



Enabling Grids for E-science

# Les besoins des applications

*C. Loomis (CNRS/LAL)*

*Atelier Grille au LAPP*

*Annecy, France*

*25-27 septembre 2007*

[www.eu-egee.org](http://www.eu-egee.org)



Information Society  
and Media



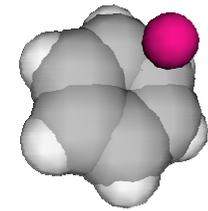
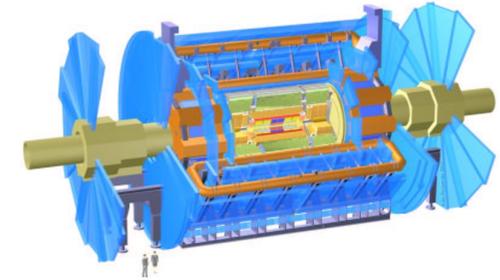
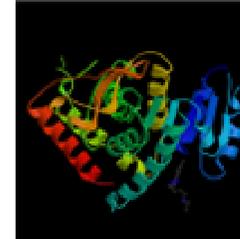
## Partage transparent de l'utilisation de ressources massivement distribuées par les utilisateurs des différentes disciplines...

“A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high computational capabilities.” (The Grid, I. Foster, C. Kesselman, 1998)

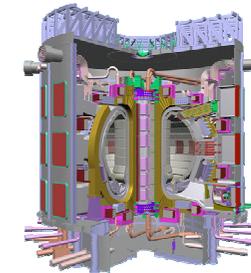
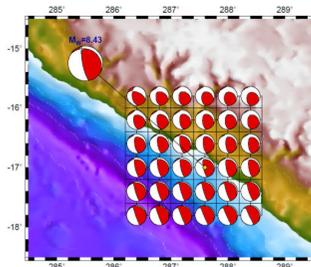
- **Introduction**
  - Qui utilise la grille?
  - Et pourquoi?
- **Fonctionnalités de base et avancées**
  - Exemples des applications
- **Conclusions**

- **La science et toujours l'équilibre entre la coopération et la compétition.**
- **Partage des ressources :**
  - Coopération : meilleure utilisation des ressources
  - Compétition : obtenir (et publier) vos résultats plus rapidement
- **Fédération des ressources (et données) :**
  - Coopération : utilisations des données plus variées
  - Compétition : produire les résultats plus précises
- **Collaboration**
  - Infrastructure peut mettre ensemble les gens avec des compétences différentes.
  - Moyenne pour publier, re-utiliser, et combiné les résultats précédents.

