

E-BIOGENOUEST: A REGIONAL LIFE SCIENCES INITIATIVE FOR DATA INTEGRATION

Datacite Annual Conference 2014 - Nancy

Olivier Collin – IRISA/INRIA

Olivier.Collin@irisa.fr

<http://www.genouest.org>



Agenda

- Context
 - Biogenouest
 - Biology
- The e-biogenouest project
 - “Bridging data, metadata and computation”
- A system of systems : collaborative portal, metadata management environment, data analysis portal

Biogenouest



Biogenouest is a network bringing together technological core facilities dedicated to Life and Environmental Sciences in the West of France



Biogenouest



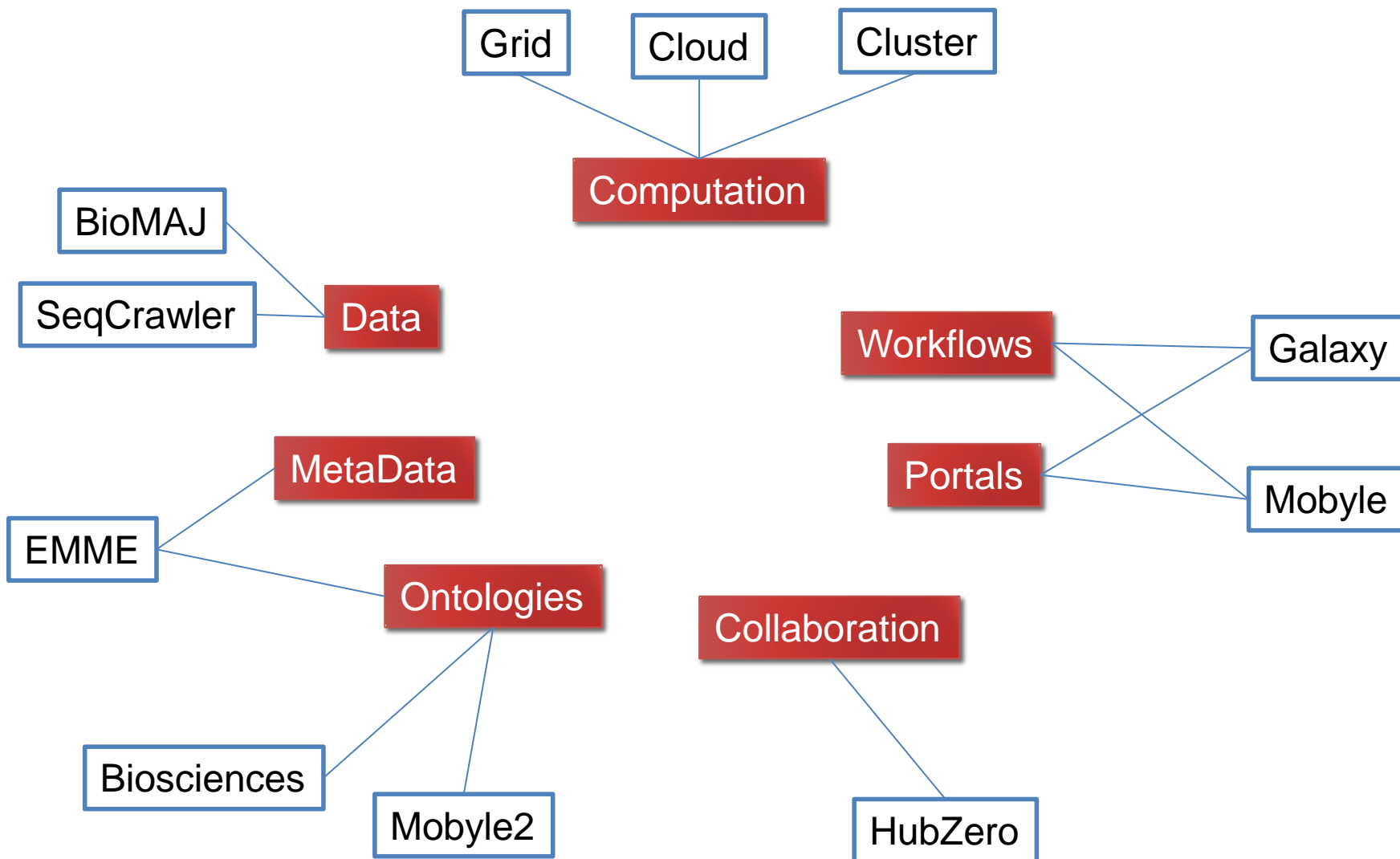
Created in 2002, **Biogenouest coordinates 31 technological core facilities** based in the regions of Brittany and Pays de la Loire, with the aim to **organize and pool interregional resources**.

