



Enabling Grids for E-scienceE

EGEE et son avenir

G. Vetois (CGGVeritas)

*3^{ème} rencontre EGEODE
Villefranche-sur-Mer
28-29 Mai 2009*

www.eu-egee.org



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

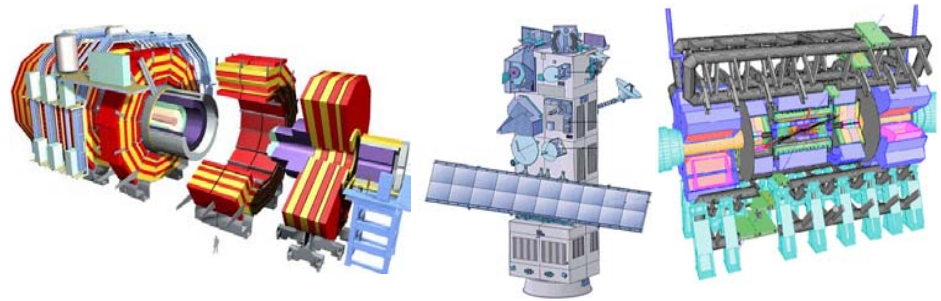
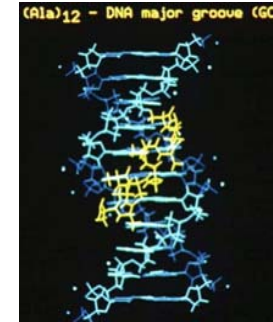
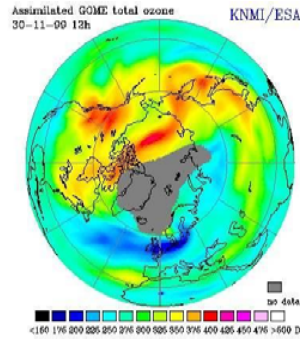
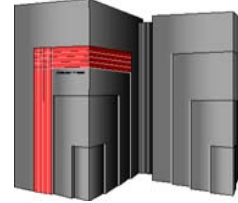
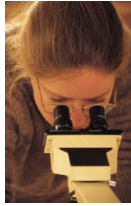
“A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high computational capabilities.” (The Grid, I. Foster, C. Kesselman, 1998)

- **Qu'est que c'est une grille?**
- **Le projet EGEE**
- **Qui l'utilise et pourquoi?**
- **Fonctionnalités de gLite**
- **L'avenir de la grille**
- **Conclusions**

- **La grille peut fédérer des ressources distribuées.**
 - CPU, stockage, algorithmes, expertises
- **Accès sécurisé entre des domaines administratives différentes.**
 - Confiance entre les administrateurs de système, développeurs des applications, utilisateurs, ...
- **Accès transparente à « toutes » les ressources.**
 - Par des portes d'entrée : User Interface
 - Les aspects de calcul distribué sont cachés.

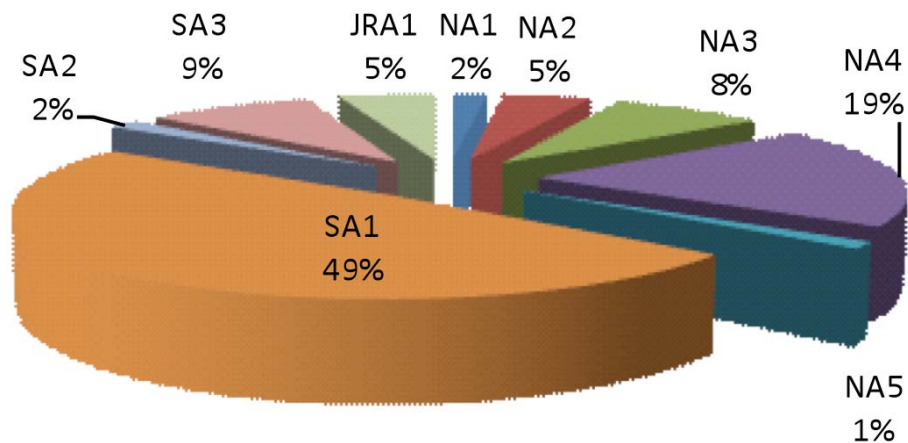
- **Ressources partagées/fédérées**
 - Meilleure utilisation des ressources
 - Obtenir (et publier) vos résultats plus rapidement
 - Utilisations des données plus variées
 - Produire les résultats plus précises
- **La grille utilise les interfaces standardisées :**
 - Coût de développement diminué
 - Plusieurs implémentations des services grilles
- **La grille peut fournir les services de haut niveau :**
 - Éviter votre propre développement des services
 - Permettre de faire une analyse plus efficacement

Grid ‘Middleware’



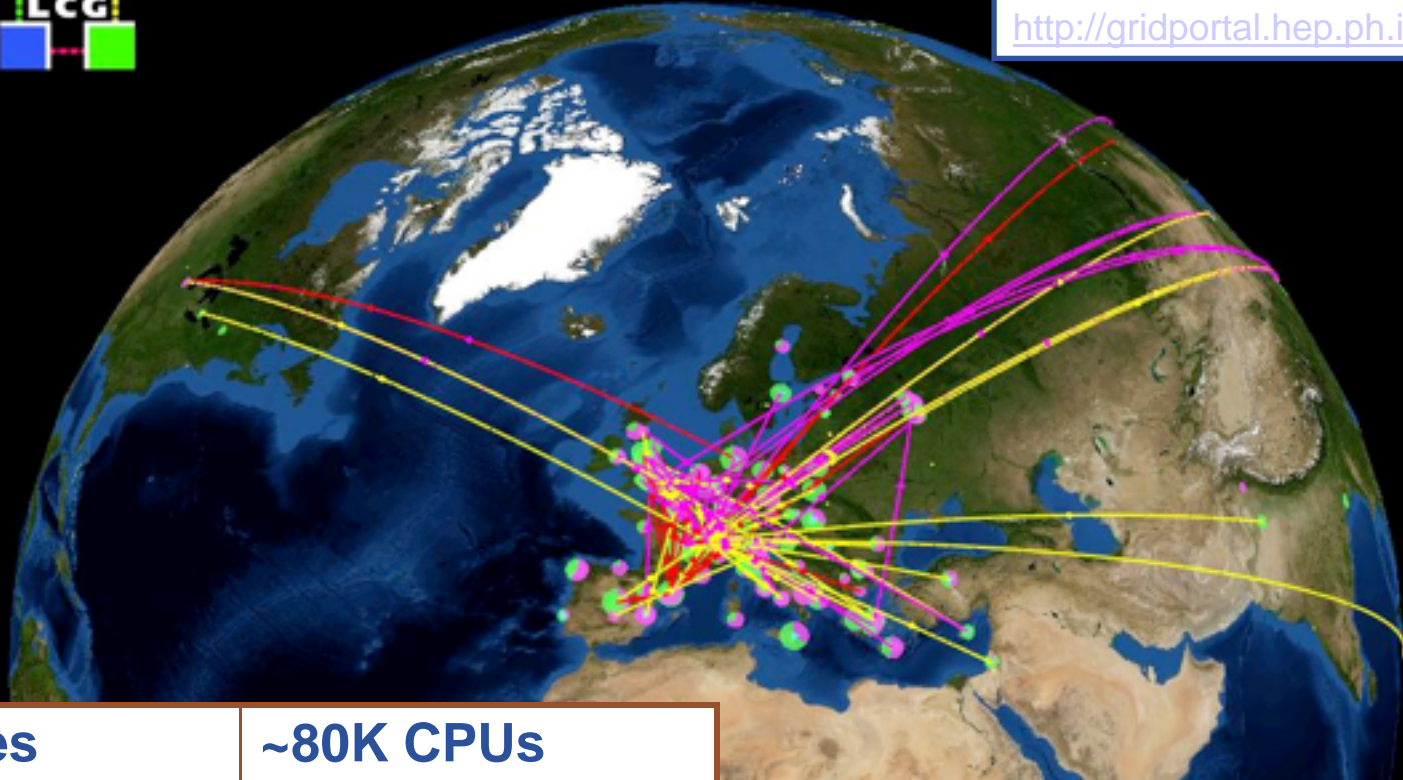
- **Grille des supercalculateurs**
 - DEISA (<http://www.deisa.org/>)
 - Cher, procédure de allocation lourde, bien adapté aux applications fortement parallèle
- **Grille institutionnelle**
 - EGEE (<http://www.eu-egee.org/>)
 - Moins cher, allocation des ressources fait par les propriétaires, peut accueillir des applications très variées
- **Grille des « desktop »**
 - BOINC (<http://boinc.berkeley.edu/>)
 - EDGeS (<http://www.edges-grid.eu/>)
 - Beaucoup moins cher, ressources des bénévoles (« cycle sharing »), applications avec les transferts des données minimaux

