Promoting Grids and Clouds for the Health Science and Medical Research community in France

Gilles MATHIEU, Inserm/DSI
gilles.mathieu@inserm.fr

EGI Community Forum 2015, Bari, Italy – 10 November 2015

Outline

- **About Inserm**

- **Scientific computing at Inserm**

- **Grids and clouds**

- **Virtual Research Environments**

What is Inserm ?

- **French National Institute of Health and Medical Research**
  - Created in 1964
  - national coordinator in biomedical research
  - high-level international stakeholder
  - develops new medical applications
  - produces and disseminates knowledge
  - promotes clinical innovations quickly off the ground
  - develops partnerships with patient associations and society

3

Inserm Key Figures

- **15,000** health research professionals
- **276** research units nationwide
- **40** clinical research centers
- **10k** scientific articles published yearly
- **€989 million** yearly budget (2014)
- Innovation & technology transfer
    - **150** patents filed each year
    - **800** R&D contracts and industrial licenses
    - **50** European and International cooperative agreements

Inserm Pluri-disciplinarity

- **9 mult-organization thematic institutes**
    - Neurosciences, cognitive sciences, neurology and psychiatry
    - Cancer
    - Immunology, Inflammation, Infectiology and Microbiology
    - Physiopathology, Metabolism, Nutrition
    - Public health
    - Health technologies
    - Molecular and structural basis of life sciences
    - Cell biology, development and evolution
    - Genetics, genomics and bioinformatics

Scientific Computing at Inserm

Medical imaging + pre-clinical imaging

E-Health, Knowledge Engineering

Bioinformatics and molecular analysis

Hadrontherapy

Ergonomics

Medical Robotics and interventional techniques

Biostatistics and Biomathematics

More coffee, please

Molecular modeling, chemoinformatics

Medical embedded systems

Numerical simulation

Challenges: Scientific and technical

- **Challenges are at every step of research workflows**
  - Initial processing, raw data storage, data structuration, metadata, data processing…

- **Human health specificities**
  - Data privacy concerns, especially clinical data
  - Legal obligations wrt data conservation, territoriality, etc.
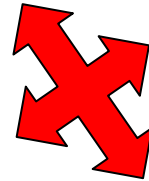 - Wide range of constraints between research and clinical area

7

Challenges: Technological

Desktop computers

Isolated servers

Exponential entropy
Many different strategies
and spread resources

Isolated clusters
or platforms

Data centres and
e-infrastructures

8

Challenges: A cultural shift

- **Transformation needed from**

© Albert Uderzo, René Goscinny - Astérix

- **To**

© "The Hobbit : the Battle of the Five Armies" - www.theonering.net

The CISI team

- **A team to support Inserm research teams in the field of scientific computing**
  - CISI = Scientific Computing Coordination at Inserm
  - Team created in 2014

- **Addressed topics**
  - Scientific computing technologies and areas of interest
    - Parallel and distributed algorithms, Grids and Clouds, HPC, Big Data
  - Resources sharing and infrastructure optimisation

CISI's development axes

MAPPING

Get a correct understanding of existing IT infrastructures at Inserm and their usage

