



Instituts
thématiques



Inserm

Institut national
de la santé et de la recherche médicale

Favoriser l'accès aux grilles et clouds à l'Inserm

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

Plan

- **A propos de l'Inserm**
- **L'informatique scientifique à l'Inserm**
- **Grille et cloud à l'Inserm**
- **Environnements de recherche virtuels**



L'Inserm

- **Institut National de la Santé et de la Recherche Médicale**
 - Créé en 1964
 - Coordinateur national pour la recherche biomédicale
 - Acteur international de haut niveau
 - Découvreur d'applications médicales
 - Producteur et diffuseur de connaissances de référence
 - Accélérateur d'innovations cliniques
 - Développeur de liens avec le tissu associatif et la société

➤ <http://www.inserm.fr>



Quelques ordres de grandeur

- **15,000** personnes
- **276** laboratoires/structures de recherche
- **40** centres d'investigation clinique
- **10,000** articles scientifiques publiés par an
- **989 millions d'euros** de budget annuel (2014)

- Innovation & transfert de technologie
 - **150** brevets déposés chaque année
 - **800** contrats R&D et licences avec l'industrie
 - **50** accords et conventions de collaborations européens et internationaux



La pluri-disciplinarité à l'Inserm

- **9 Instituts thématiques**

- Neurosciences, sciences cognitives, neurologie, psychiatrie
- Cancer
- Immunologie, Inflammation, Infectiologie et Microbiologie
- Physiopathologie, Métabolisme, Nutrition
- Santé Publique
- Technologies pour la Santé
- Bases Moléculaires et Structurales du Vivant
- Biologie Cellulaire, Développement et Evolution
- Génétique, Génomique et Bioinformatique

- **Chacun a un équivalent multi-organisme (AVIESAN)**



L'informatique Scientifique à l'Inserm

Imagerie médicale + préclinique

E-santé

Bioinformatique et analyse moléculaire

Hadronthérapie

Ergonomie

Robotique médicale et techniques d'interventions

Biostatistiques et Biomathématiques

Encore du café svp

Modélisation moléculaire, chémoinformatique

Systèmes médicaux embarqués

Simulation numérique



Les enjeux de l'informatique scientifique aujourd'hui

Réduire le temps de calcul

Pour obtenir des résultats plus vite, plus de résultats en un même temps, voire obtenir un résultat tout court

Gérer les données efficacement

Pour faire face à la taille et au volume croissant des données générées, traitées, stockées, archivées

Partager les résultats et les outils

Pour permettre des collaborations efficaces



Les challenges

- **Un volume de données toujours croissant**
 - Accroissement exponentiel pour de nombreux domaines
- **Des contraintes fortes de sécurité**
 - En particulier à l’Inserm pour ce qui concerne les données médicales
- **Des collaborations multiples**
 - Besoin de partager à grande échelle
- **Une diversification du métier de chercheur**
 - Des compétences informatiques pour compléter les compétences métier



Challenges scientifiques et techniques

- **Challenges à toutes les étapes**
 - Acquisition, stockage de données brutes, structuration des données, méta-données, analyse...
- **Spécificités de la santé humaine**
 - Contraintes de confidentialité (données cliniques)
 - obligations légales sur la rétention, la territorialité, etc.
 - De nombreuses contraintes pour la liaison entre le domaine clinique et le domaine de la recherche



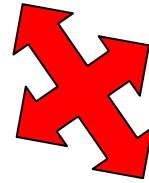
Challenges technologiques



PC de bureau



Serveurs isolés



Entropie exponentielle

Nombreuses stratégies
Ressources éparpillées



Clusters ou
plateformes isolés



Data Centers,
e-infrastructures



Challenges culturels

- Besoin de transformer l'approche générale de :



© Albert Uderzo, René Goscinny - Astérix

- vers



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



CISI: Coordination de l'Informatique Scientifique de l'Inserm

- **une structure du DSI fonctionnant avec un pilotage scientifique des Instituts thématiques de l'Inserm**
 - Conduite par Isabelle Perseil, constituée de 6 personnes
 - Lancée et validée par la Direction Générale de l'Inserm en 2014
- **Missions**
 - Support et accompagnement aux équipes de recherche :
 - Algorithmique parallèle et distribuée
 - Grille et Cloud Computing
 - HPC
 - Big Data
 - Mutualisation et optimisation des infrastructures



Les 4 axes de développement de CISI

CARTOGRAPHIE

Assurer une bonne connaissance de l'existant en infrastructures informatiques à l'Inserm

FORMATION

Faciliter la montée en compétence des ingénieurs et chercheurs dans les domaines de l'informatique scientifique

