



Enabling Grids for E-scienceE

Utilisation de la Grille EGEE en science de la planète

M. Petitdidier (IPSL/CETP)

Coordinatrice des applications en science de la planète dans EGEE

monique.petitdidier@cetp.ipsl.fr

En collaboration avec D. Weissenbach (IPSL) et les partenaires des projets Européens EGEE et DEGREE

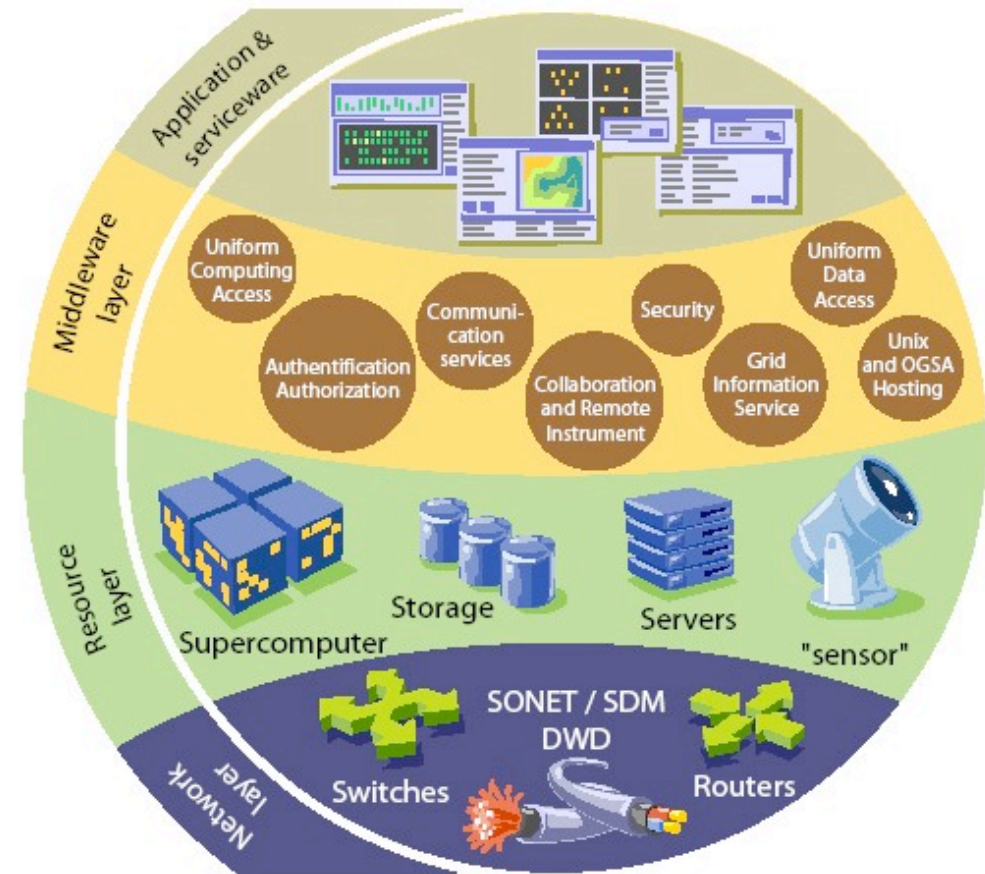
www.eu-egee.org

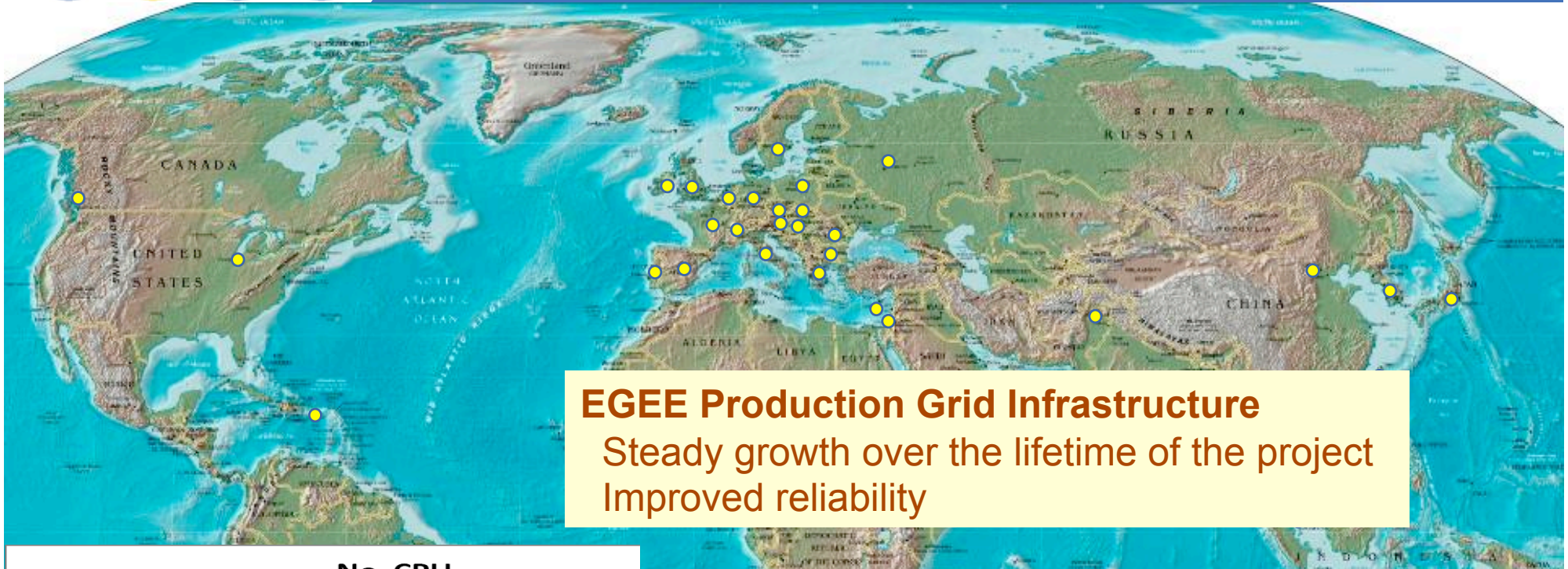


- Ressources distribuées sur un réseau rapide
- Intergiciel (middleware) pour le dialogue entre les différents éléments répartis

