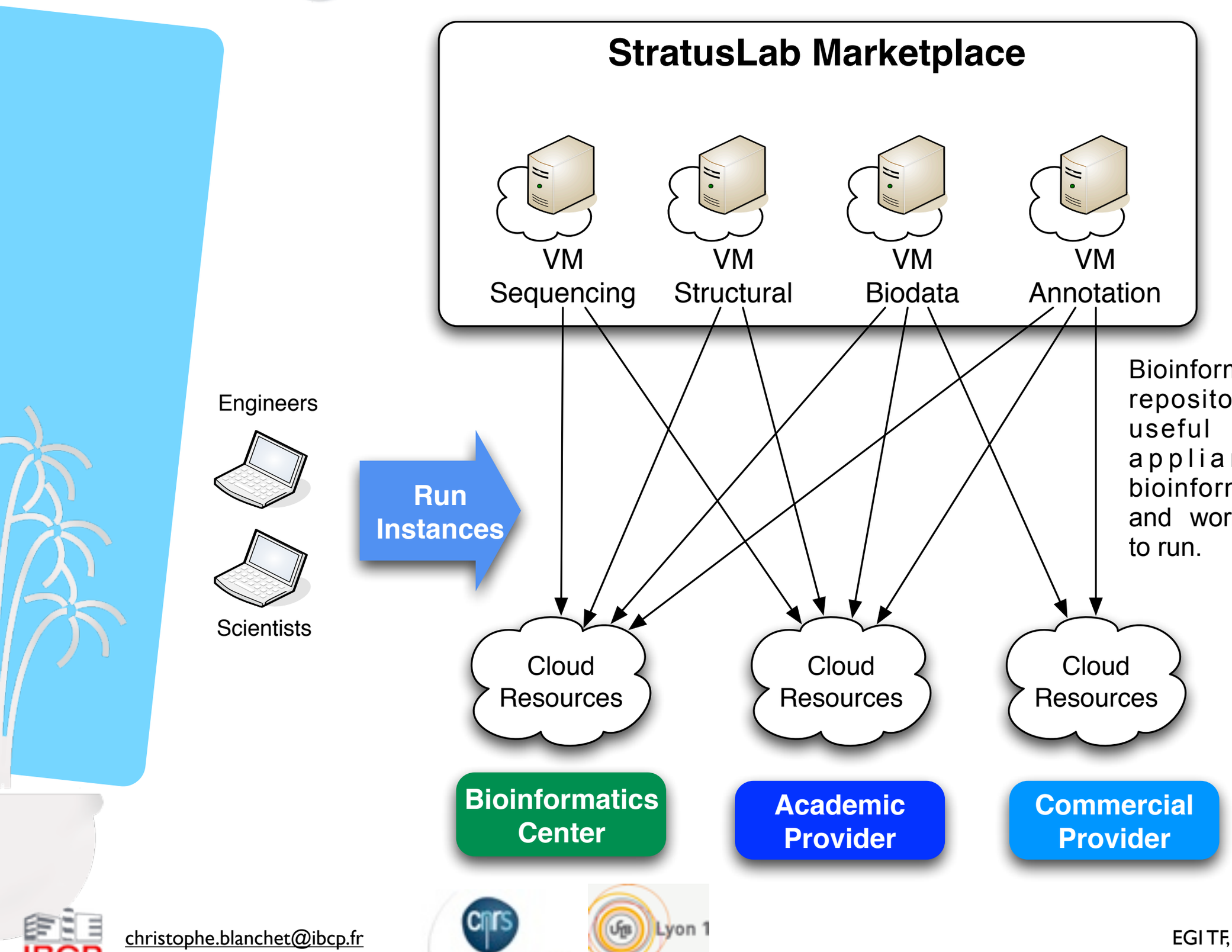
StratusLab Create Bioinformatics Appliances



Create and register bioinformatics appliances registered with metadata to ease their identification by scientists.