Biogenouest also **federates 70 research units** involved in thematic research covering 4 areas of activity : **Marine resources, Agri-food, Health and Bioinformatics**.

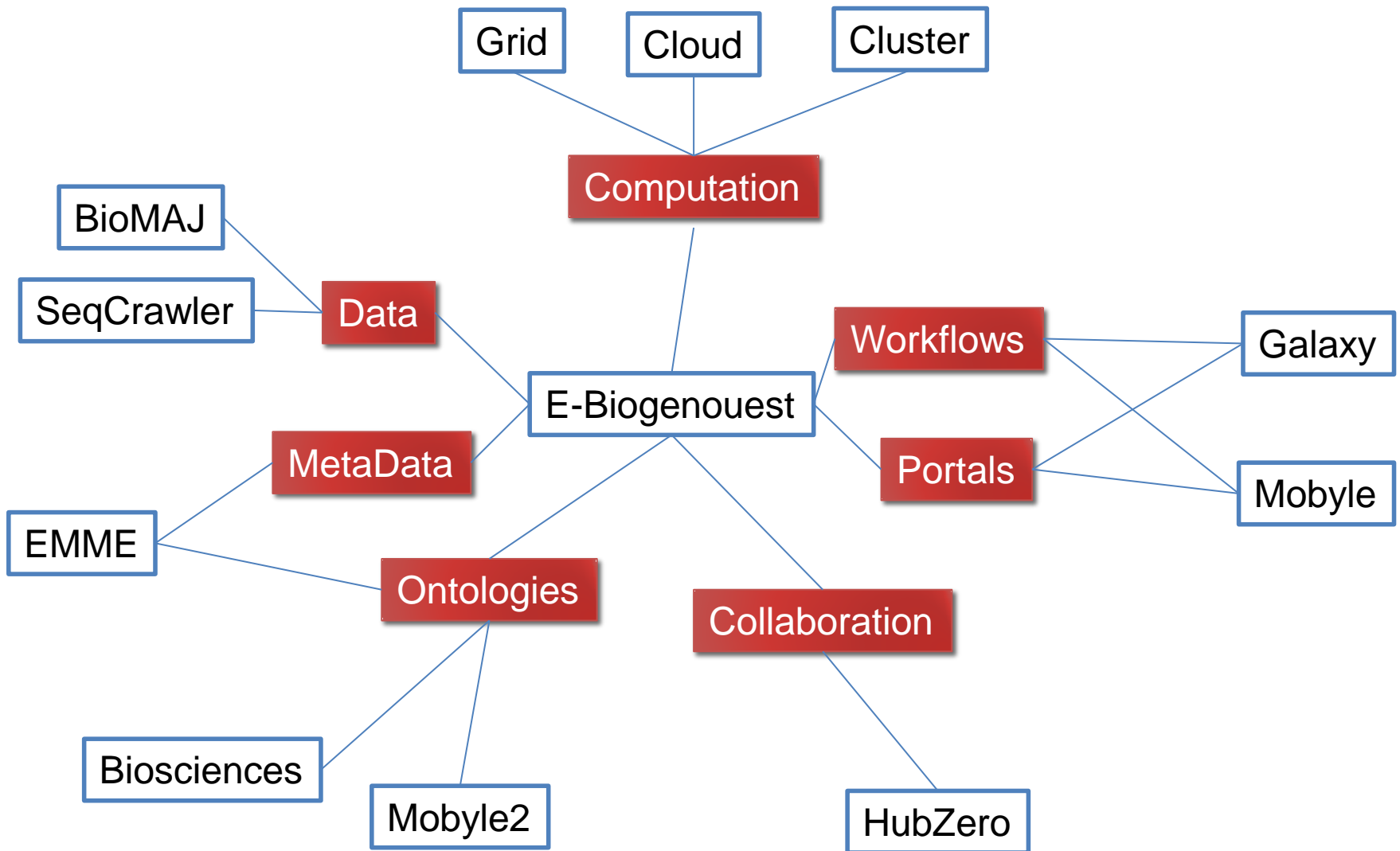
GenOuest : Bioinformatics core facility

- Member of the Biogenouest network
 - Member of the IFB : French Bioinformatics Institute
 - National recognition : IBiSA platform
 - Regional strategic facility for INRA (National Institute of Agronomical Research)
 - ISO9001:2008 certified
-
- Established since 2002
 - 10 to 12 people
 - Computing infrastructure, storage, software development, expertise, R&D projects