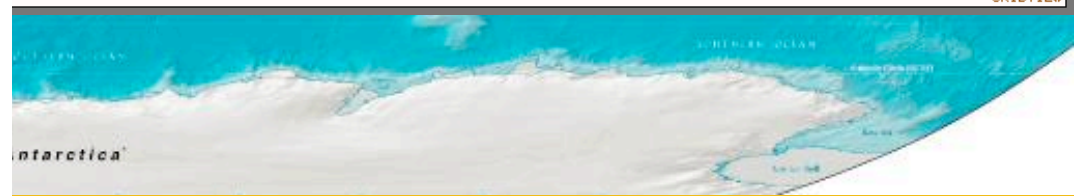
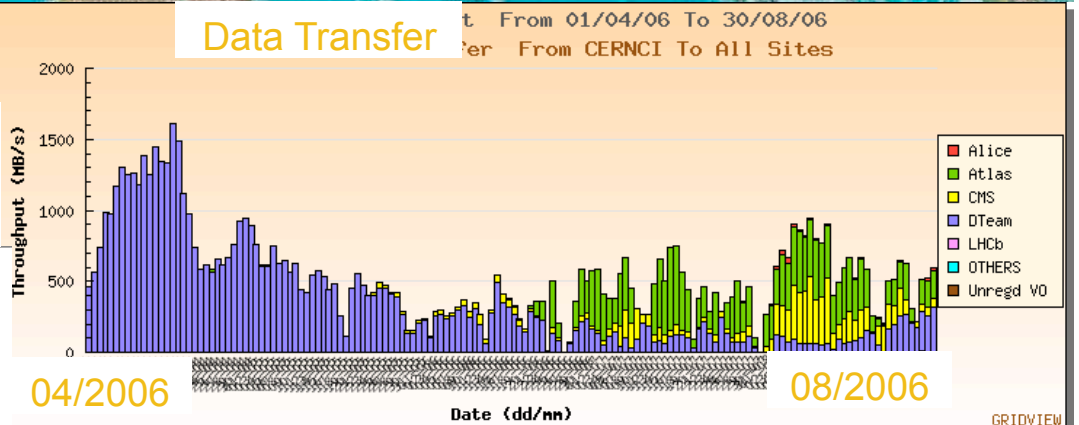
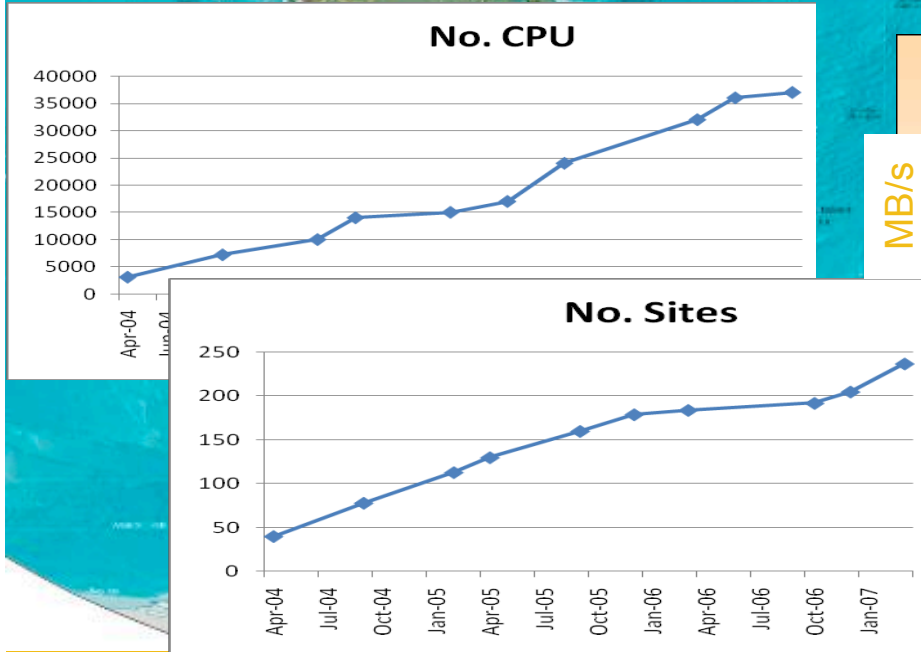
Différentes grilles