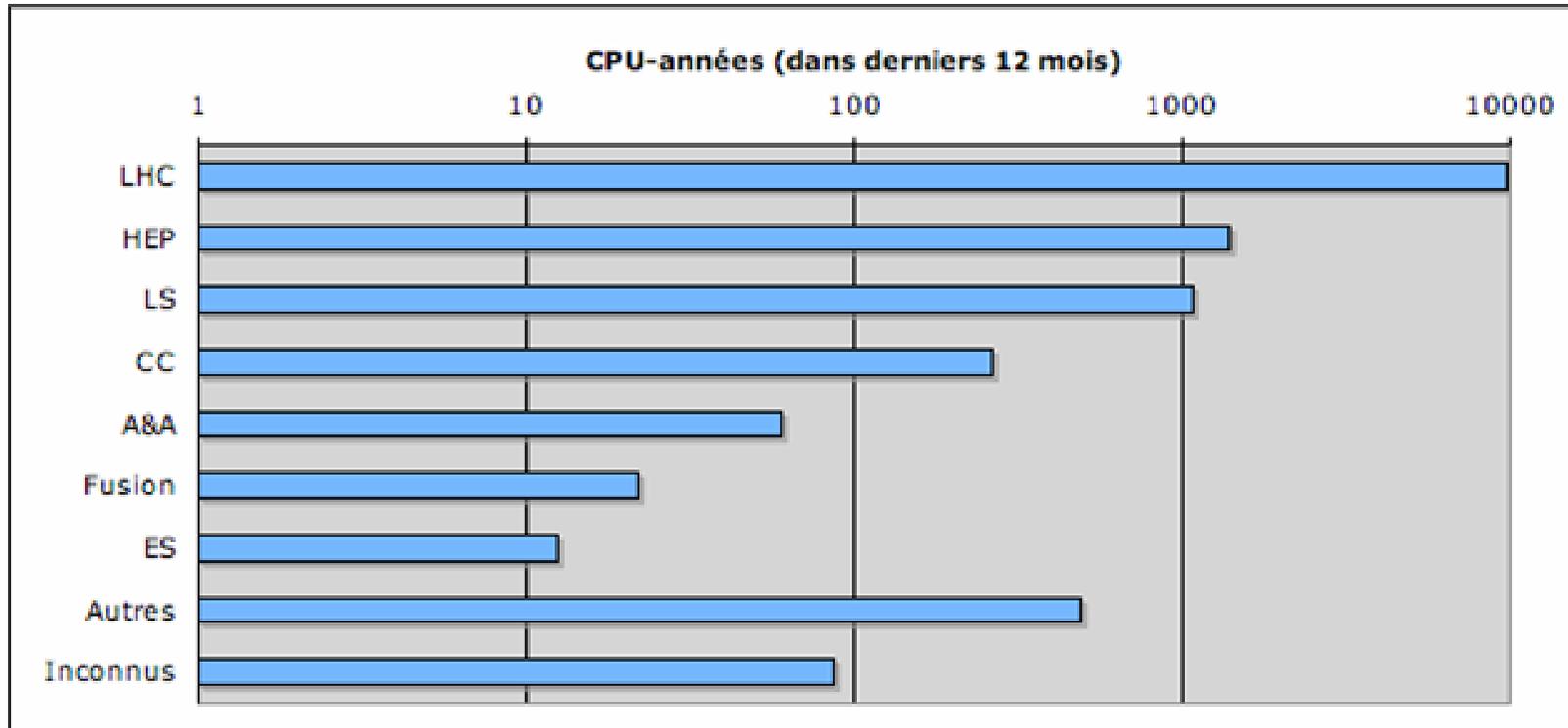
- Disciplines: 10
- Sub-disciplines: 36
- Agrandissement importante des applications et disciplines.
- Dans nos rapports  $\Rightarrow$  **Sous-estimé!**



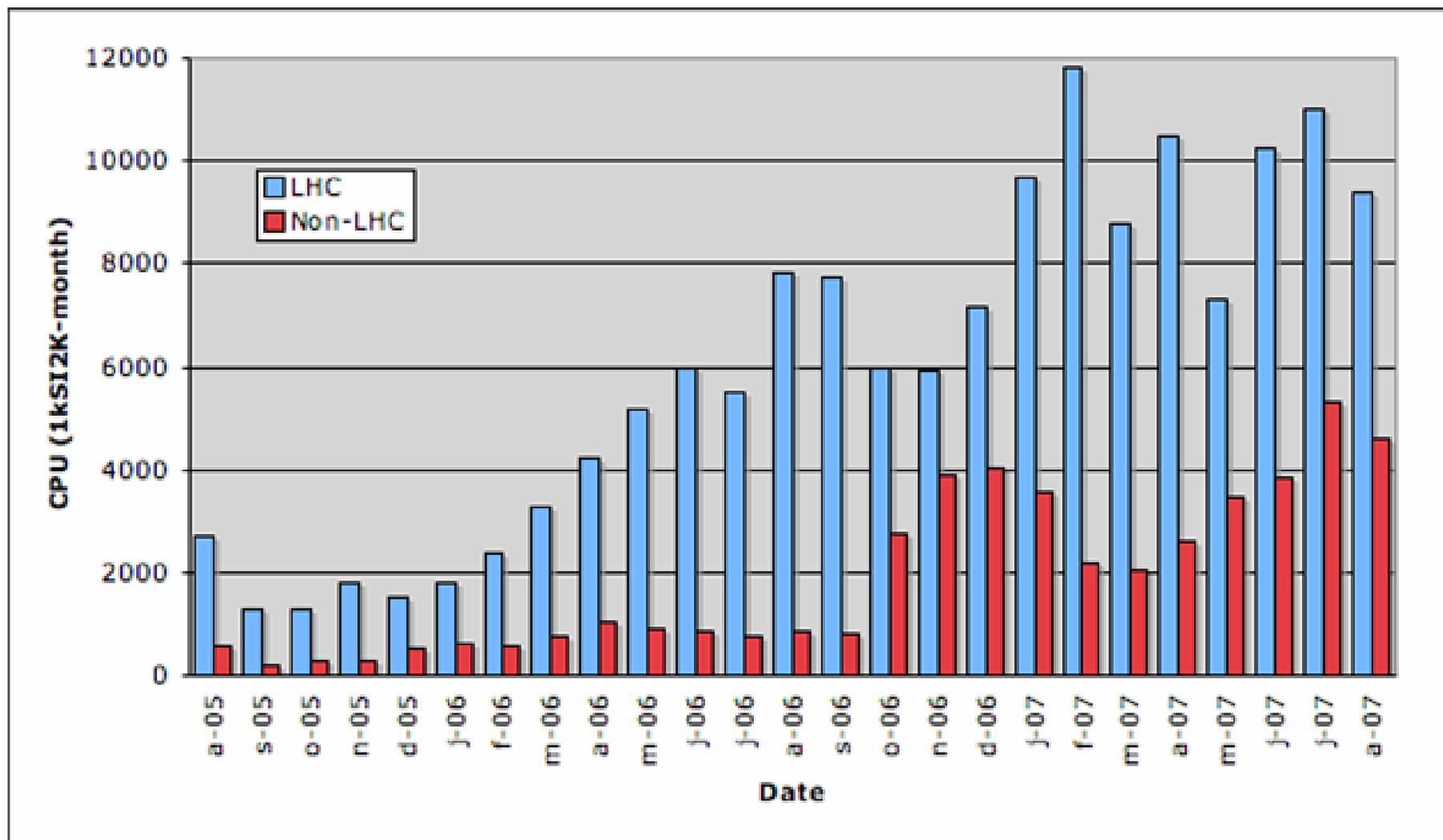
	6/2006	2/2007
Astronomie & astrophysique	2	8
Computational chemistry	6	27
Sciences de la terre	16	16
Fusion	2	3
Physique des hautes énergies	9	11
Sciences de la vie	23	39
Autres	4	14
<b>Totale</b>	<b>62</b>	<b>118</b>