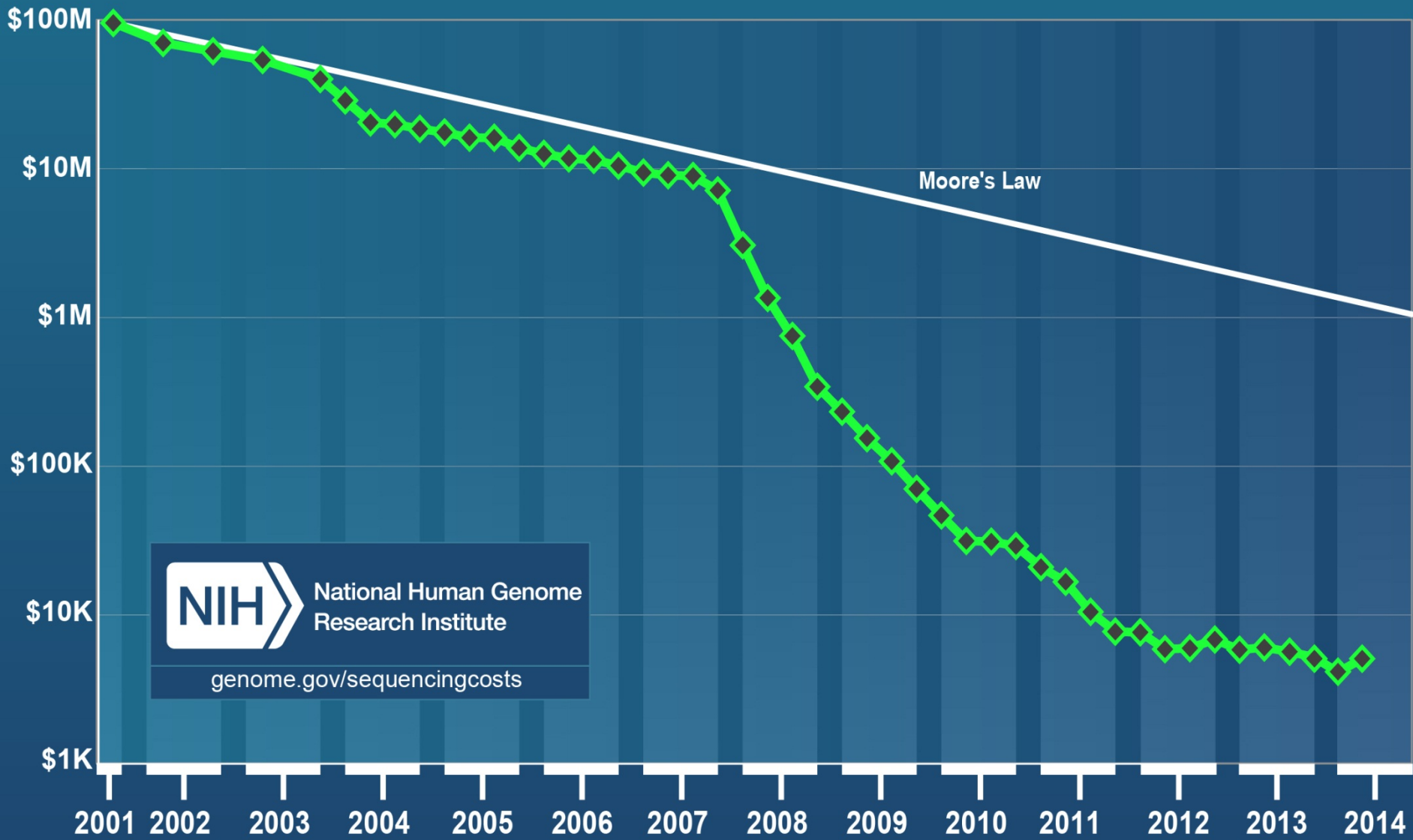
R&D projects



R&D projects



Cost per Genome



Context

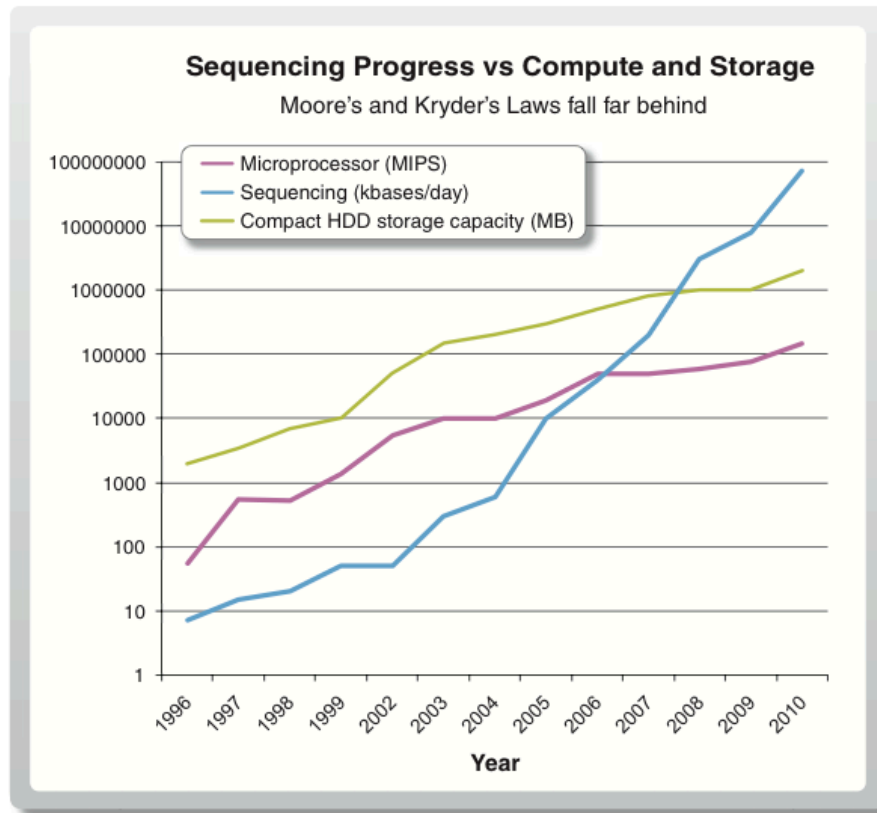


Fig. 1. A doubling of sequencing output every 9 months has outpaced and overtaken performance improvements within the disk storage and high-performance computation fields.

Kahn. On the future of genomic data. *Science* (2011) vol. 331(6018) pp. 728-9

➤ Now : Genomics : Next Generation Sequencing

➤ Next : Proteomics

➤ Next : Bio-imaging

➤ Digital data

➤ Huge amount

➤ Heterogenous

➤ Critical situation for some laboratories

E-BIOGENOUEST

E-Biogenouest

- Started in May 2012 for 3 years
- Funded by Brittany and Pays de la Loire
- E-science initiative for the Biogenouest network

- Community building
- Training/workshops
- Roadmap preparation
- Experimentation/Pilot project : Virtual Research Environment (VRE)

A system of systems

- Combination of various tools
 - A data analysis portal : Galaxy
 - A metadata management tool : ISAtools suite
 - A collaborative portal : HubZero
 - Additional utilities :
 - Pydio : file transfer
 - Some software glue to make it work...
 - BioBlend : Galaxy API
 - In-house developments



Galaxy portal

- Galaxy : a web based portal for biomedical data analysis
 - Intuitive interface
 - Workflows
- Galaxy@Genouest
 - 800 tools (transcriptomics, population genetics, quantitative genetics, metagenomics, proteomics, etc.)