- **Une série des projets européens (2001 - présent)**
 - European DataGrid (EDG)
 - Enabling Grids for E-Science (EGEE)
 - EGEE-II (terminé fin avril 2008, ~36 M€)
 - EGEE-III (jusqu'a mai 2010, ~32 M€)
- **Buts :**
 - Créer une grille européenne pour la communauté scientifique
 - Aider les gens qui veulent intégrer leurs ressources dans la grille
 - Aider les gens qui veulent utiliser la grille
 - Planifier la transition au EGI





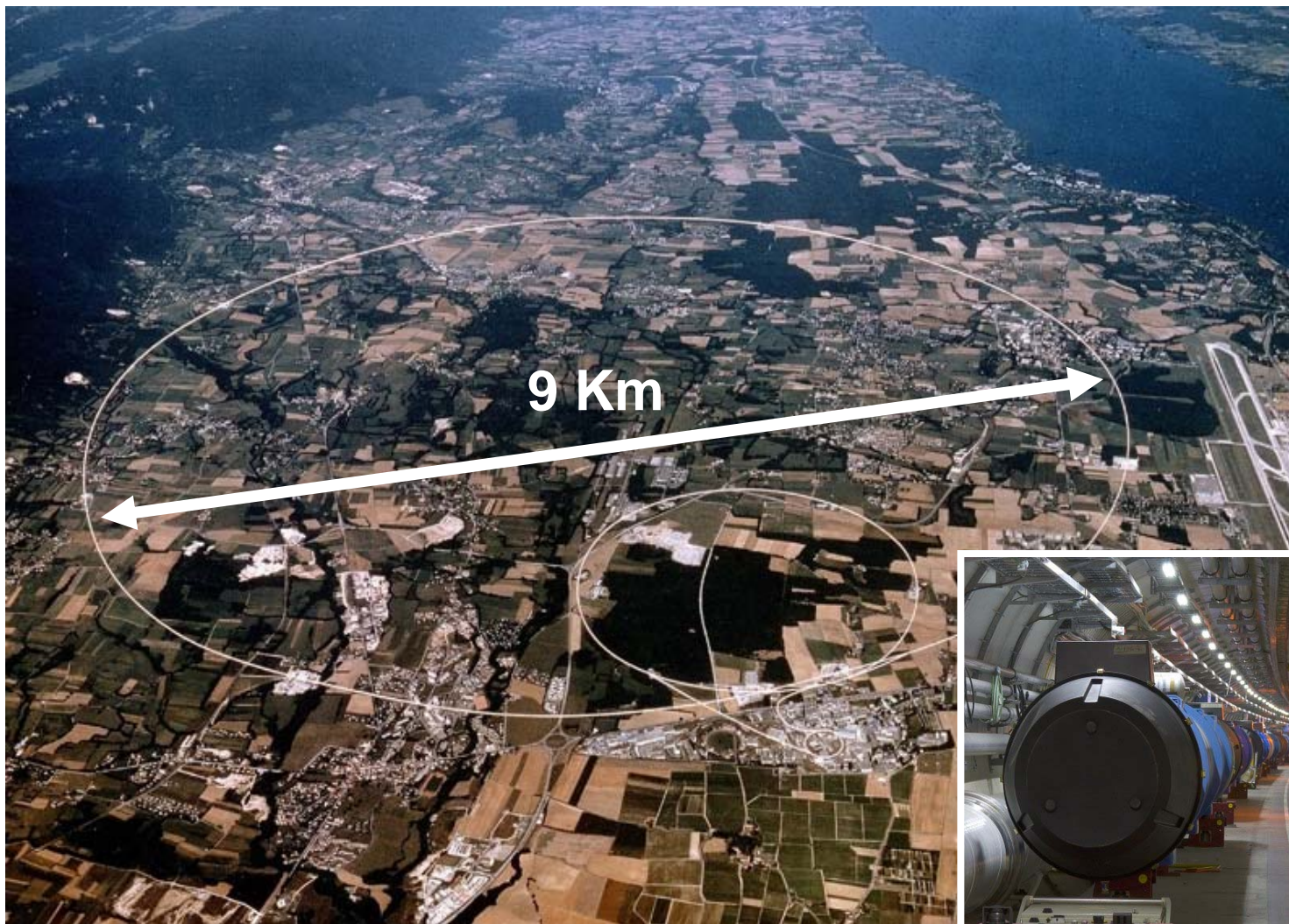
Real Time Monitor
<http://gridportal.hep.ph.ic.ac.uk/rtm/>



| | |
|----------------------|---------------------------|
| 280 sites | ~80K CPUs |
| 45 countries | > 20 PB disk |
| 13000 users | > 250K jobs/day |
| 110 (200) VOs | |

**Infrastructure de production
 → 24h/24h Operationelle !**

Large Hadron Collider (LHC)