TRAINING

Help researchers and engineers acquire needed competencies in the various scientific computing domains
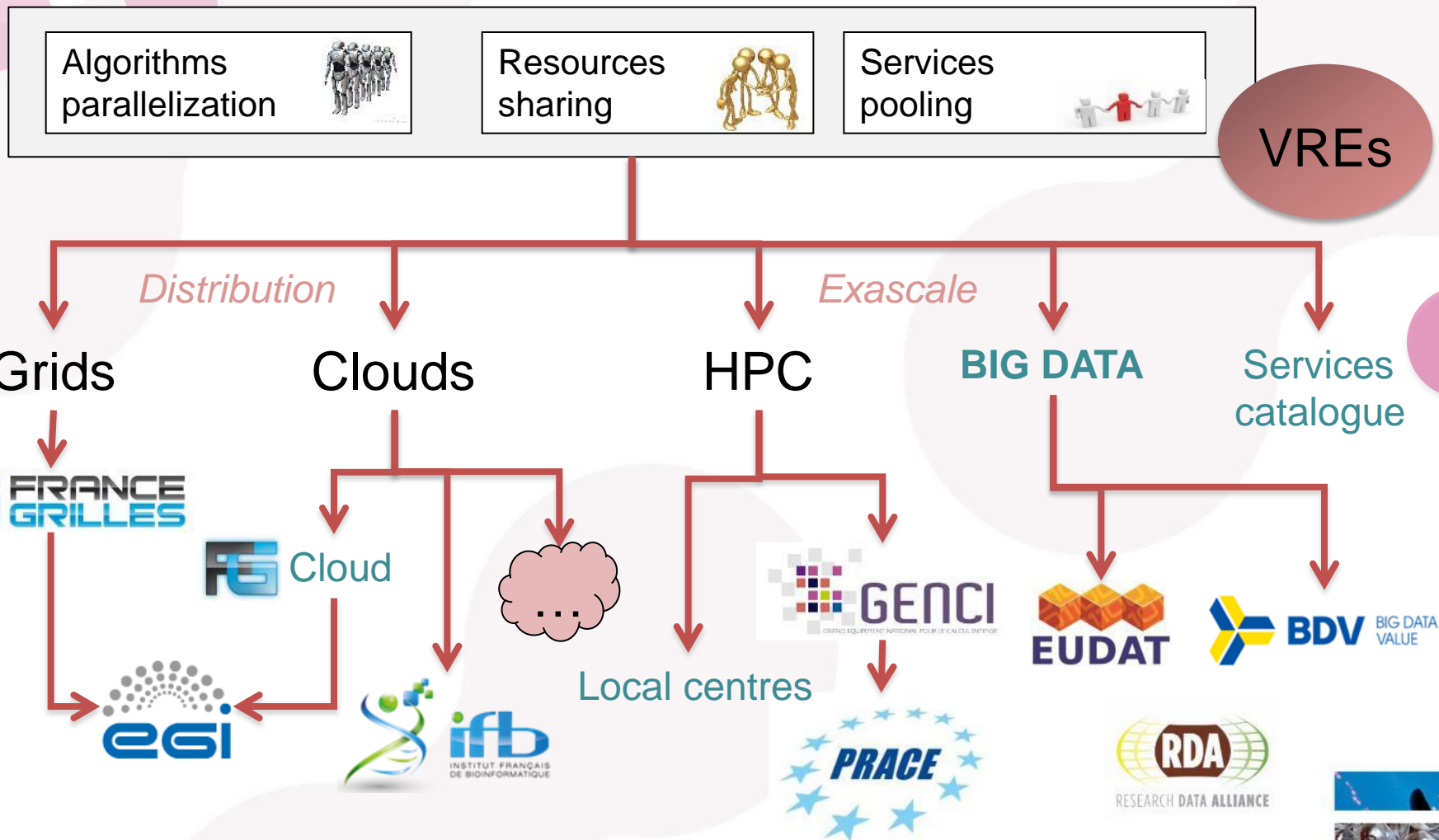
COUNCELLING

Provide high quality expertise and support to Inserm researchers and engineers on identified areas

PROJECTS

Foster national and international collaborations

Grids & Clouds at Inserm: current state

- **Life Science Grid Community in France is very active, but Inserm does not take full benefit of it yet**
  - Less than 10 Inserm research units are currently using EGI
  - Most are doing so at small scale (1-2 users, often PhD students)

- **And yet, Inserm…**
  - … is a partner of the French NGI
  - … is involved in many ESFRIs that collaborate with EGI
  - … has many research topics that are typical grid/cloud usecases

- **There is room for improvement!**

13

Potential grid/cloud users at Inserm

- **Medical imaging communities**
  - Imaging applications (simulation, analysis) work well on grids/clouds

- **Genomics, genetics and bioinformatics communities**
  - NGS, molecular analysis, chemo-informatics, …

- **E-health and public health communities**
  - Widely distributed applications, embarqued medical equipments, connected objects, biobanks, …

14

Grid/cloud diffusion strategy (1)

- **Identification of existing users**
  - Build up examples
  - Identify success stories
  - List problems and possible improvements
  - Organise bases for an efficient knowledge transfer

- **Detection of potential new users**
  - "pre-selection" on the discipline/domain
  - Needs analysis through our "mapping" action
  - Identify workflows similar to those identified above, that could benefit from using a grid or cloud infrastructure

15

Grid/cloud diffusion strategy (2)

- **Information**
  - Organise talks and seminars
  - Promote related events
  - Propose communication tools (website, forums, collaborative documentation)

- **Training and follow-up**
  - Organise trainings with the help of France Grilles, our NGI
  - Provide support
  - Encourage knowledge transfer through the creation of working groups and groups of interest

16

First concrete actions

- **Identification of existing and potential users**
  - Mapping action now in progress

- **Information**
  - Redaction of a "Grids and clouds for dummies" guide in French
  - 2 seminars this month given to Inserm research units
  - Website and documentation platform ready