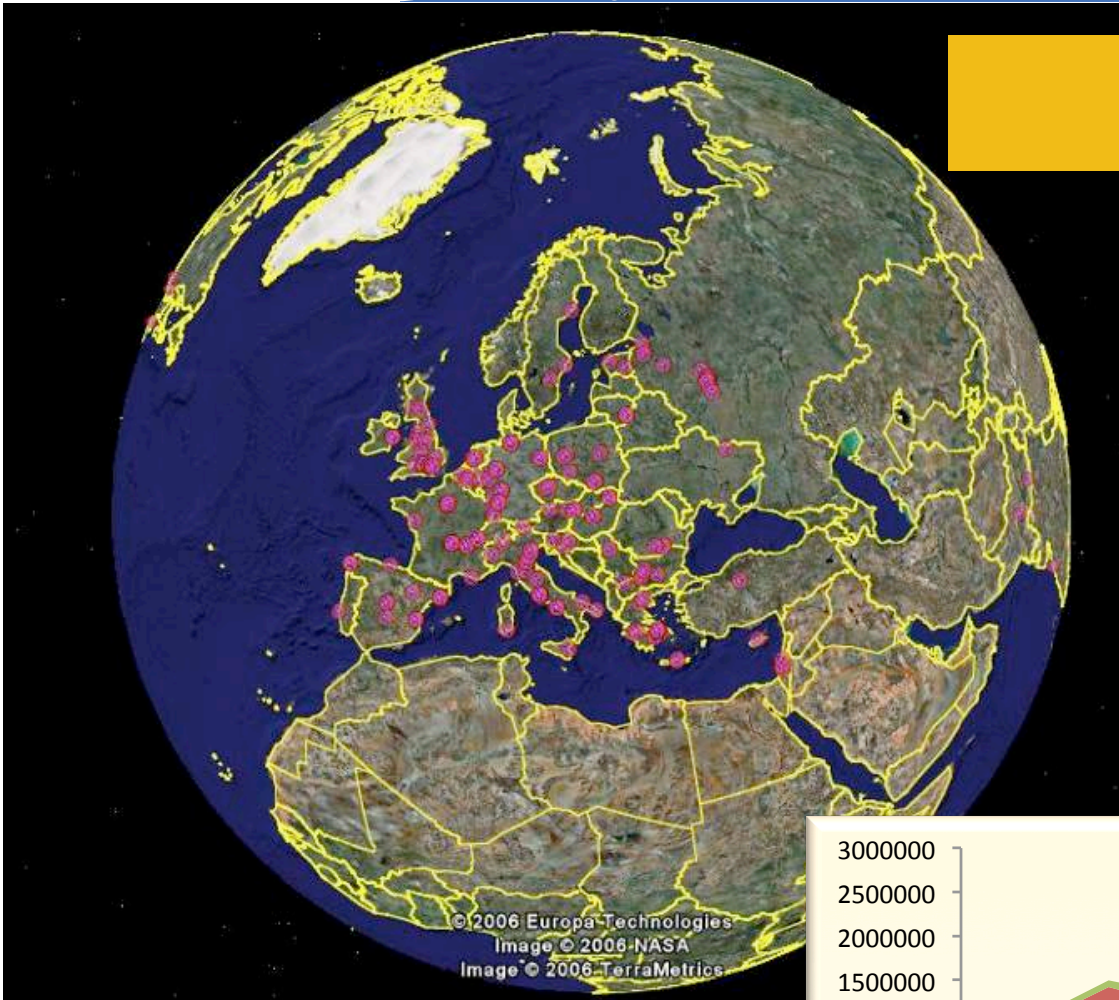
- Adaptées à différents types d'application
- TeraGrid: superprocesseurs
- EGEE: Cluster et ferme de PCs sous Linux- Grille de production
- Grid 5000: Grille de recherche





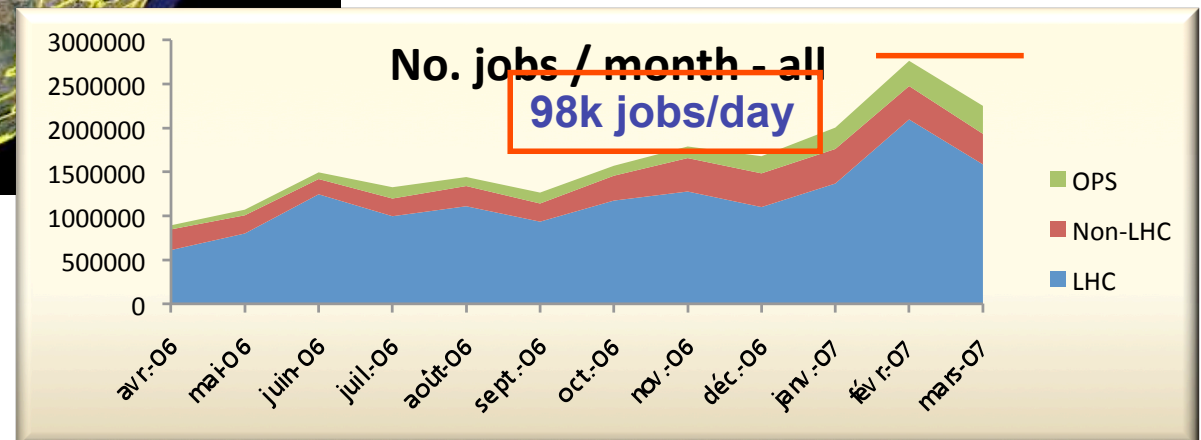
EGEE Production Grid Infrastructure
 Steady growth over the lifetime of the project
 Improved reliability