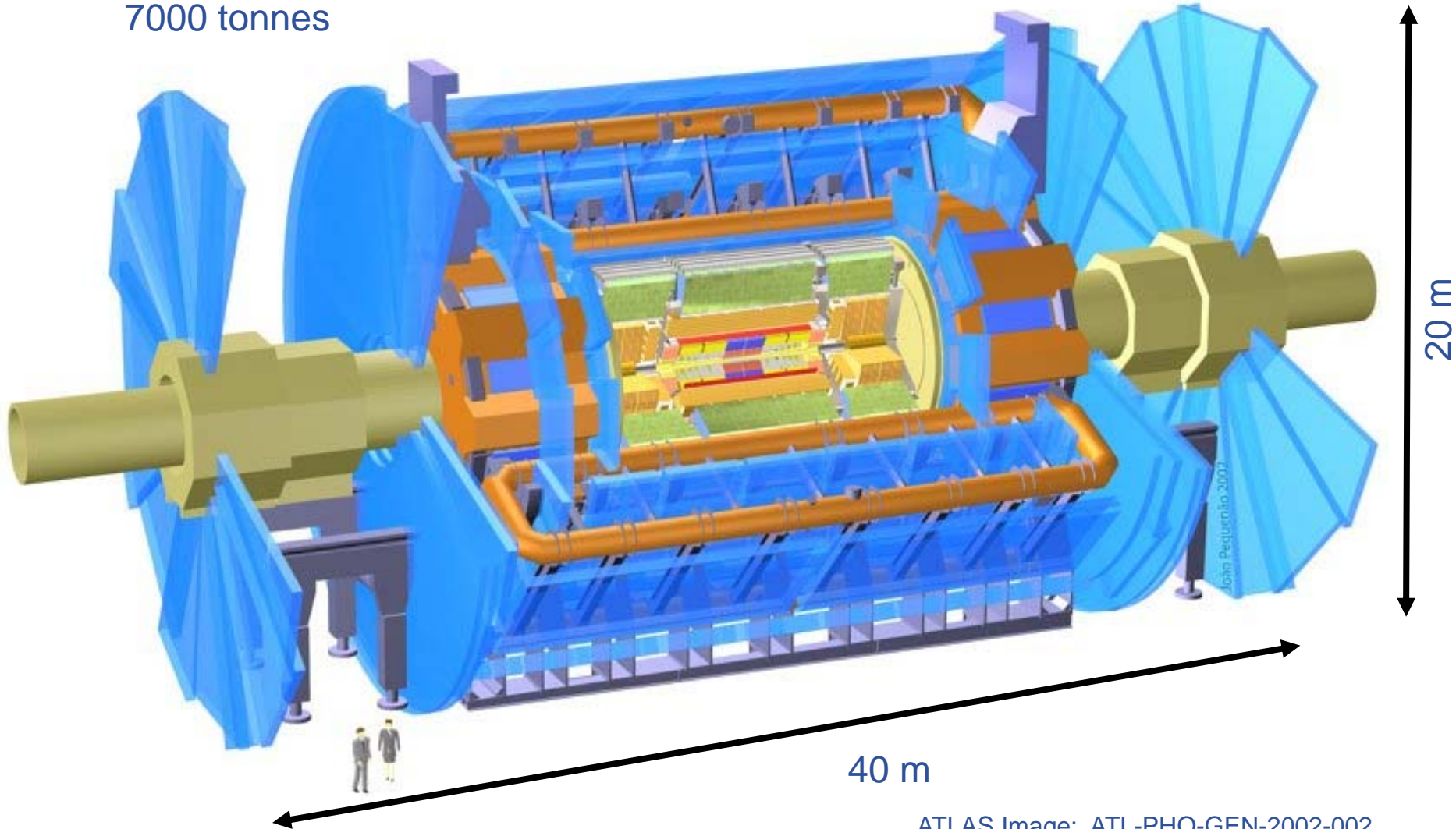


© CERN Geneva



© CERN Geneva

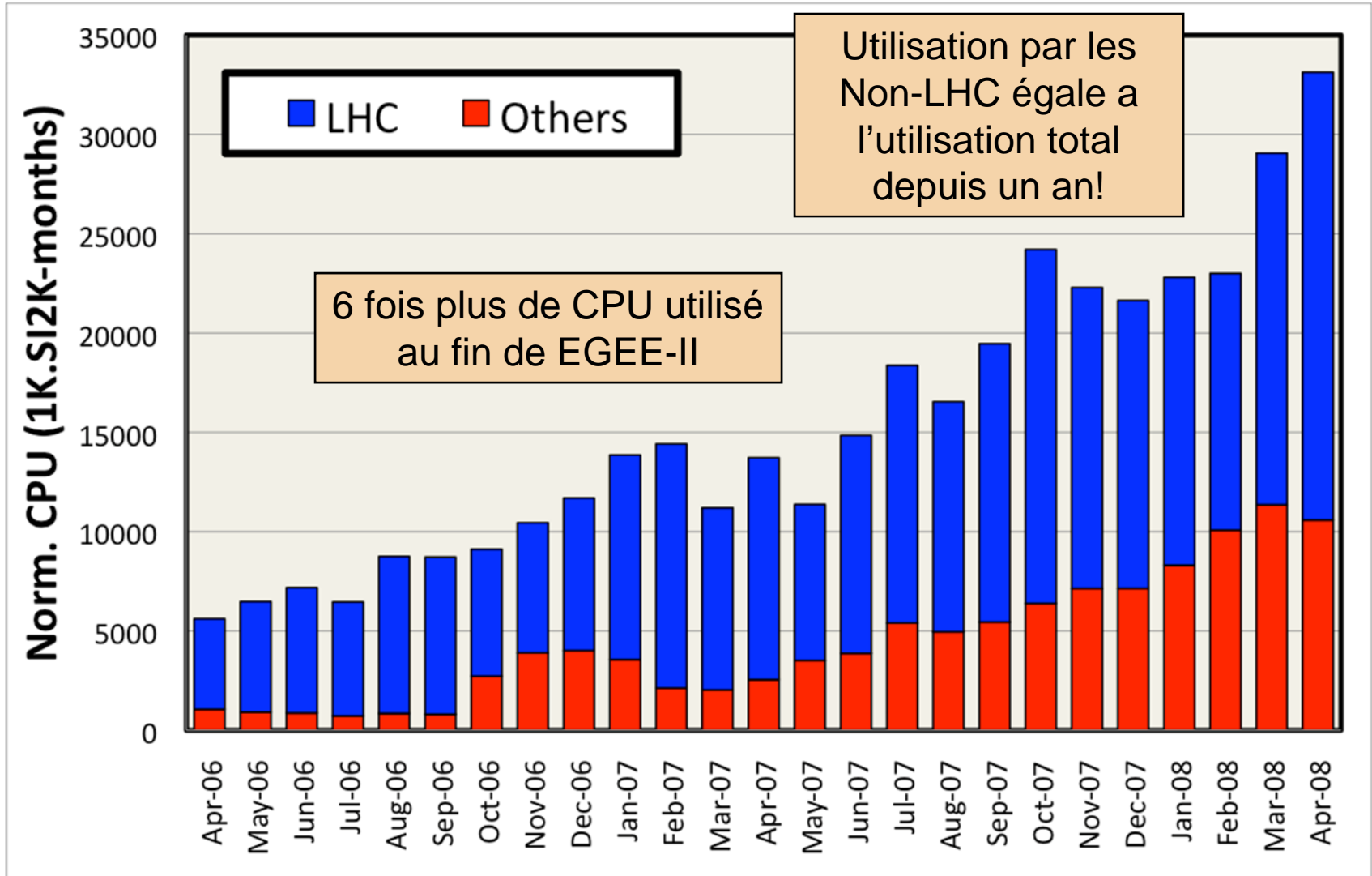
7000 tonnes



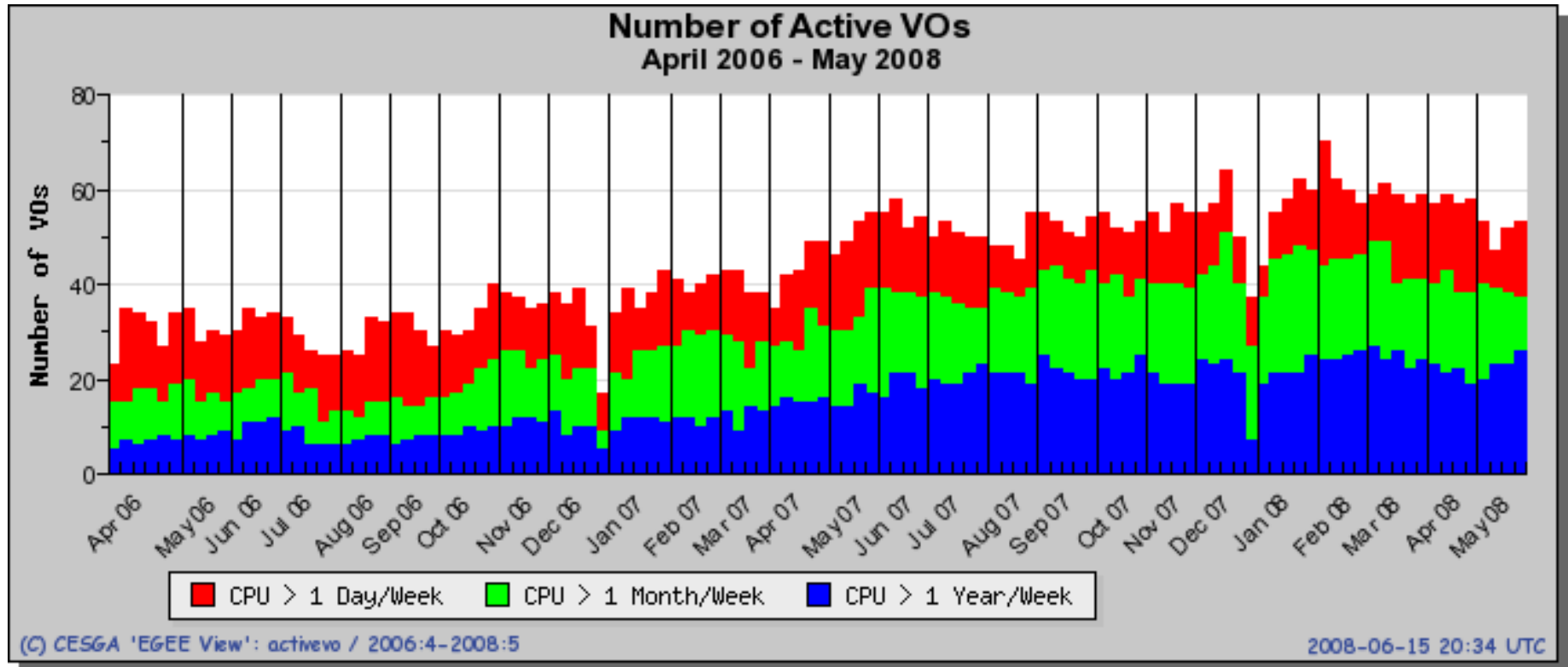
ATLAS Image: ATL-PHO-GEN-2002-002

- **Taux des données :**
 - 40 MHz pour les interactions
 - 100 Hz des événements filtrés
 - 1-10 Mo per chaque événement
 - 0.1-1 Go/second
- **Volume des données :**
 - LHC : 24h/24, 7j/7 (septembre 2008)
 - LHC génère 15 Po des données chaque année !
 - Durée de vie : 15-20 ans
 - Données simulées plus grande
- **Challenge : gestion des données**
 - Enregistrement et lecture des données
 - Gestion des métadonnées
 - Supporte pour les analyses chaotique