- **Trainings**
  - DIRAC training in june 2015 (12 participating Inserm users)
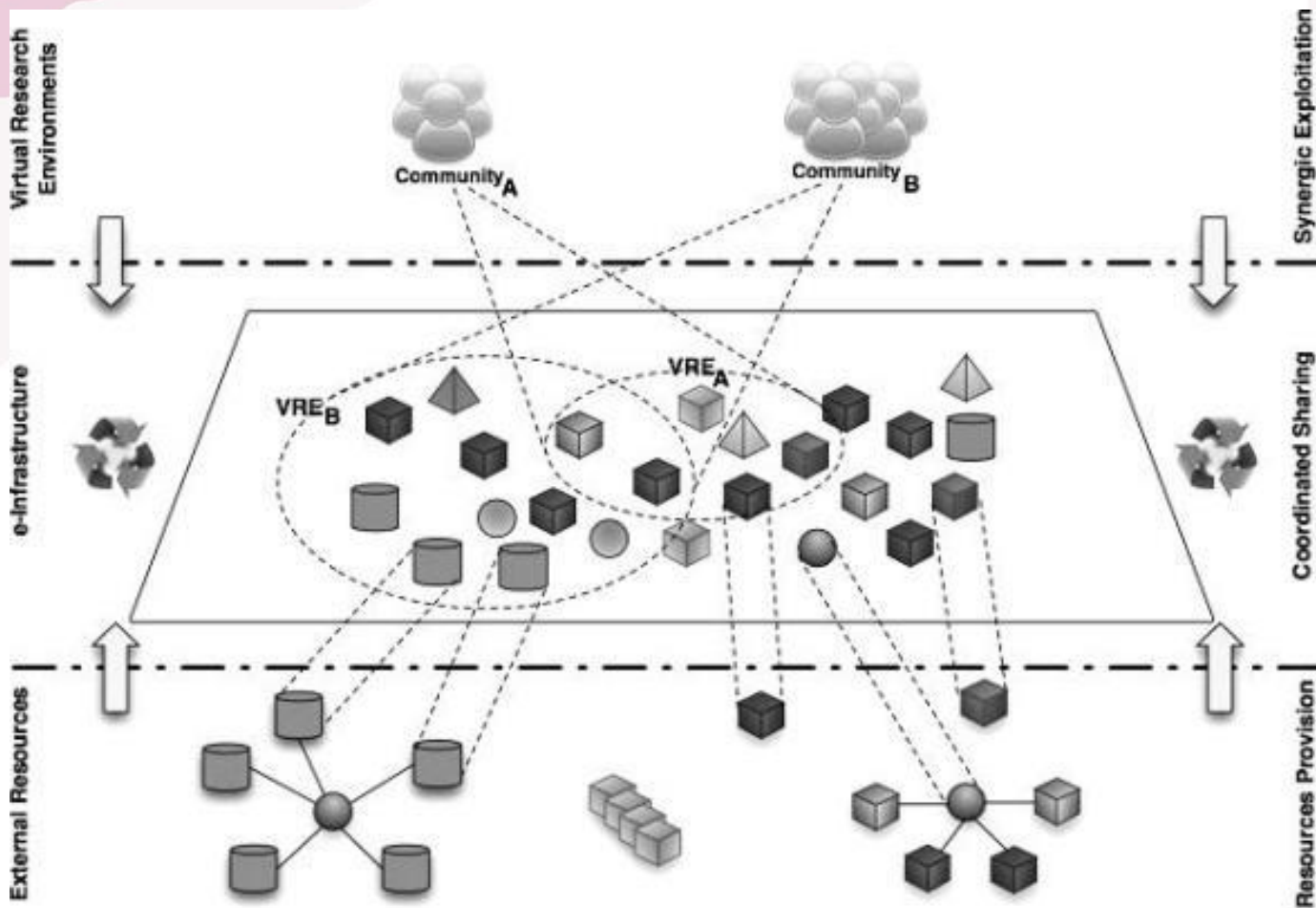 - Promotion of France Grilles cloud trainings – fall 2015

These actions are not specific to grids

- **Engagement toward the HPC community in a similar way**
  - Through our local/Regional HPC centres (Mesocentres)
  - Through our national coordination body (GENCI) as a gateway to PRACE

- **CISI can act as a broker**

18

Vision: empowering users through Virtual Research Environments

Vision: empowering users through Virtual Research Environments

- **Well identified general benefits**
  - Mask infrastructures heterogeneity
  - Foster collaborative work
  - Allow efficient data and resources sharing
  - Improve user experience

- **Adapted to our case**
  - Allow cross-platforms workflows (e.g. grid/HPC)
  - Provide single entry points to distributed e-infrastructures
  - Facilitate needed cultural shift
    - *"Scientific Computing with a human face"*

Vision: empowering users through Virtual Research Environments

- **Identified challenges**
  - Handle Inserm specificities, especially wrt security and data privacy issues
  - Requires an integrated handling of underlying resources
  - Match user needs with available resources in a coherent way
  - Should involve the whole stack, from researchers to technology providers
  - Requires a smart combination of bottom-up and top-down approaches

- **This is a long term, large scale project**

Credits – thanks – links

- **The CISI Team**
  - Isabelle Perseil, Itebeddine Ghorbel, Gwénaël Dumont, Daniel Salas, Nathalie Loriod
  - http://cisi.inserm.fr
- **Inserm**
  - http://www.inserm.fr
- **France Grilles**
  - http://www.france-grilles.fr

- **Questions welcome !**