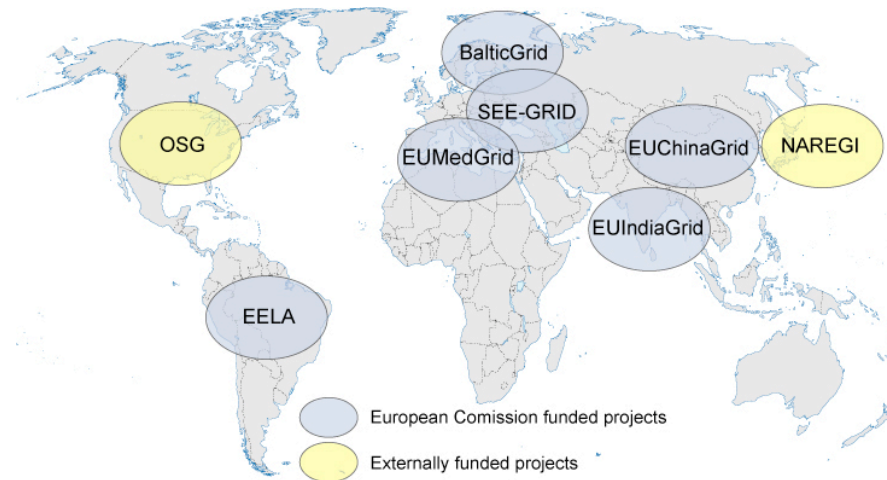


Size of the infrastructure today:

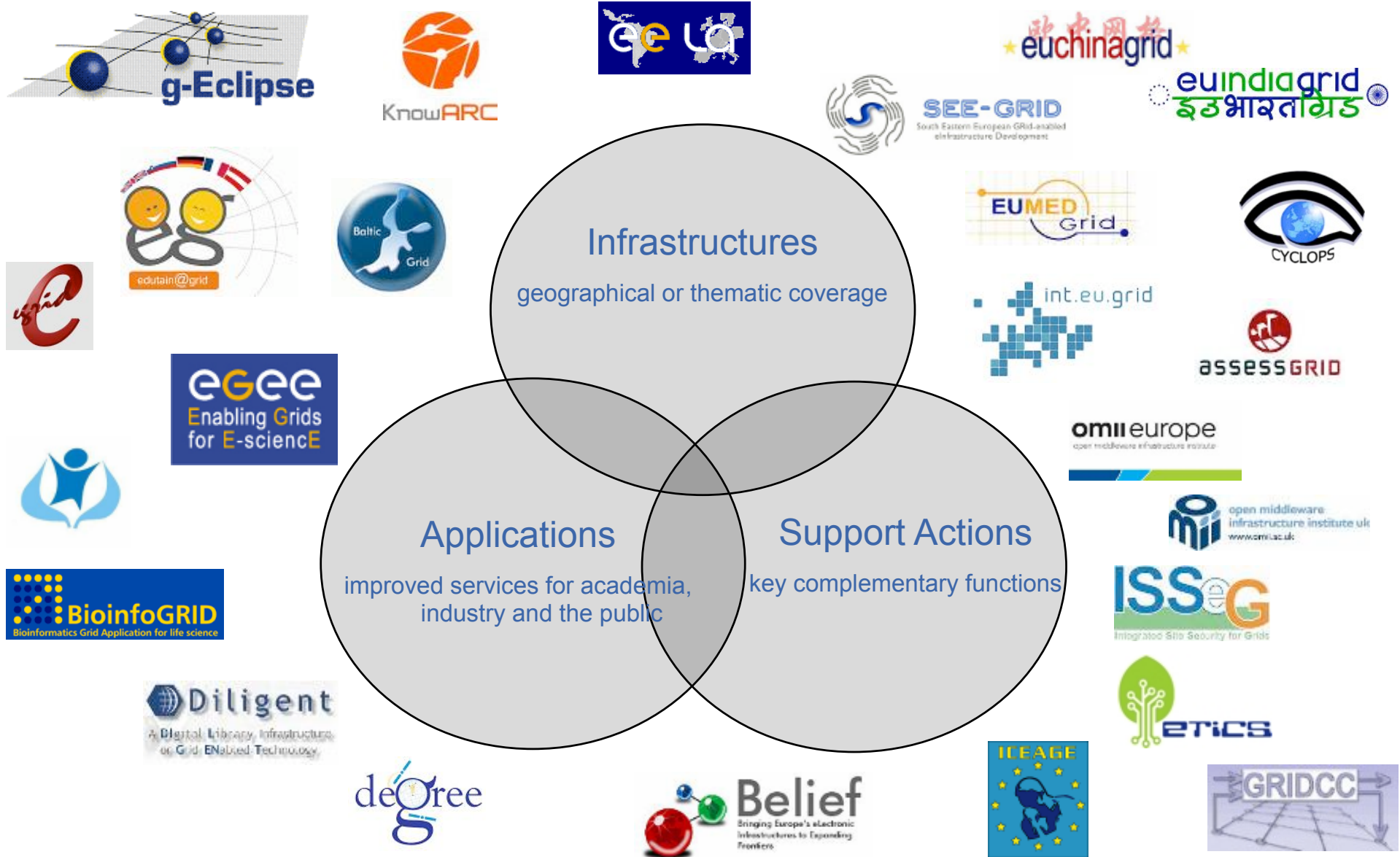
- 237 sites in 45 countries
- ~36 000 CPU
- ~ 5 PB disk, + tape MSS
- distributed operations
- copes well with increase in size and usage