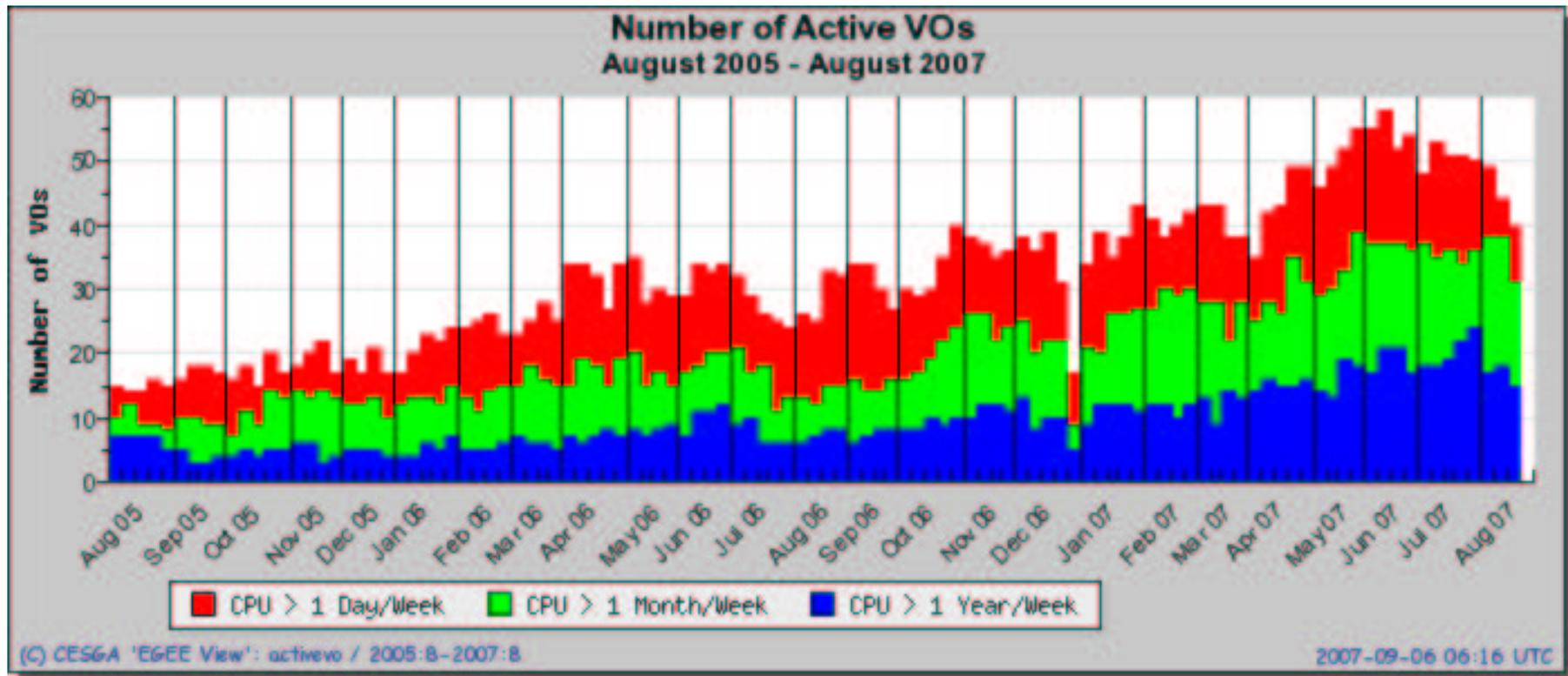
Condensed Matter Physics  
 Comp. Fluid Dynamics  
 Informatique  
 Protection Civile