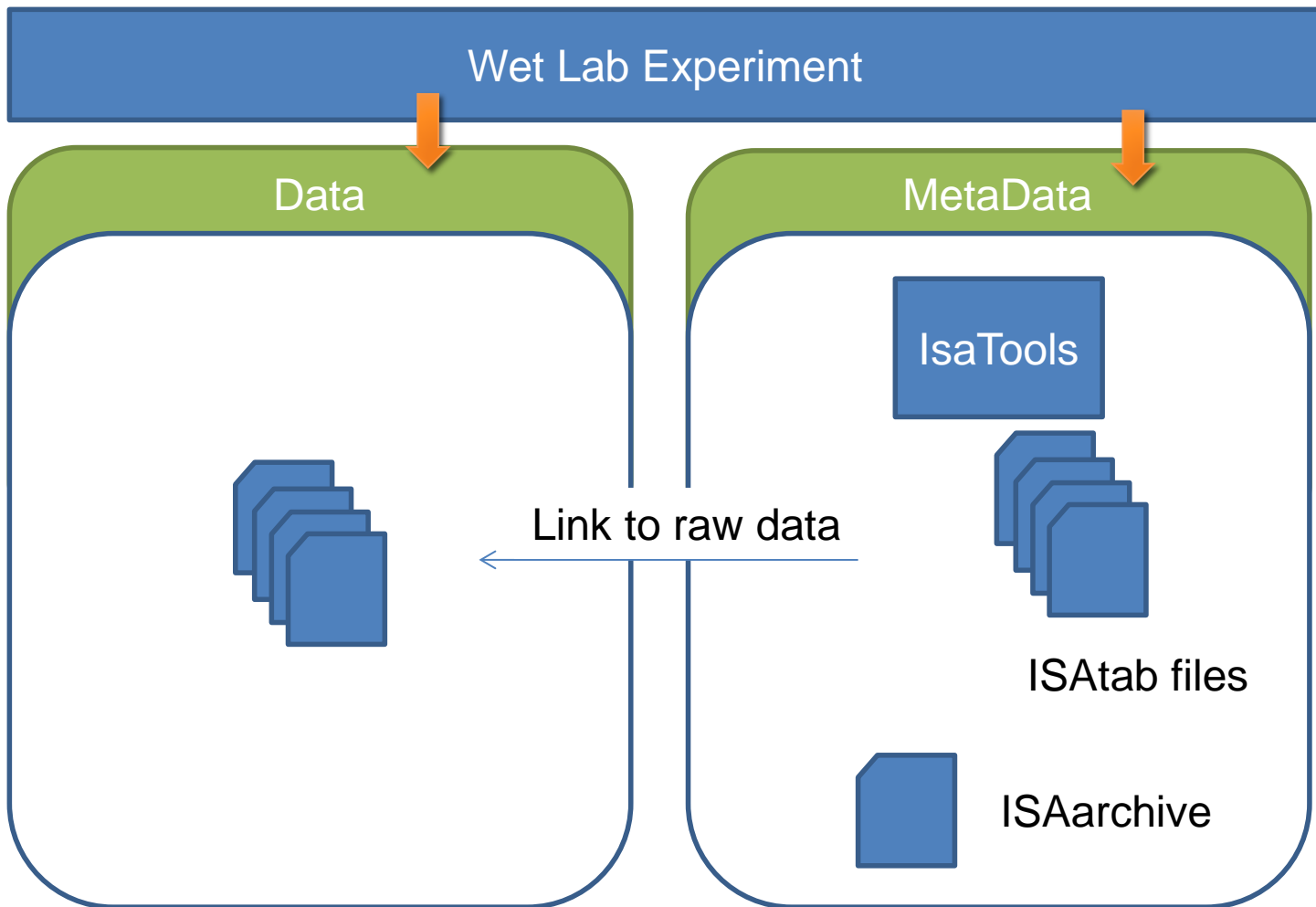
- <http://galaxyproject.org/>

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A. "Galaxy: a platform for interactive large-scale genome analysis." *Genome Research*. 2005 Oct; 15(10):1451-5.