- **European Union**
 - § **BalticGrid**
 - § **EELA**
 - § **EUChinaGrid**
 - § **EUIndiaGrid**
 - § **EUMedGrid**
 - § **SEE-GRID**
 - § **NorduGrid**
- **USA**
 - § **OSG**
 - § **TeraGrid**
- **Japan**
 - § **Naregi**
- **Africa- Unesco Programme**
 - § **Algeria, Ghana, Nigeria,**
 - § **Senegal, Zimbabwe**
- **Sub-Saharan Africa**
 - § **Possible pilot project**



24 projects have registered as on February 2007: [web page](#)

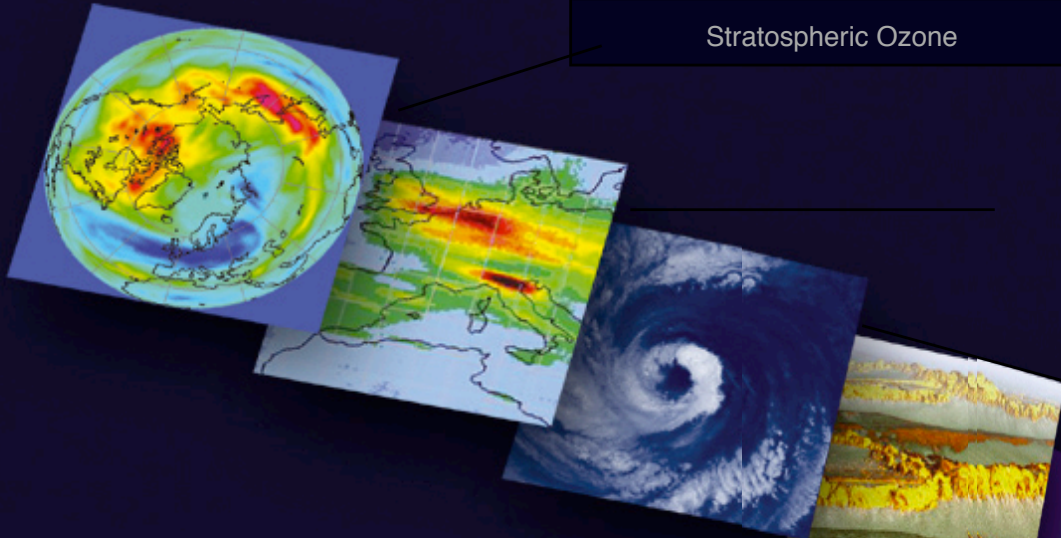


- **Nouveau partenaire**
- **Identification et parrainage**
- **Reconnaissance par l'autorité de certification (CNRS/ UREC)**
- **Enregistrement**
- **Apport de ressources car partage des ressources**
 - Création d'un site
 - Apport de ressources à un site déjà existant