CONSEIL

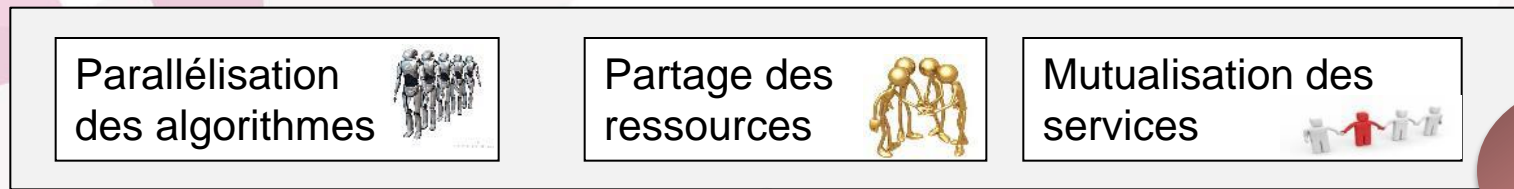
Fournir une expertise et un support de qualité aux chercheurs Inserm sur des thématiques techniques identifiées

PROJETS

Favoriser les collaborations au niveau national et international



Outils



VREs

Distribution

Exascale

Grilles

Clouds

HPC

BIG DATA

Catalogue de services



Mésocentres



14



Grille/Cloud à l'Inserm : état des lieux

- **La communauté française en sciences de la vie est très active sur la grille, mais l'Inserm est peu présent**
 - Moins de 10 équipes de recherche Inserm utilisent actuellement les ressources d'EGI
 - Utilisation à petite échelle (1-2 utilisateurs, souvent en thèse)
- **Et pourtant, l'Inserm...**
 - ... est un partenaire du GIS France Grilles
 - ... est impliqué dans plusieurs ESFRIs qui collaborent avec EGI
 - ... travaille sur des sujets qui sont de bons candidats pour la grille et le cloud
- **Il y a donc des améliorations possibles !**



Potentiels utilisateurs

- **Communautés d'imagerie médicale**
 - Les applications d'imagerie (simulation, analyse) fonctionnent bien sur la grille et le cloud
- **Communautés génomique/génétique/bioinfo**
 - NGS, analyse moléculaire, chemo-informatique, ...
- **E-santé et santé publique**
 - Applications distribuées, systèmes embarqués, objets connectés, bio-banques...



Stratégie de diffusion (1)

- **Identification des utilisateurs existants**
 - Construire des exemples
 - Identifier les *success stories*
 - Lister les problèmes et améliorations possibles
 - Organiser les bases pour un transfert de compétences efficace
- **Détection de potentiels nouveaux utilisateurs**
 - "pré-sélection" sur la discipline et le domaine
 - Analyse de besoin à travers notre action de cartographie
 - Identification de workflows semblables à ceux des équipes qui utilisent déjà la grille ou le cloud



Stratégie de diffusion (2)

- **Information**

- Organisation de présentations et séminaires
- Promotion des événements associés
- Mise en place d'outils de communication (site web, forums, documentation collaborative)

- **Formation et suivi**

- Organisation de formation avec France Grilles
- Mise en place de structure de conseil et de support
- Facilitation du transfert de compétence par la création de groupes de travail et de groupes d'intérêt



Premières actions concrètes

- **Identification des utilisateurs**
 - Action de cartographie en cours
- **Information**
 - Rédaction d'un guide « la grille pour les nuls »
 - Plusieurs séminaires donnés à l'automne 2014 dans des unités
 - Site web en ligne : <http://cisi.inserm.fr>
- **Formations**
 - Formation DIRAC en juin 2015 (12 participants Inserm)
 - Diffusion d'information sur les prochaines formations FG