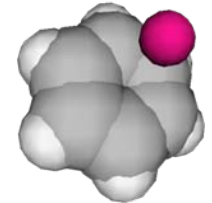
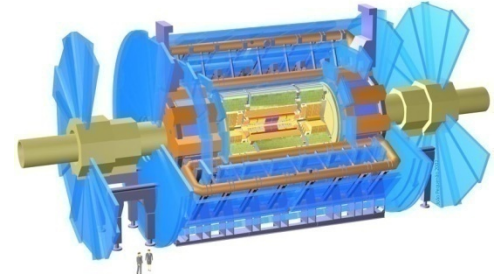
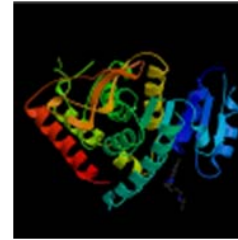
| | | |
|-------|---|------------------|
| kilo- | K | 10 ³ |
| mega- | M | 10 ⁶ |
| giga- | G | 10 ⁹ |
| tera- | T | 10 ¹² |
| peta- | P | 10 ¹⁵ |
| exa- | E | 10 ¹⁸ |



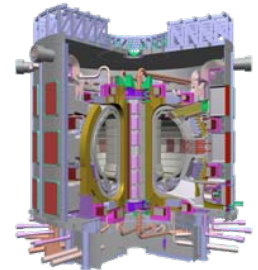
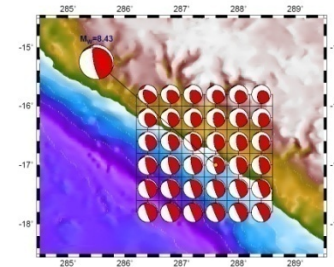
- **Nombre des VOs « actives » agrandisse!**
 - Très souvent pas les mêmes VOs chaque semaines
 - Total VOs : 130 enregistrées, 258 visibles



- Agrandissement importante des application et disciplines
- EGEE est une infrastructure ouverte.



| | 6/2006 | 2/2007 | 1/2008 |
|------------------------|-----------|------------|------------|
| Astron. & Astrophysics | 2 | 8 | 9 |
| Comp. Chemistry | 6 | 27 | 21 |
| Earth Science | 16 | 16 | 18 |
| Fusion | 2 | 3 | 4 |
| High-Energy Physics | 9 | 11 | 7 |
| Life Sciences | 23 | 39 | 37 |
| Others | 4 | 14 | 21 |
| Total | 62 | 118 | 117 |



Condensed Matter Physics
 Comp. Fluid Dynamics
 Computer Science/Tools
 Civil Protection

- **WISDOM** (<http://wisdom.healthgrid.org/>)
 - Développer les nouveaux médicaments *in silico*
 - Coûts réduits
 - Développement accélérée
 - Malaria, grippe aviaire, ...
- **Trois grosse calculs :**
 - WISDOM-I (Q3'05), 1M molécules, 1 To
 - Avian Flu (Q2'06), 300k molécules, 750 Go
 - WISDOM-II (Q4'06), 125M molécules, 2 To
- **Les calculs WISDOM utilisent FlexX de BioSolveIT (3-6k licences gratuites) et Autodock.**

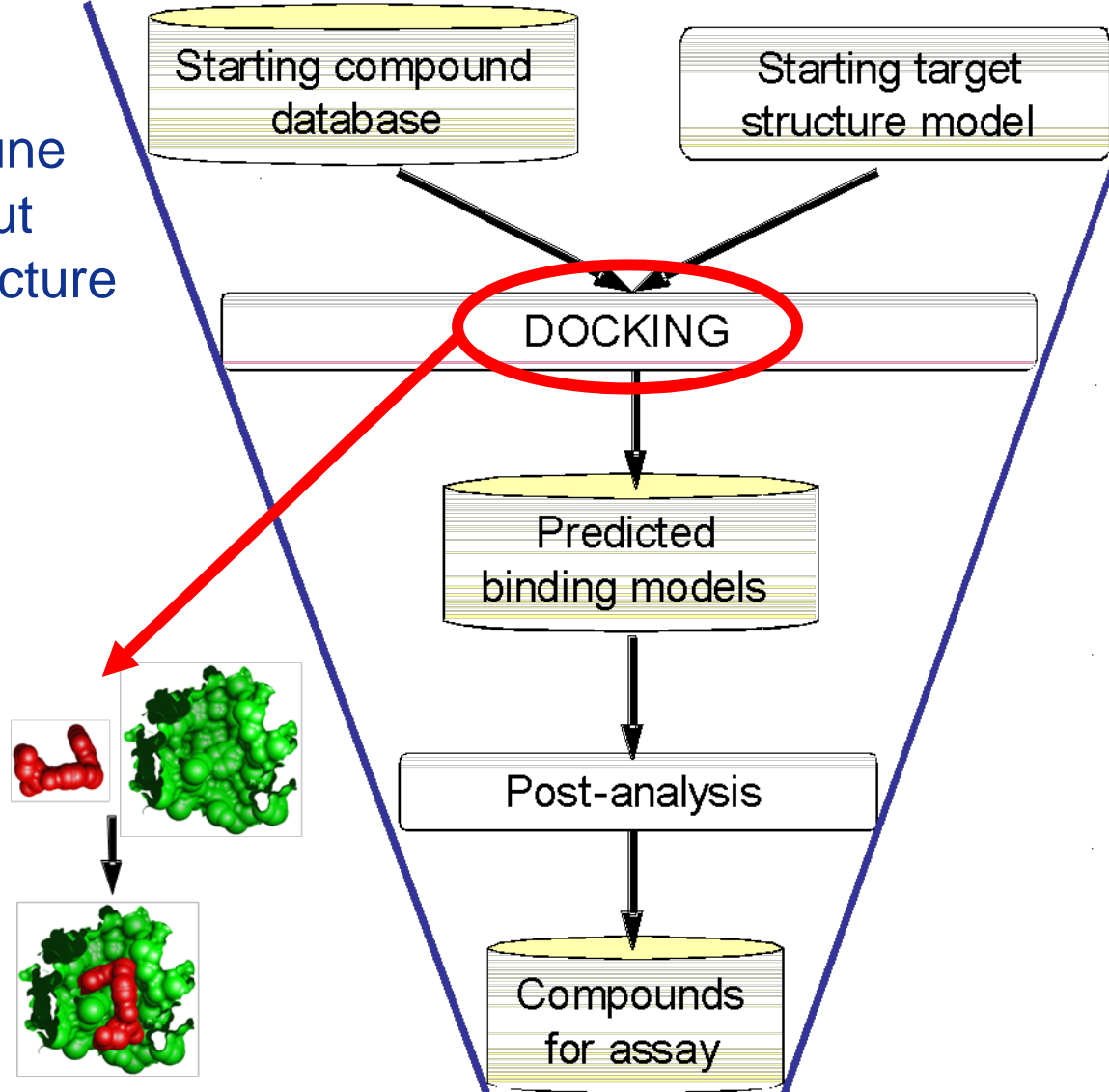


- **Docking**

- Predire comment une petite molécule peut attacher d'une structure 3D

- **Projets**

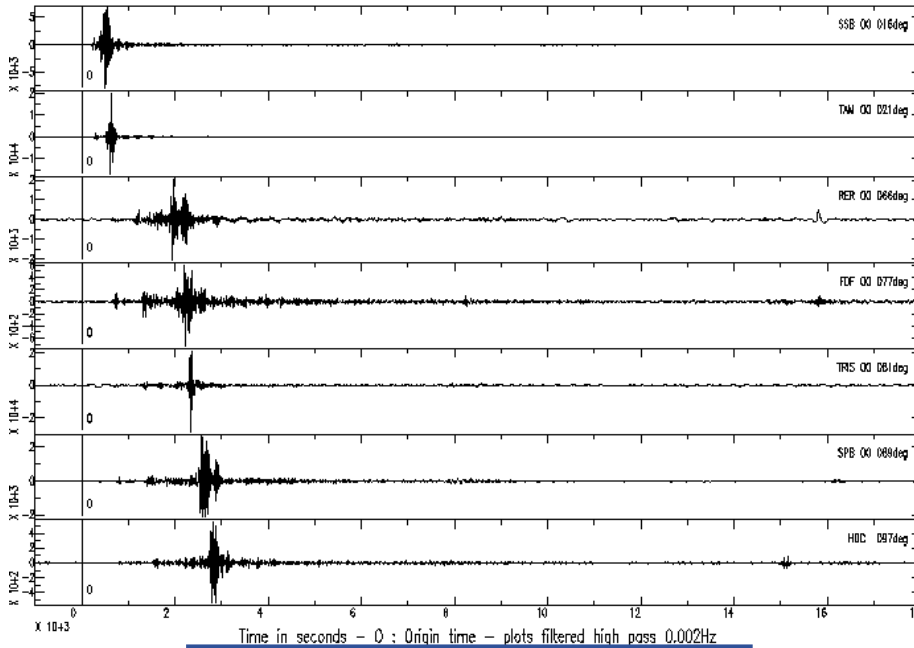
- Proteins@Home
- Rosetta@home
- Docking@Home
- AFRICA@home
- malariacontrol.net
- WISDOM