- **Gestion centrale des communautés (Organisations virtuelles) qui fournissent des ressources et qui y ont accès**
 - Confiance entre les membres d'une VO et des VOs entre elles
 - Il existe d'autres modèles
- **EGEE:**
- **Certificat personnel par l'autorité de certification**
 - En France CNRS/UREC
 - Vérification des déclarations des candidats demandant un certificat auprès d'un responsable de l'entité
- **Appartenance à une Organisation Virtuelle (VO)**
 - Accès aux ressources (machines et stockage)
 - Accès aux données et algorithmes
 - Accès aux données limitées à des sous groupes

GOME total ozone assimilation

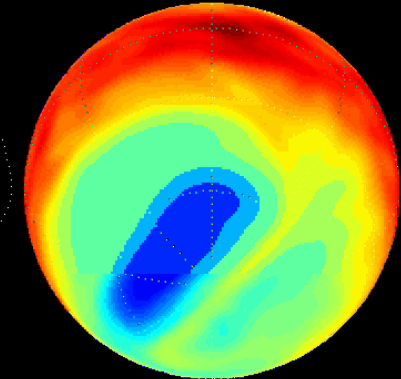
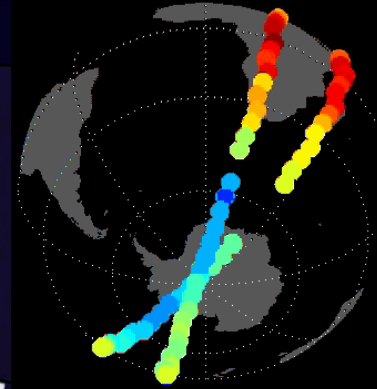
Stratospheric Ozone



Ozone at 10hPa 08:00:00 22-Sep-2002

Envisat

Assimilation



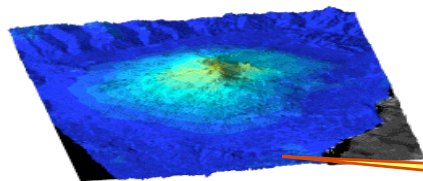
Data Assimilation Research Centre
MIPAS data (c) ESA 2002

Topography & Motion

Land cover & vegetation

Marine SST, SSH& colour

10 y displacement of Etna 1992-01



0 range displacement 14 cm

1992 year 2001

- **Année Géophysique Internationale (1957)**
 - Déploiement d'instruments,
 - mise en commun des données
- **Données satellitales :**
 - vision à une échelle globale
 - Nouveau domaine en informatique
 - Comment explorer toutes ces données ?
- **Développements de réseaux d'observations**
- **Modélisation de plus en plus complexe**
- **Internet – transmission de données et accès aux bases distribuées**
- **Services Web très développés: accès aux données, exploitation...**
- **Grille de calcul**
 - Nouvelles perspectives

- **Plateforme pour gérer des ressources de calcul, manipuler des données, fournir des services...**
 - Nombre impressionnant de ressources (exemple EGEE)
 - § Un partenaire utilise beaucoup plus de ressources qu'il n'en apporte
 - Distribution sur presque l'ensemble du monde
 - § Facilite les collaborations entre équipes dans le cadre de projets
 - Sécurité et accès contrôlé aux ressources, données...
 - Confidentialité des travaux lié à la très forte compétitivité
- **Motivations:**
 - Partage des ressources informatiques
 - § Plus de ressources utilisables que celles qu'on apporte
 - Partage des algorithmes et des données
 - Collaboration interactive pour des développements communs

2 Organisations Virtuelles: appartenance à une communauté

- ESR : Earth Science Research

VO Manager: W; Som de Cerff (KNMI) & D. Weissenbach (IPSL)
~50 personnes, Académie, Organisation, industrie pour tests
FR, BG, CH, DE, IT, NL, RU, SK, SP
Grande variété d'applications

- EGEODE (Expanding Geosciences On Demand)

VO manager : Gael Youinou, G. Vetois
~30 personnes CGGVeritas + CNRS/INSU
Geocluster
e-collaboration

- **Cas des jobs sur alerte**
 - Tremblements de terre
 - Crue des fleuves
- **Approche statistique**
 - Nécessité de nombreux runs pour avoir de bonnes estimations
- **Etude paramétrique**
 - Evaluation de très nombreux scénarios