- Utilisation récente : ~14000 CPUs.



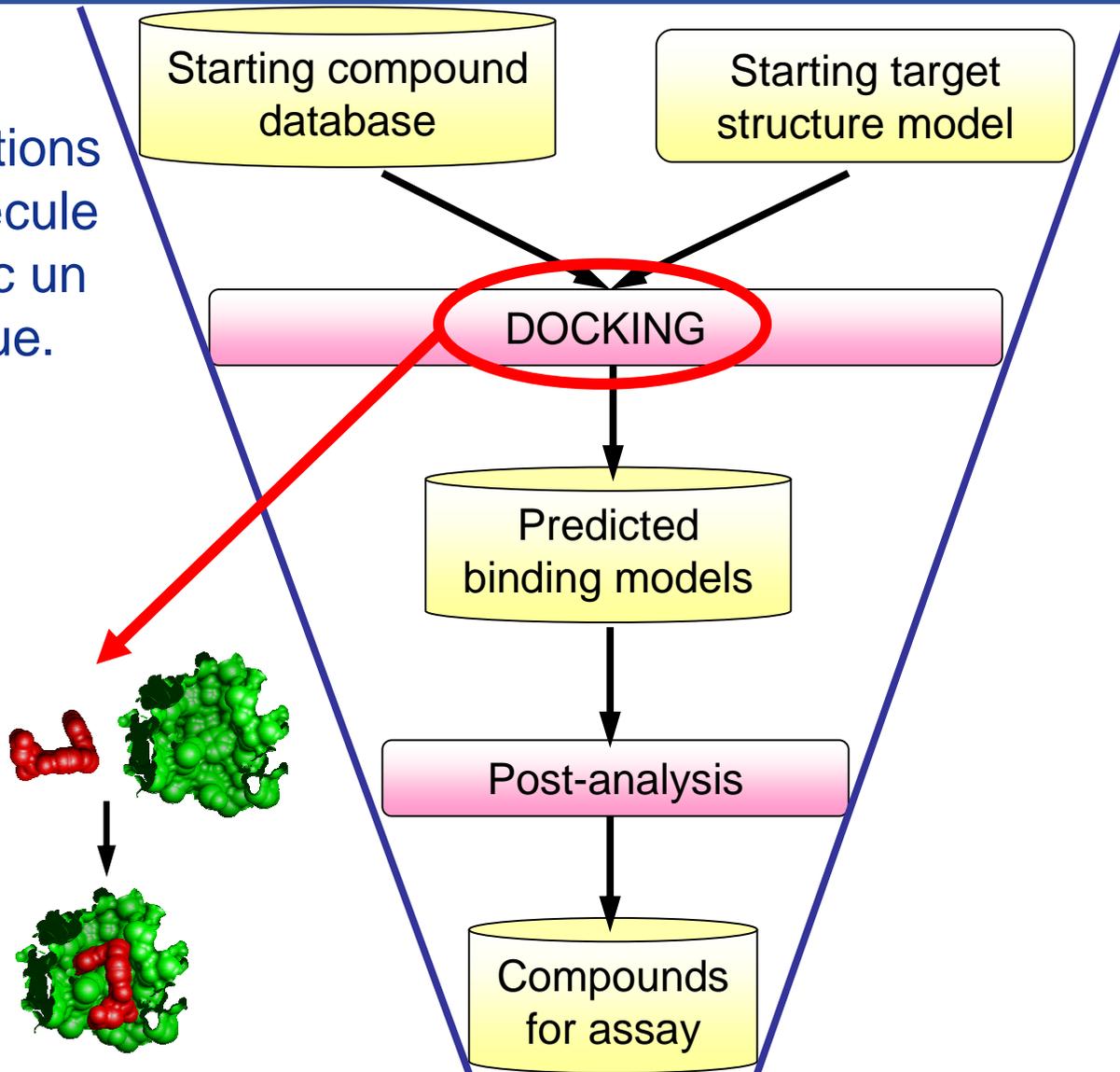
- **Nombre de VOs « active » agrandisse.**
  - Turnover (derniers 6/12/24 mois) : 84 / 93 / 102 VOs différentes
  - Total : 104 enregistrées, 258 visibles