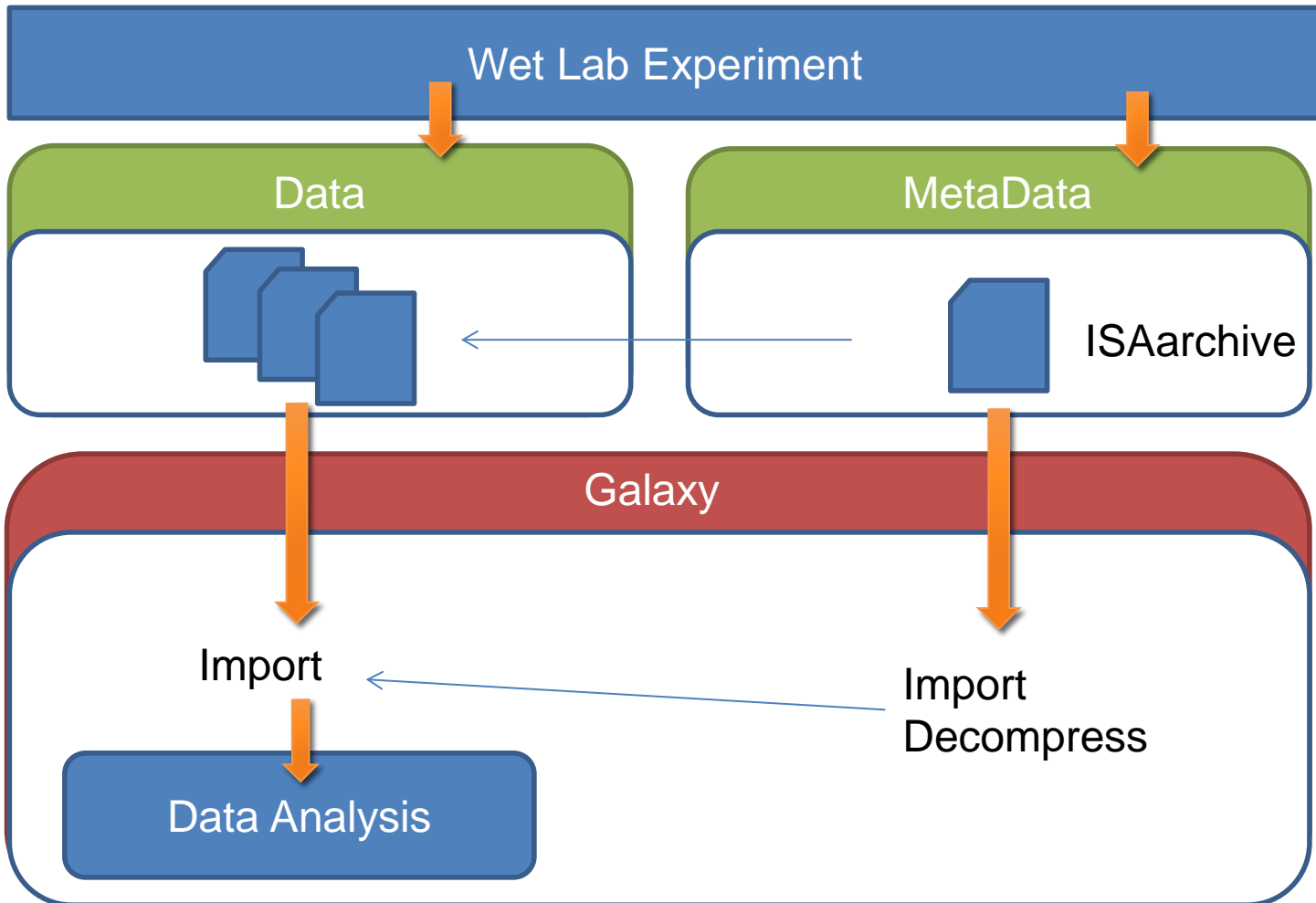
ISAtools Suite

- Open Source tools for experimental metadata management
 - Enforces the description of experiments with standards or ontologies
 - Creates local repository
 - Allows publication to public repositories
- ISA@GenOuest = EMME
 - Additional developements and auxiliary tools.
- <http://www.isa-tools.org/>
- Rocca-Serra, P. *et al.* ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level. *Bioinformatics* **26**, 2354–6 (2010).