- **GEOSCOPE** (<http://geoscope.ipgp.jussieu.fr/>)
 - Analyses rapides des tremblements de terre

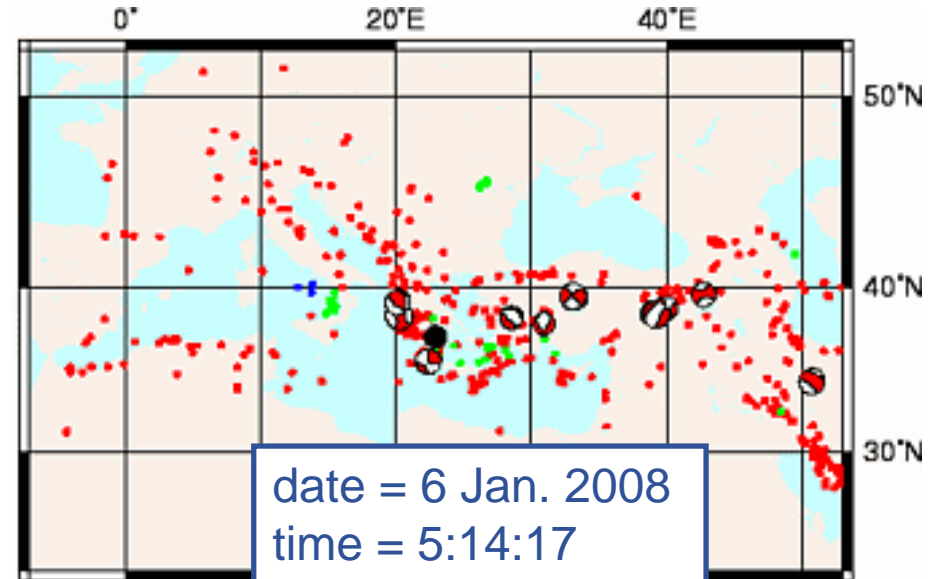


January 6th 2008 – Southern Greece, M=6.1, origin time 05h14m17s



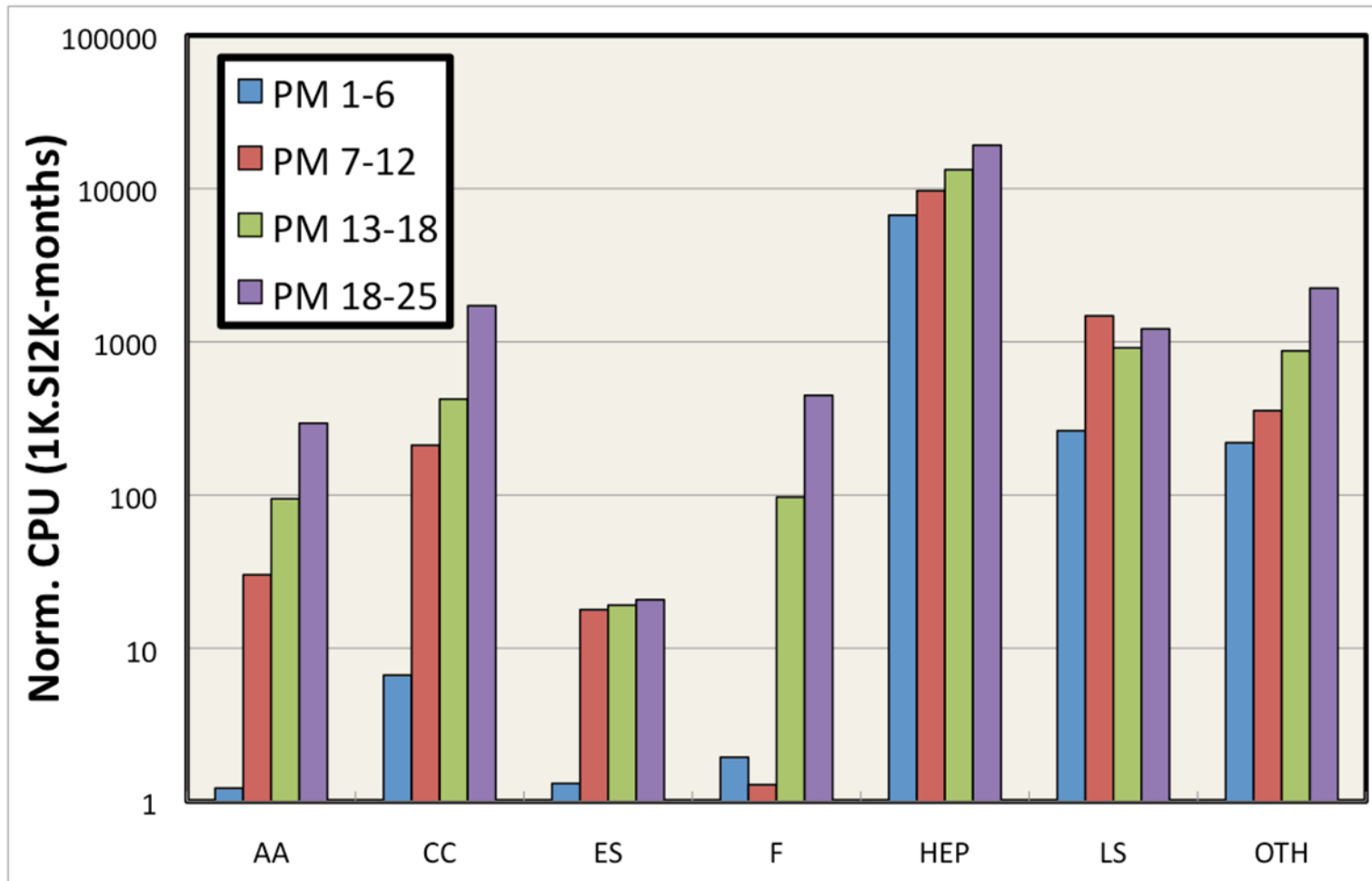
Nouveau calcul:

- Détermination du bruit
- 25 années de données
- 28 stations sismiques



date = 6 Jan. 2008
 time = 5:14:17
 depth = 50.9 km
 magnitude = 6.1
 latitude = 37.150°
 longitude = 22.934°

- Utilisation totale très différente entre les disciplines
- Technologie grille adopté par beaucoup des domaines

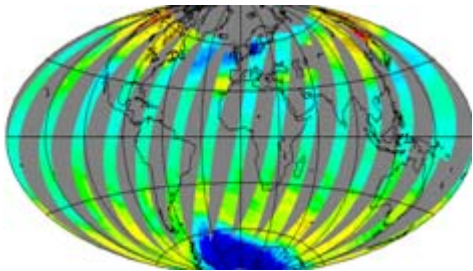
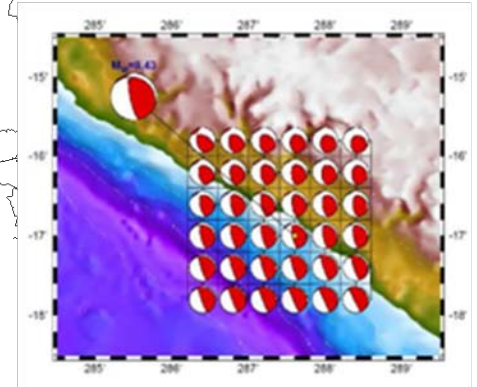
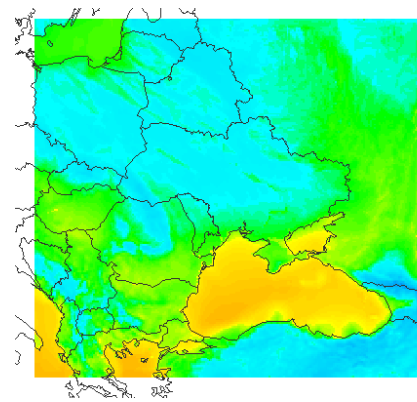
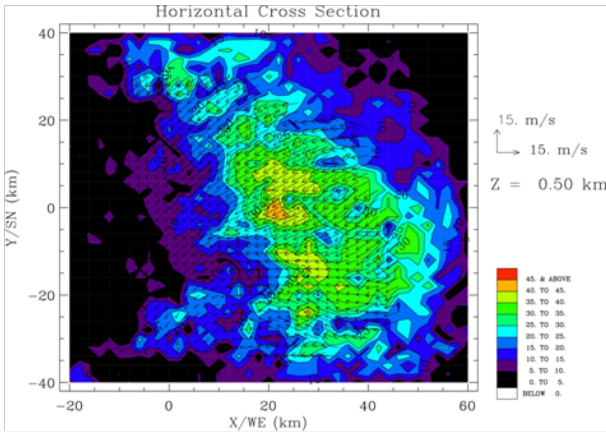


- Les laboratoires décident qui peut utiliser leur ressources.