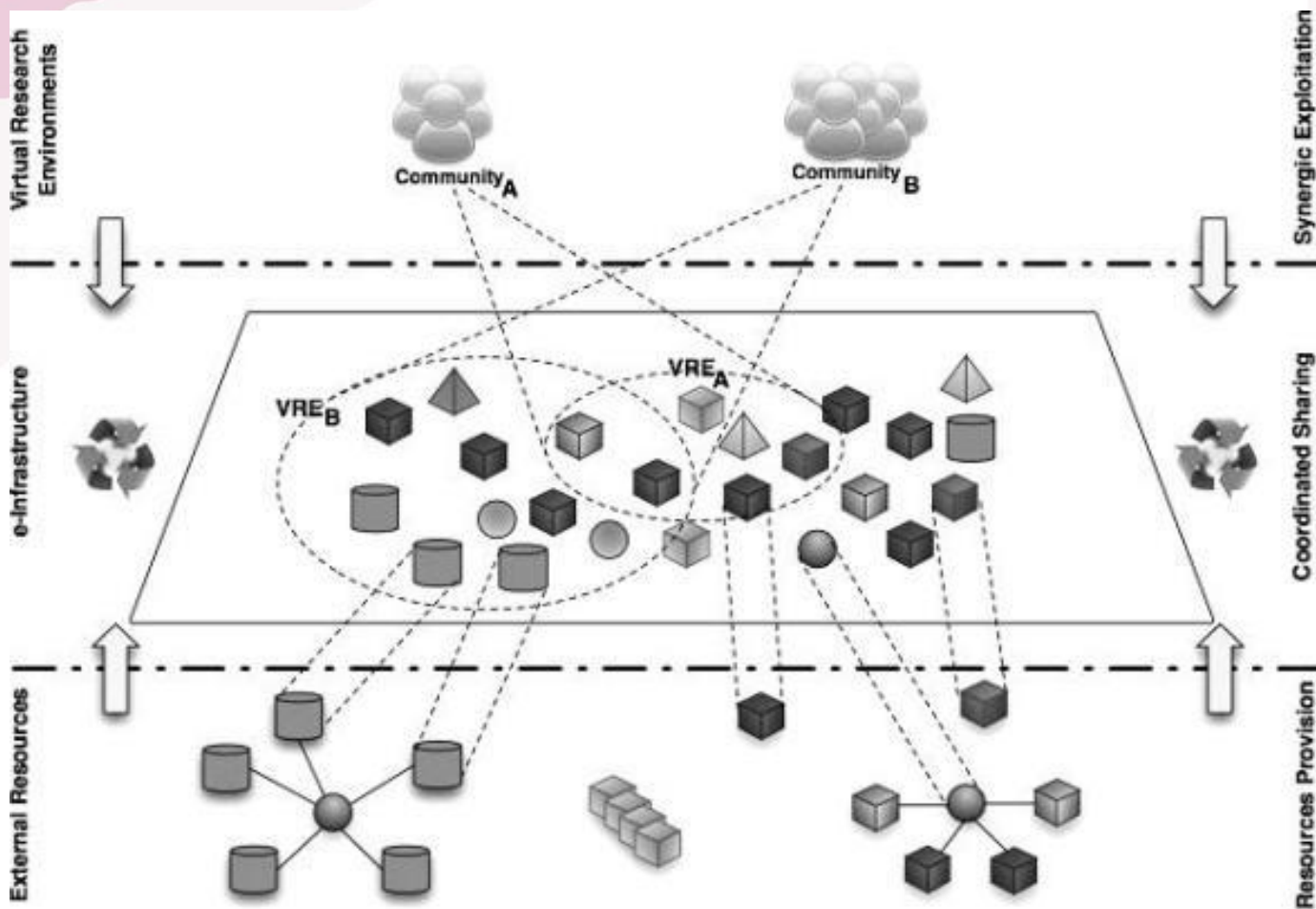


Ces actions ne sont pas spécifiques à la grille

- **Un engagement similaire est prévu pour les solutions HPC**
 - A travers les mésocentres
 - A travers les centres nationaux et GENCI
- **L'équipe CISI peut ainsi agir en intermédiaire**



Vision: des environnements de recherche virtuelle (VRE)



Vision: des environnements de recherche virtuelle (VRE)

- **Bénéfices bien identifiés**
 - Masquer l'hétérogénéité des infrastructures
 - Faciliter le travail collaboratif
 - Permettre un partage efficace des données et des ressources
 - Faciliter la vie des utilisateurs
- **Adaptés à notre cas**
 - Permet de créer des ponts entre technos (e.g. grid/HPC)
 - Fournit des points d'entrée uniques vers les infras distribuées
 - Facilite le changement culturel nécessaire
 - *"Scientific Computing with a human face"*



Vision: des environnements de recherche virtuelle (VRE)

- **Défis identifiés**

- Comment prendre en compte les spécificités Inserm, surtout relatives aux contraintes de confidentialité des données ?
- Nécessité d'une gestion intégrée des ressources sous-jacentes
- Associer de manière cohérente les besoins avec les ressources disponibles
- Besoin d'impliquer toute la pile, depuis les chercheurs jusqu'aux fournisseurs de solutions
- Nécessité d'une combinaison intelligente des approches *bottom-up* et *top-down*

- **C'est un grand projet à long terme**



Merci !

- **The CISI Team**

- Isabelle Perseil, Gilles Mathieu, Itebeddine Ghorbel, Gwénaél Dumont, Daniel Salas
- <http://cisi.inserm.fr>

- **Inserm**

- <http://www.inserm.fr>

- **France Grilles**

- <http://www.france-grilles.fr>