EMME



EMME



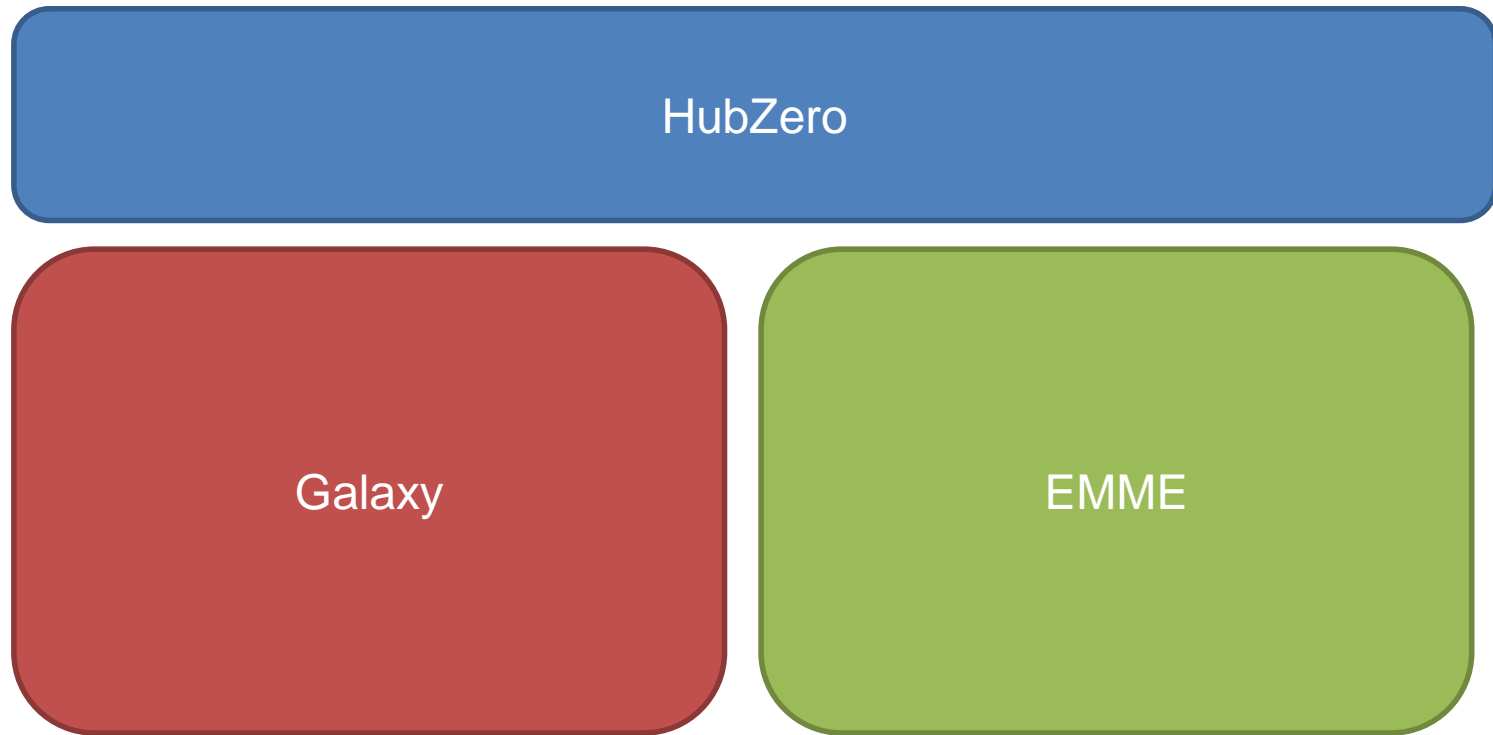
HubZero

- Scientific web portal
 - Collaboration: wiki, blog, etc.
 - Resources : results, articles, presentations, etc.
 - Lightweight project management

- <https://hubzero.org/>