- Les sites du EGEE supportent bien les disciplines variées
 - Les sites souvent déploient plus d'un CE ou SE.
 - Nombre (*pas* taille) des ressources!

- Les sites typiquement mettent à disposition leurs ressources pour les analyses bien justifiées

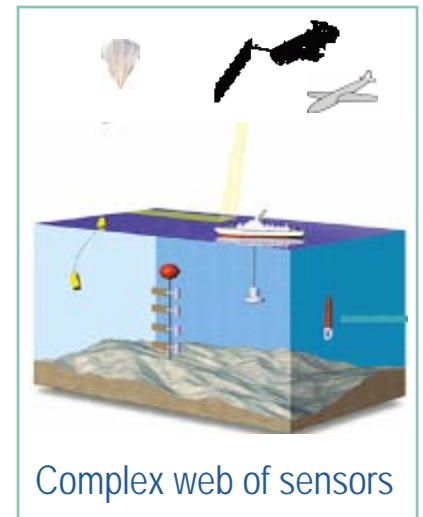
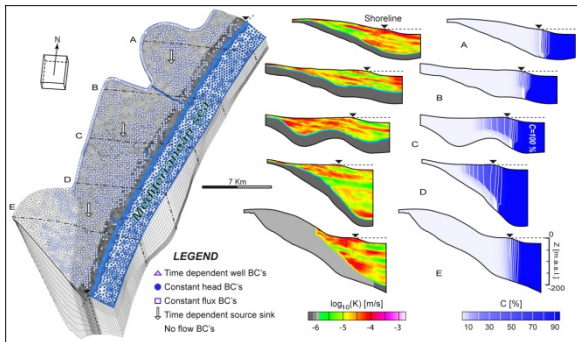
| | # CEs | # SEs |
|--------------|------------|------------|
| HEP | 292 | 299 |
| LS | 113 | 123 |
| CC | 25 | 41 |
| AA | 57 | 83 |
| Fusion | 19 | 21 |
| ES | 42 | 65 |
| Others | 143 | 149 |
| Unknown | 288 | 327 |
| Infra. | 282 | 306 |
| Total | 366 | 334 |



Large variety of ES disciplines
(ESR,EGEODE,... VOs)

<http://www.eu-degree.eu/>

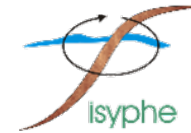
<http://www.euearthsciencegrid.org/>



- **EGEODE VO**

- 6 CNRS laboratories

- ISTEP - Laboratoire de Tectonique Paris6
- UMR – GéoAzur
- UPMC - Sisyphe
- IPG/EOST - Strasbourg
- IPGP Jussieu Paris
- ENS Géologie Paris



- UNIL – Université de Lausanne



- ~30 users

- 2000+ Geocluster batch jobs in 2008

- 200 Gb of seismic data

<http://www.egeode.org>

- **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
- 

<http://glite.web.cern.ch>
- **Le « middleware » gLite fournisse ces fonctionnalités**
 - Tous les VOs utilisent les services de base de gLite.
 - HEP demande une fiabilité et un « scalability » extrêmes.

APPLICATION-LEVEL

- Web Portal
- Pilot Job Frameworks
- ...
- Interactivity
- Database Access
- Job Management

HIGH-LEVEL

- Replica Catalog
- File Transfer Service
- Meta-scheduler
- Encrypted Data Tools
- Workflow
- Metadata Catalog

RESOURCE

- Computing
- Storage

FOUNDATION

- Registry
- Security Infrastructure

- **Simulation : batch, pas de gestion de donnée**
- **Analyse de données : batch + gestion de données**
- **Pseudo-interactif : temps de réponse court**
- **Workflow : enchainement de tâches complexes**
- **Applications parallèles : MPI**
- **Utilisation croissante d'applications commerciales**
- **...**

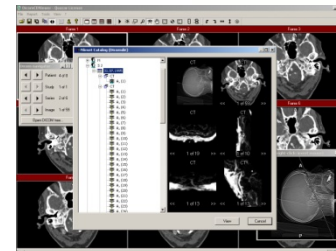
- **Listes référencés des applications sur la grille :**
 - <http://appdb.eu-egee.org/>
 - http://grid.ct.infn.it/egee_applications/

- **gLite fournisse les fonctionnalités du base.**
- **Plusieurs analyses requissent les services de haut niveau aussi.**

- **Développement au niveau des applications :**
 - NA4 : HEP, LS, ES, CC, Fusion, A&A, GO
 - RESPECT : « middleware » externe
 - § http://egeena4.lal.in2p3.fr/index.php?option=com_content&task=view&id=71&Itemid=63

- **Collaborations avec les industriels :**
 - Gaussian, Turbomole
 - MathWorks (MATLAB)
 - Interactive Supercomputing

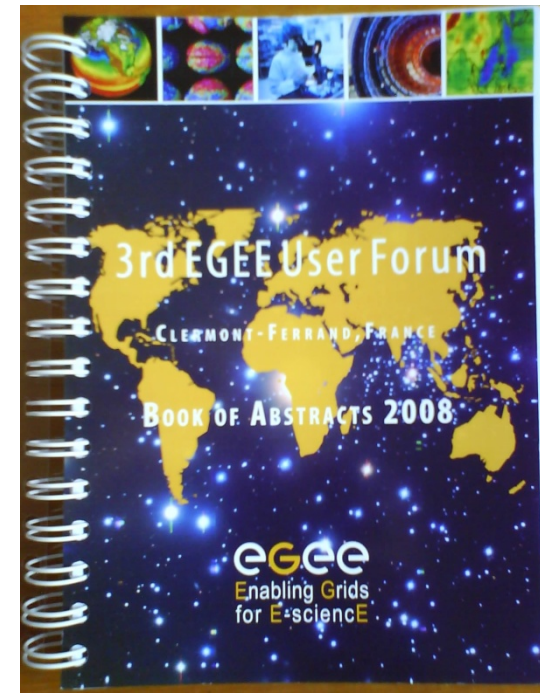
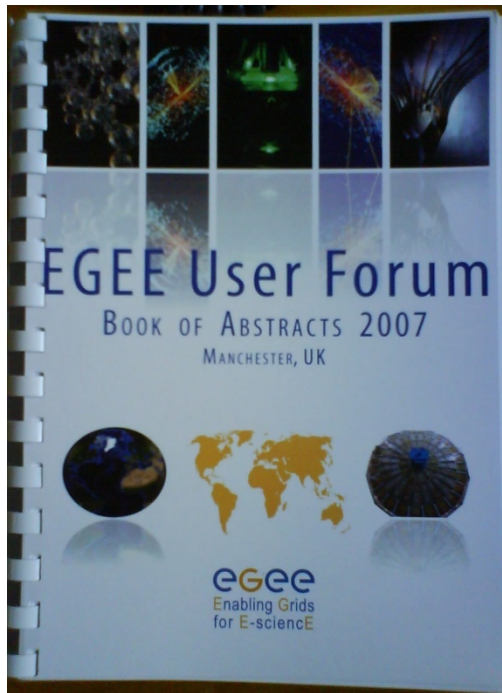
- **AMGA: Metadata catalog.**
- **Ganga: Job submission framework.**
- **DIANE: Master/slave task manager.**
- **Dashboard: VO and user-level monitoring.**
- **MOTEUR: Workflow engine.**
- **MDM: Medical Data Management**



- **Rec. External Software Pkgs. for the EGEE Community**
 - Identifier des applications tiers fonctionnant correctement avec gLite
 - Disséminer ses informations pour éviter de réinventer la roue
 - http://egeena4.lal.in2p3.fr/index.php?option=com_content&task=view&id=71&Itemid=63
- **Les softs dans la list :**
 - GridWay: Grid metascheduler.
 - Ganga: Job submission framework.
 - DIANE: Master/slave task manager.
 - i2glogin: Interactive login to grid nodes.
 - GRelC: Database access and management.
- **On discute avec int.eu.grid pour ajouter leur produits dans RESPECT.**

- **Garantir l'utilisation quotidienne et efficace.**
- **Support aux utilisateurs**
 - Documentation
 - GGUS = Global Grid User Support
 - <http://ggus.org>
- **Support pour les organisations virtuelles**
 - Déclaration des VO
 - Création et vérification des outils pour les « VO managers »
- **Aide aux portages d'applications pour la grille**
 - GASuC = Grid Application Support Centre
 - <http://www.lpds.sztaki.hu/gasuc/>