- **Gestion des tâches (jobs) :**
  - Interface uniforme pour les systèmes de batch différents
  - Ordonnancement intelligente
  
- **Gestion des données :**
  - Transferts des données
  - Réplication de fichiers
  - Gestion des meta-données associées
  
- **Ces fonctionnalités sont fournis par le « middleware » gLite.**
  - Tous les VO's utilisent les services de base de gLite.
  - HEP demande une fiabilité et un « scalability » extrêmes.



- **Docking :**
  - Prédire les interactions entre un petit molécule et un protéine avec un structure 3D connue.
- **Projets :**
  - Proteins@Home
  - Rosetta@home
  - Docking@Home
  - AFRICA@home
  - malariacontrol.net
  - WISDOM





Coûts réduits pour les maladies délaissées  
R&D accéléré pour nouvelles maladies

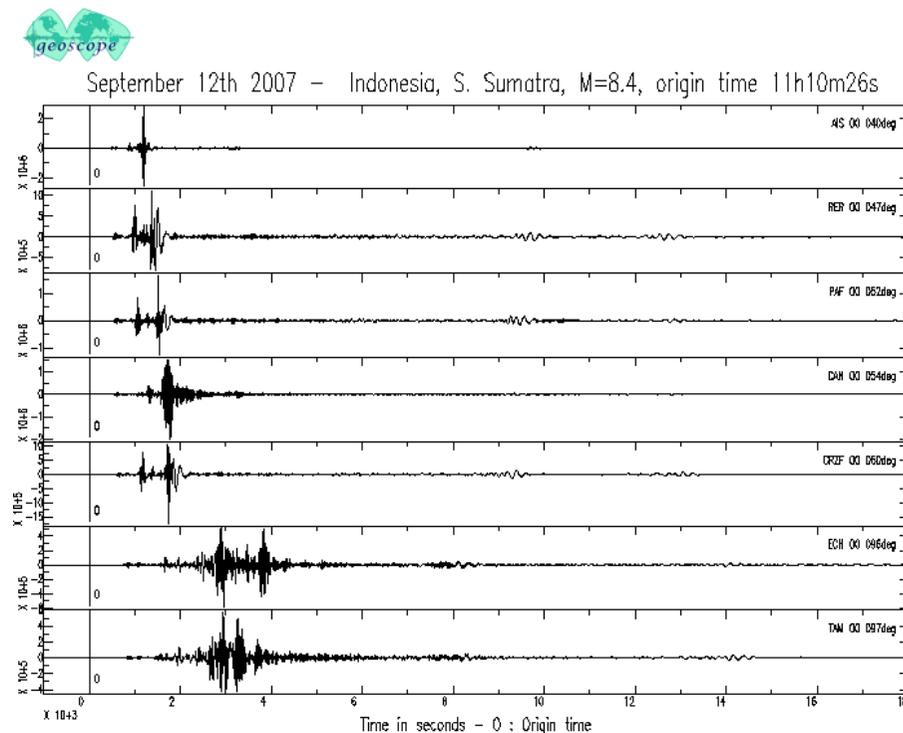
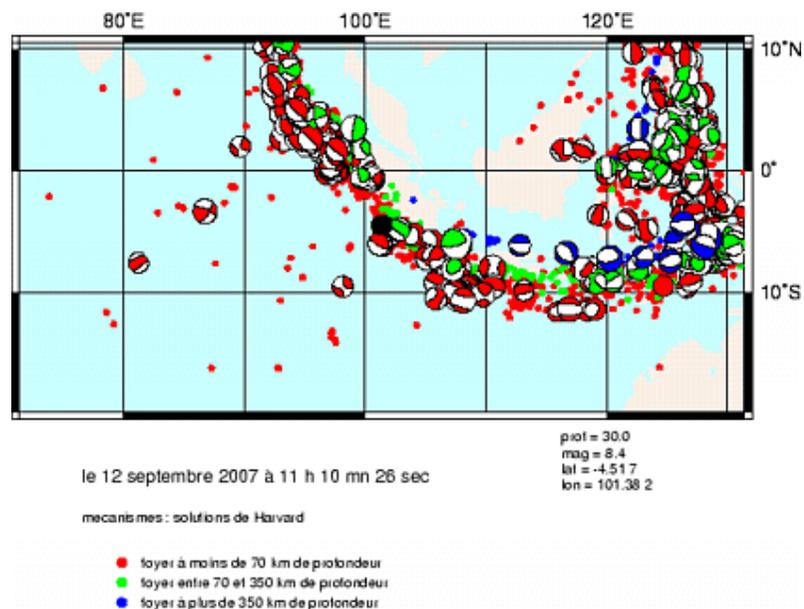
<http://wisdom.healthgrid.org/>