christophe.blanchet@ibcp.fr

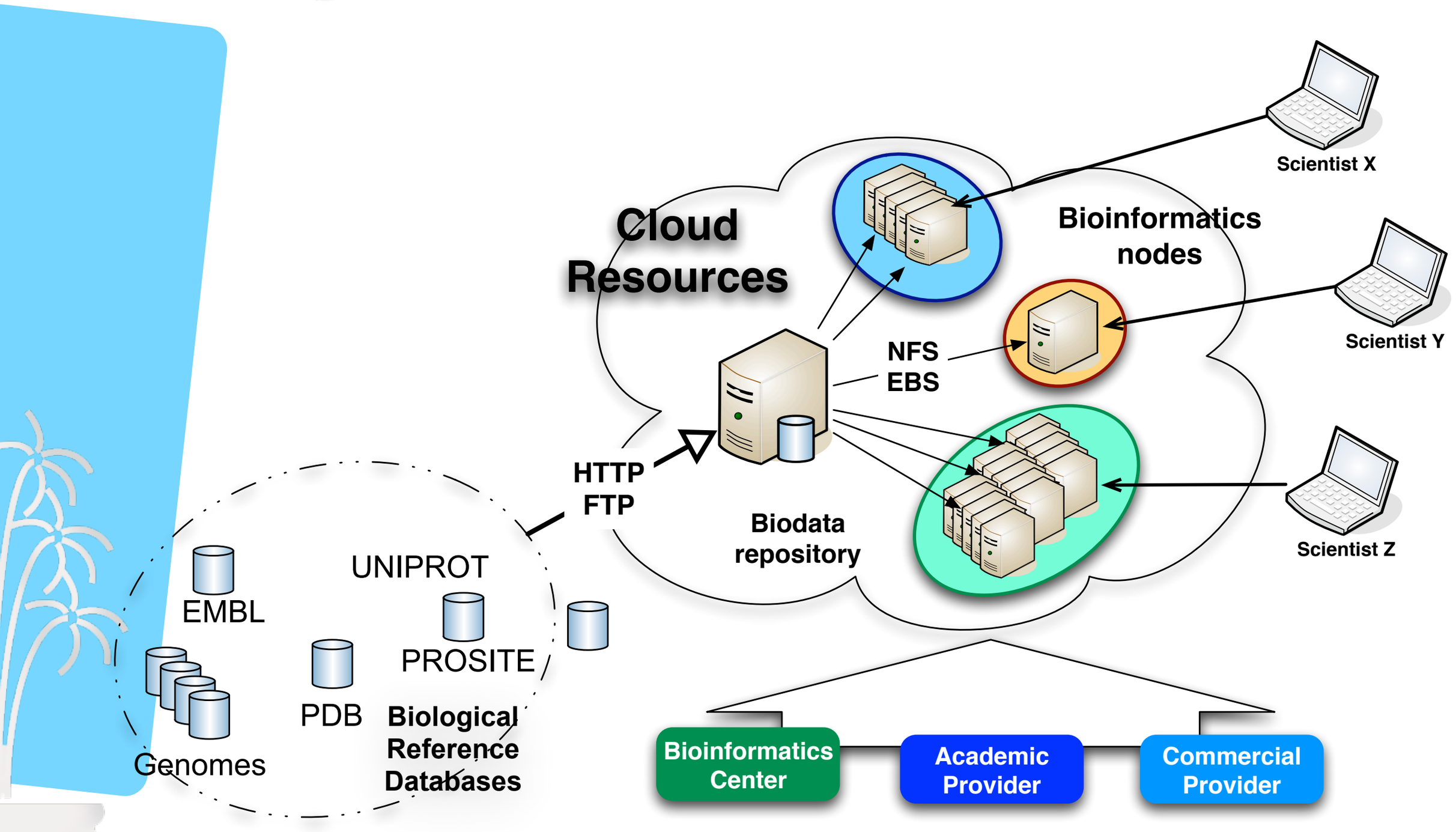
StratusLab Deploy Appliances on-demand



christophe.blanchet@ibcp.fr

StratusLab

Connected to Biological Data



christophe.blanchet@ibcp.fr

StratusLab Experimental Cloud for Bioinformatics

- Current deployment in CNRS IBCP
  - StratusLab release 1.0
  - 6 cloud nodes (48 cores, 192 GB RAM) + front-end
  - usual constraint of not having enough public IPs
    - port translation on the front to reach the VMs
- Provide bioinformaticians with an experimental platform to evaluate the cloud technology
  - in collaboration with the RENABI GRISBI services
  - will be open to the French community (initially)
  - access based on national Educ-Research federations
    - CNRS GRID2-FR (X509), Renater (shibboleth)
  - provide users with an interactive method to create a new appliance from a base one

What next ?

- Help Bioinformatics centres to provide scientists with services on the cloud
- Platform-as-a-Service
  - Provide relevant Bioinformatics appliances
  - Easily searchable: the marketplace
  - To deploy on academic/commercial datacenters
    e.g. RENABI centres, StratusLab clouds (GRNET, LAL)
  - or on their own computer/cloud
- Infrastructure-as-a-Service
  - Complete infrastructures
    - web services and portal with backoffice cluster,
    - multi-nodes applications (e.g. ARIA/ISD),
    - GRISBI grid sites
  - Complete pipelines for bioinformatics analysis

Thanks

- Questions ?

- Acknowledgment:
  - C. Gauthey, StratusLab members
  - The European Commission through the project FP7 StratusLab -- 2010-12 -- RI-261552

christophe.blanchet@ibcp.fr