- Programme scientifique très riche et variée



<http://indico.cern.ch/conferenceDisplay.py?confId=7247>

<http://indico.cern.ch/conferenceDisplay.py?confId=22351>

- **EGEE'09**
21-25 Sep 2009 : Spain
<http://egee09.eu-egee.org/>

- **UF4/OGF25**
<http://egee-uf4.eu-egee.org>
 - 2- 6 mars 2009 : Catania, Italie
 - <http://indico.cern.ch/conferenceTime>

- **EGEE'08**
 - 22-26 septembre 2008 : Istanbul, Turquie
 - <http://egee08.eu-egee.org/>

EGEE '09 CONFERENCE
 21-25 September 2009
 Hotel Barceló Sants - Barcelona, Spain
www.egee09.eu-egee.org

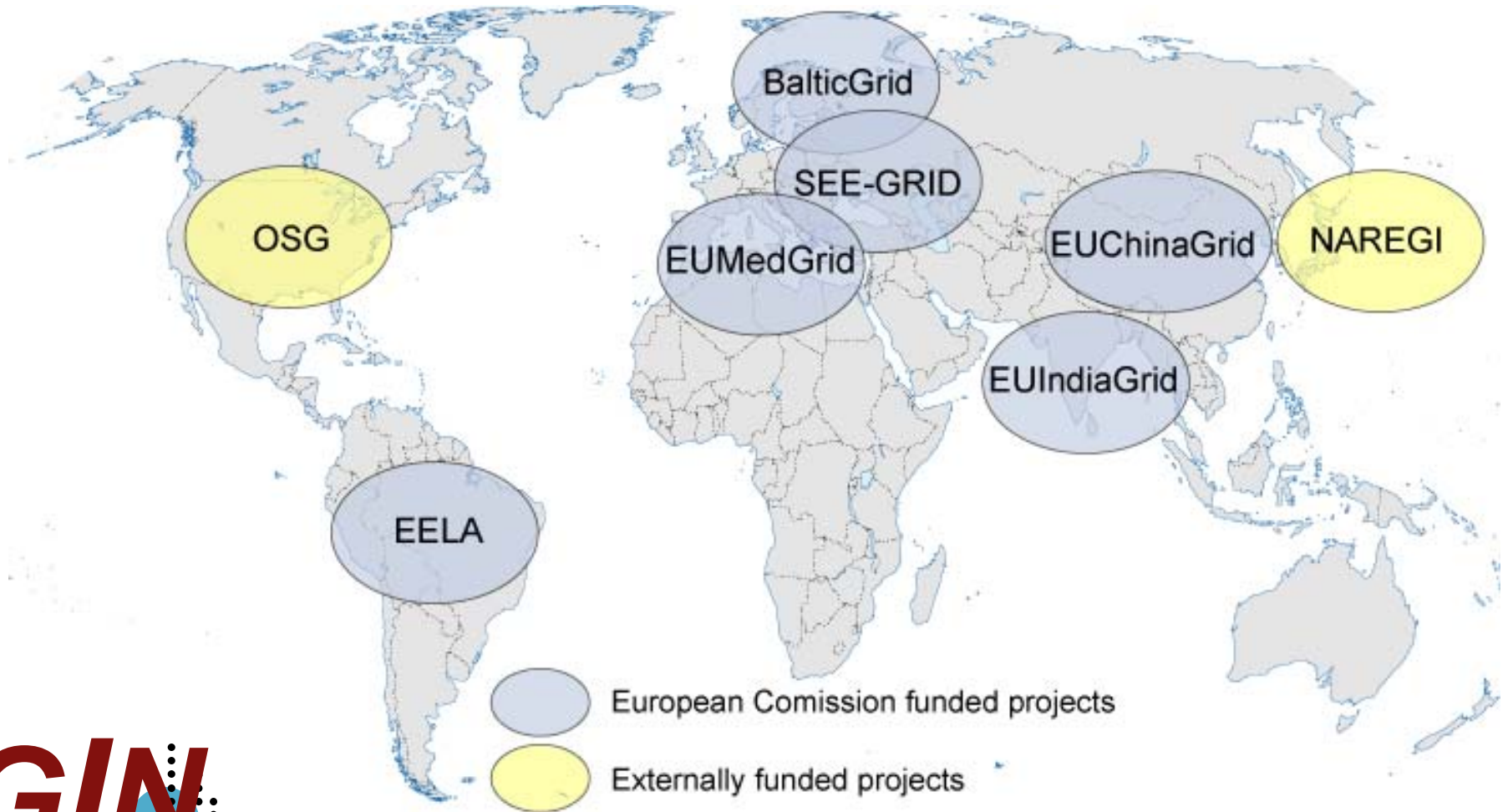


*Don't miss the last
 Conference of the EGEE project!*

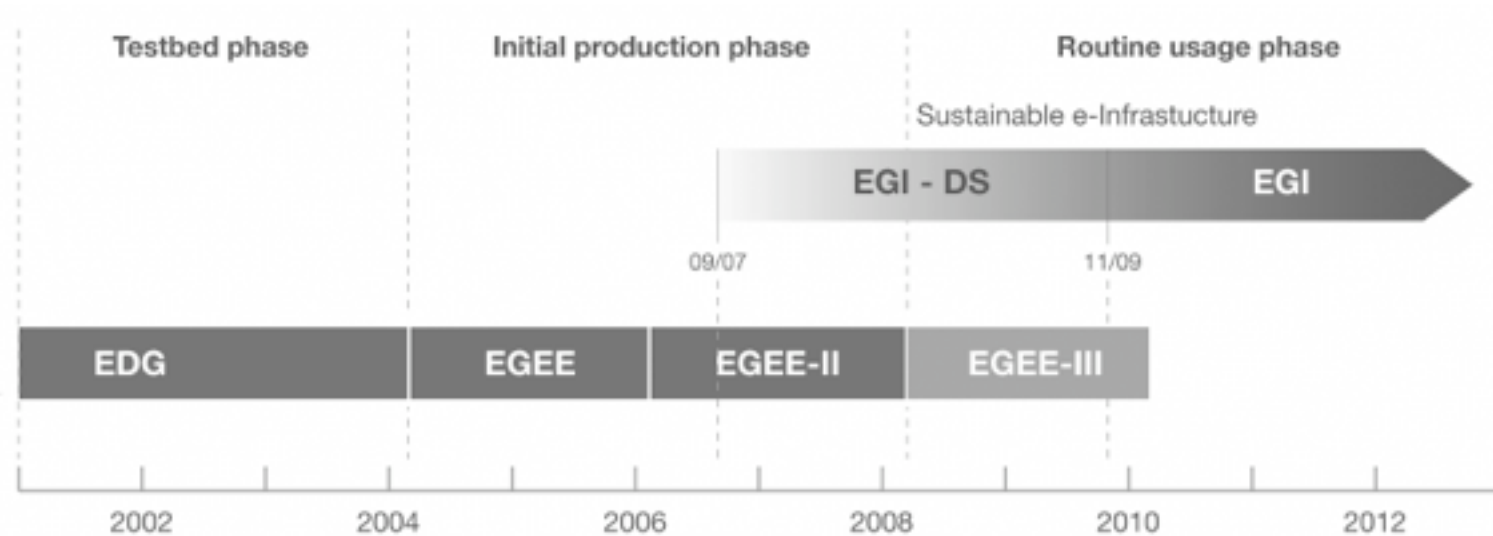
EGEE is a project co-funded by the European Commission, NFSO-RI 222667



Registration and Information
Mondial
 Mondial & Cititravel Congresses, S.L.
 Rosselló, 303 4ºico L - 08037 Barcelona, Spain
 T. +34 932 212 955 - Fax +34 934 592 059
egee09@mondial-congress.com



- **Le projet doit :**
 - Trouver une financement pérenne.
 - Adopter une modèle comme fournisseur des services.
 - § NA4 : supporte, « community building », fonctionnalités avancés.
- **En France :**
 - L’Institut des Grilles (<http://www.idgrilles.fr/>)
 - « Joint Research Unit » pour EGEE



- **EGI.org**

- Small central coordination group.
- Central office will be located in Amsterdam.



- **Resource Providers**

- Federation of National Grid Infrastructures (NGIs).