	Cibles	Molécules	CPU-années	Durée	CPUs max.	Résultats (TB)
WISDOM-I (Q3'05)	PBD	1M	80	6 sem.	1700	1
Avian Flu (Q2'06)	H5N1	300k	105	6 sem.	1700	0.750
WISDOM-II (Q4'06)	GST DHFR DHFR Tubulin	125M	240	8 sem.	5000	2

- **Ressources de calcul**
  - La grille a fourni plus de CPU que le projet peut acheter.
- **Ressources de stockage**
  - Mettre les ressources de stockage dans la grille.
  - Faire les sauvegarde permanents.
- **Outils**
  - Gérer millions de jobs.
  - Gérer les fichiers produits et les meta-données associes.
  - Les moyennes pour partager les résultats avec des autres.
- **Collaboration**
  - L'infrastructure grille a engendré une nouvelle collaboration humaine. Cette collaboration agrandisse et commence à traiter les autres maladies.

- **WISDOM-I: Molecular dynamics**
  - WISDOM va re-analyser le 5000 meilleur molécules avec des algorithmes « molecular dynamics ».
  - Il y a besoin des ressources parallèles : utilisation de MPI dans EGEE ou DEISA.
- **Grippe aviaire :**
  - Raffiner 5% des molécules avec les autres méthodes.
  - Analyser 50+ molécules dans le laboratoire (« wet lab assay »)
- **WISDOM-II:**
  - Faire les analyses suivantes avec la grille.

- **GeoScope : Analyse des tremblements de terre**
  - Algorithme complexe : Non-linéaire, 4D inversion avec données des plusieurs stations
  - Temps de calcul : 300-600 heures du CPU
- **But : Analyse complète dans quelques heures.**



- **GPSA (Grid Protein Sequence Analysis)**
  - <http://gpsa-pbil.ibcp.fr/>
- **Les utilisateurs de ce portail peuvent faire leur analyse de protéines facilement :**
  - Plusieurs algorithmes
  - Plusieurs bases des données
  - Sans certificat grille
- **La grille fournit une réponse plus rapide.**
- **Doit utiliser la configuration « SDJ » qui garantit une haute qualité de service pour les jobs GPSA. Les jobs exécutent immédiatement ou sont tués.**

- **Middleware du base : gLite**
  - Gestion des tâches avec l'ordonnancement intelligent.
  - Gestion des données : transferts, réplication des fichiers
- **Services/configurations supplémentaires :**
  - AMGA : gestion des meta-données pour les fichiers
  - MPI : configuration des sites qui permet les jobs parallèles
  - SDJ : ordonnancement avec une courte latence
  - GANGA/DIANE : outils pour gérer les grands calculs
- **La grille continue évoluer :**
  - Plus stable, plus « scalable », ...
  - Ajouter les fonctionnalités avec les services complémentaires

- **Les chercheurs utilise la grille à :**
  - Produire meilleurs résultats, plus rapidement
  - Partager les ressources informatiques
  - Collaborer efficacement avec les autres
- **Le « middleware » continue évoluer pour mieux servir ses utilisateurs.**
  - Le base est plus stable, plus « scalable », ...
  - Plus des services complémentaires disponible
- **La meilleur utilisation de la grille est comme infrastructure collaboratif : produire, partager, et re-utiliser les résultats facilement.**