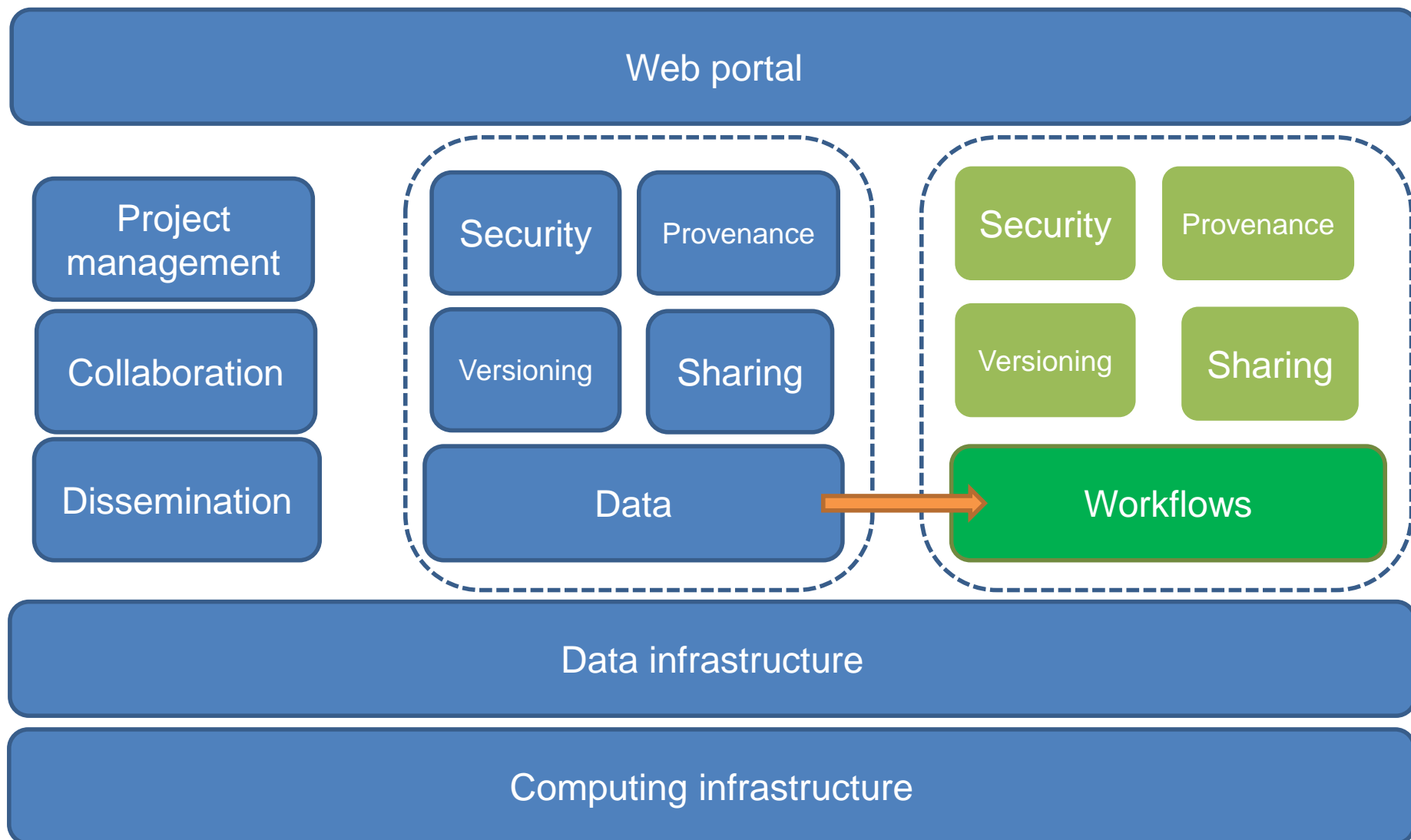
M. McLennan, R. Kennell, "HUBzero: A Platform for Dissemination and Collaboration in Computational Science and Engineering," *Computing in Science and Engineering*, 12(2), pp. 48-52, March/April, 2010