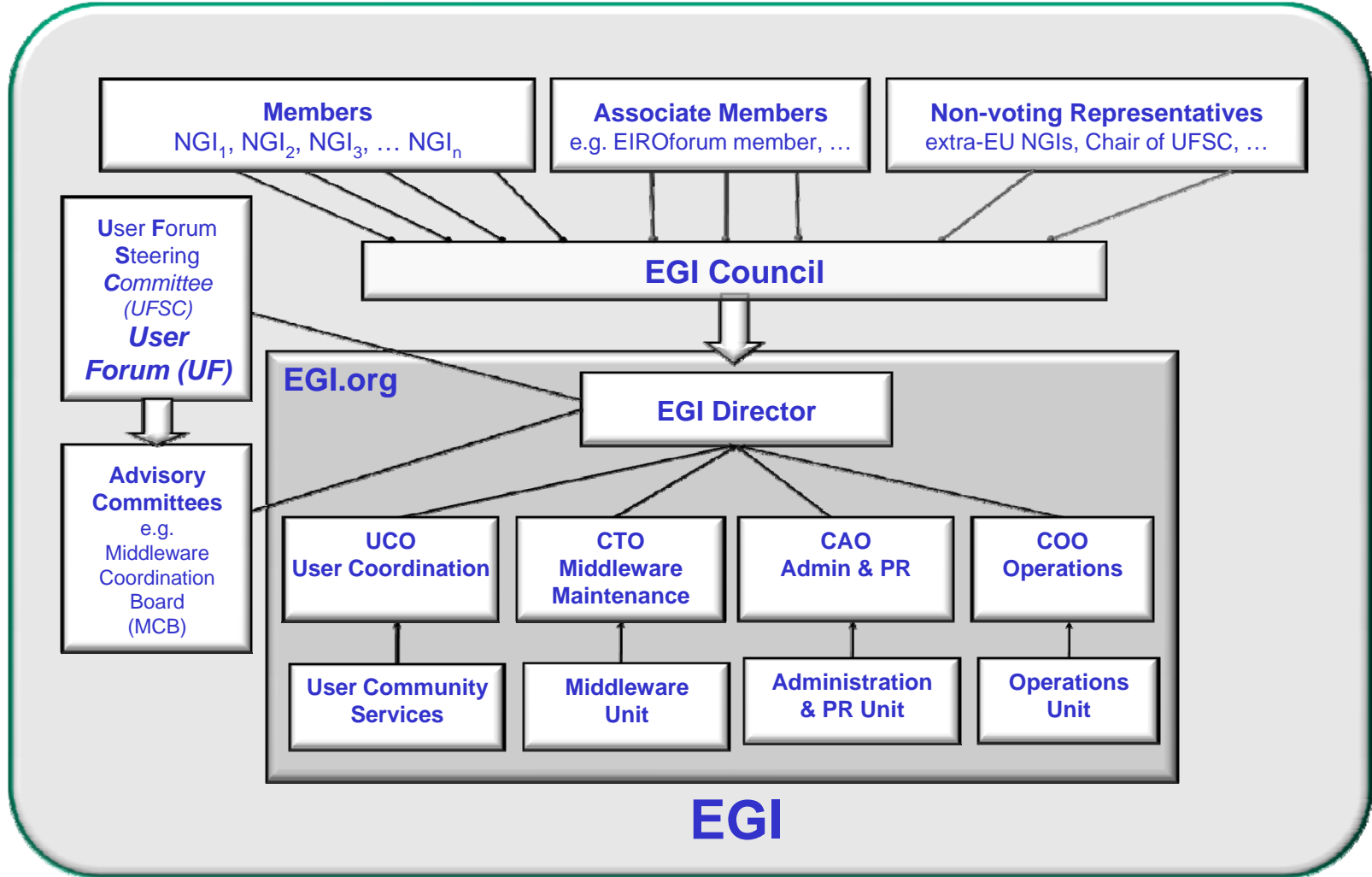
- **User Communities**

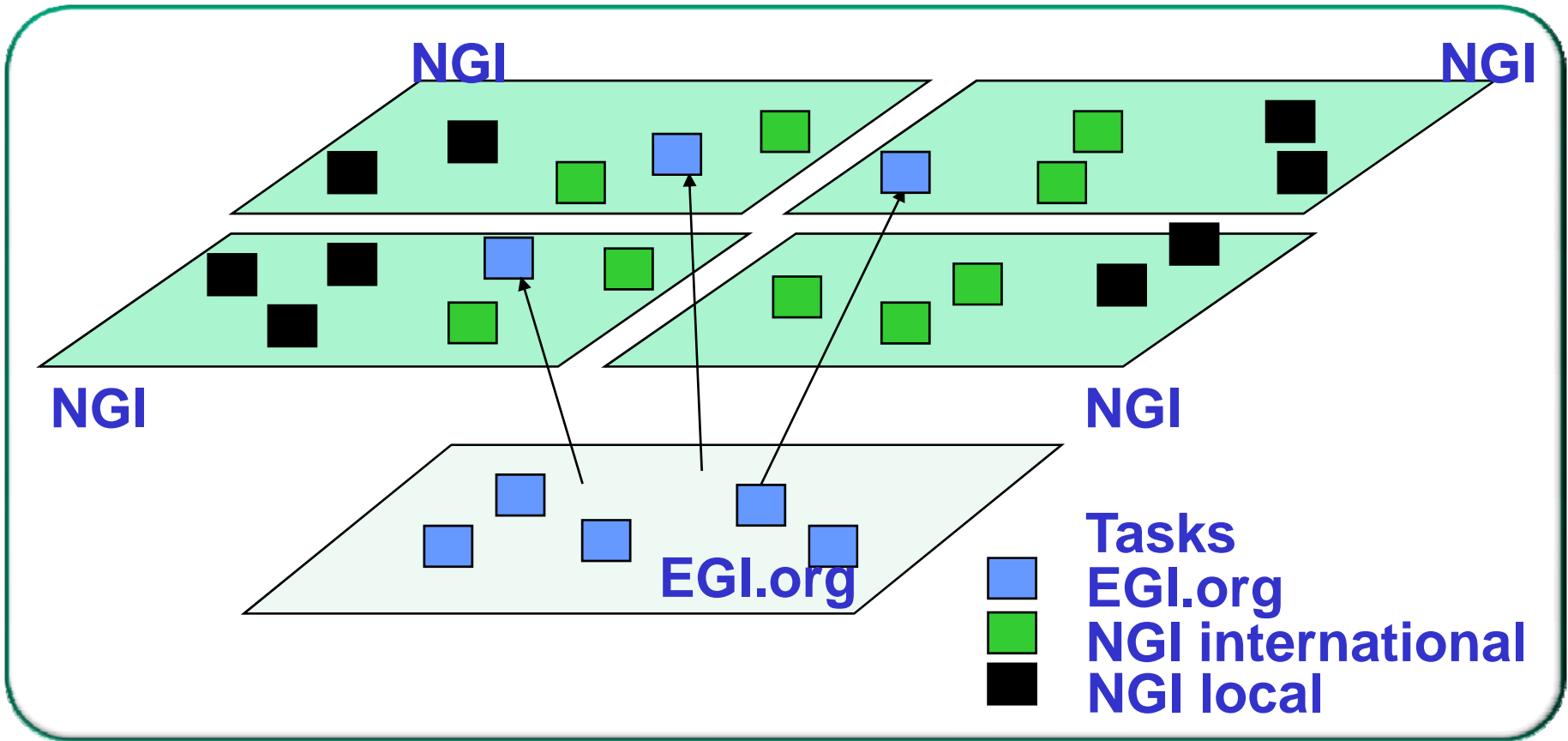
- Collection of Specialized Support Centers (SSCs).
- Probably both “Scientific” and “Support” SSCs.

- **Middleware Providers**

- gLite consortium
- ARC consortium
- Unicore consortium

<http://web.eu-egi.eu>





- **Les chercheurs utilisent la grille :**
 - Multi-science
 - Collaborer
 - Partager les ressources informatiques
 - Produire meilleurs résultats, plus rapidement

- **Evolution :**
 - gLite est encore plus stable, plus « scalable », ...
 - Plus des services complémentaires disponible
 - Projet réorienté « fournisseur des services »
 - Plan pour les financement pérenne : EGI

- **Grille est prête pour vos travaux.**

- **Site web de l'activité NA4 :**
 - <http://egeena4.lal.in2p3.fr/>
- **Informations gLite :**
 - <http://glite.web.cern.ch/glite/documentation/userguide.asp>
 - <http://glite.web.cern.ch/glite/documentation/default.asp>
- **« User information Group » :**
 - <http://egee-uig.web.cern.ch/egee-uig/>
- **Matériel formation de l'activité NA3 :**
 - <http://www.egee.nesc.ac.uk/trgmat/index.html>