Continuum



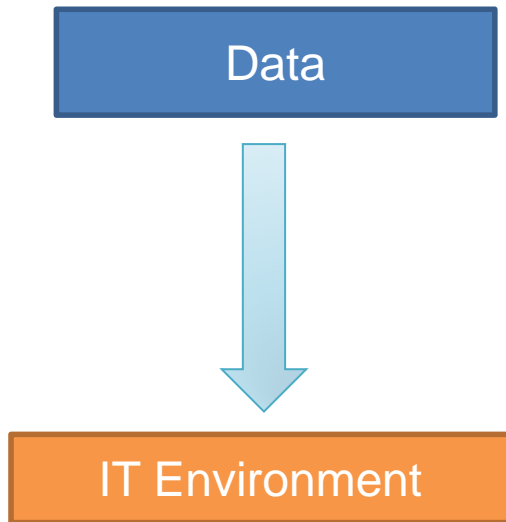
- Continuum for the management and analysis of biological data
- Collaborative environment

VRE : Virtual Research Environment

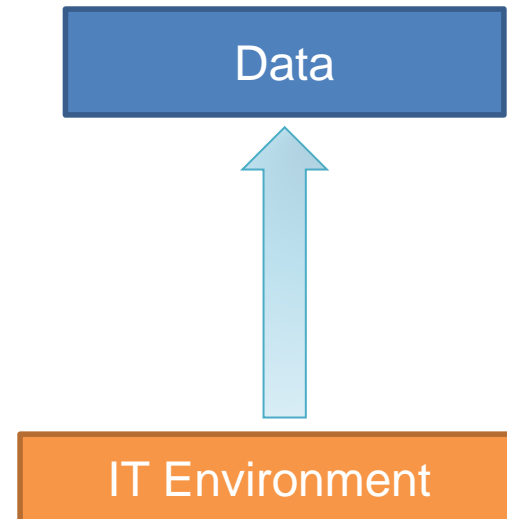


A paradigm shift

From...



To...



Next steps

- What we learned :
 - Acceptance / adoption issues are key issues
- What we will do :
 - Switch to a production environment
 - Identity federation
 - ISA-Dataflow : metadata for bioinformatics workflows

What we need to do :

- To connect to other initiatives
- To define the perimeter :
 - Big changes for bioinformatics facilities

Conclusion

- Biology becomes a digital science
- New technologies with lower costs create a dangerous situation
- A system of systems :
 - « metadata + collaborative tool + analysis portal »
- Continuum : data centered philosophy
 - « Bring back Biology to the biologist »

Questions ?

- Olivier.Collin@irisa.fr
- <http://www.genouest.org>
- <https://www.e-biogenouest.org>