Cas du tremblement de terre de la Martinique : 29 Novembre 2007 à 19h09

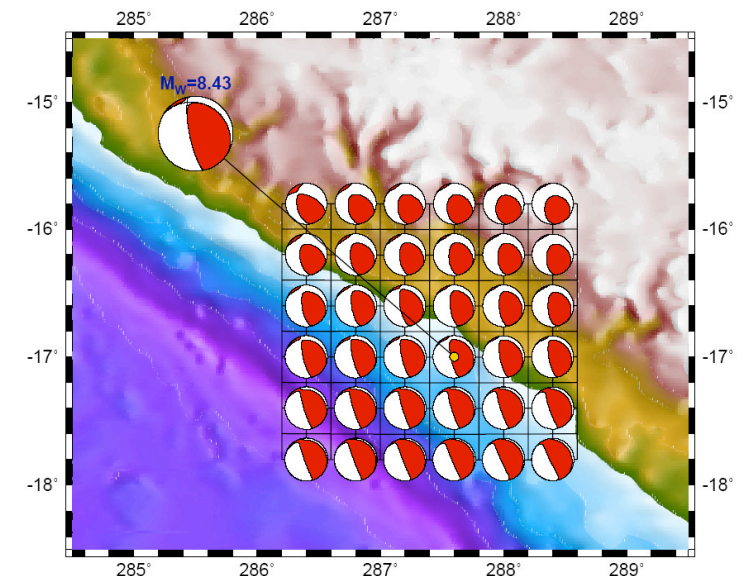
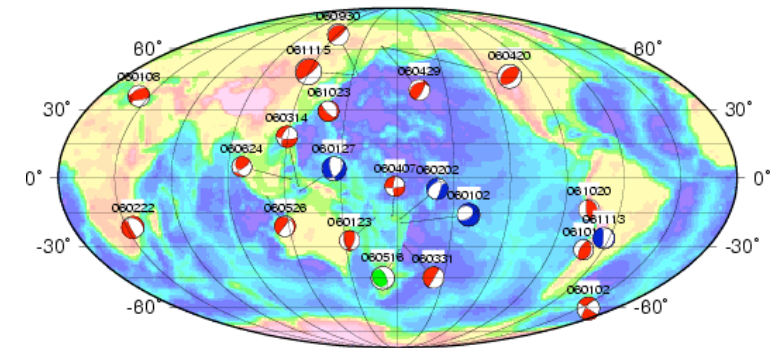
2008 - GEOSCOPE MOMENT TENSOR SOLUTIONS

- ✓ Données 28 stations sismiques de GEOSCOPE arrivées le 30/11 à 17h
- ✓ Sélection des stations et des données
- ✓ Définir une grille spatiale 3D centrée sur la localisation préliminaire de l'épicentre (1100 positions -lat, long, prof.), et la durée
- ✓ Soumettre simultanément 200-400 jobs simultanément

Résultats obtenus ~3hr après l'arrivée des données, 24h après le séisme.

Très important pour études des sources...

<http://geoscope.ipgp.jussieu.fr>: catalogue des résultats



Projet de recherche européen FP6 coordonné par le BRGM

3 ans, 15 partenaires, 9 pays

Objectif:

Développer des outils logiciels pour estimer le risque environnemental lié aux pesticides

Méthodologie

- Développement d'un grand nombre de scénarios couvrant la diversité des conditions environnementales en Europe (climats, sols, cultures)
- Modélisation du transfert de pesticides dans ces scénarios
- Intégration des résultats dans une BD et les logiciels FOOTPRINT



⇒ 10 millions de runs d'environ 2 heures chacun

GEOCLUSTER

CGG-Veritas,
France

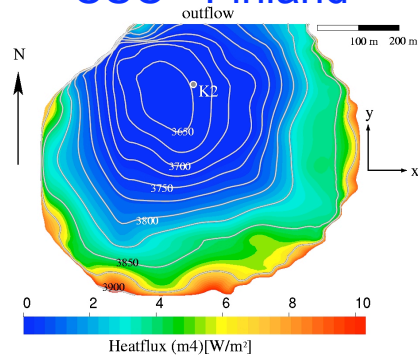


Logiciel
Plateforme
seismique
(400 modules,
serveur de
licence

VO- EGEODE

ELMER

CSC - Finland

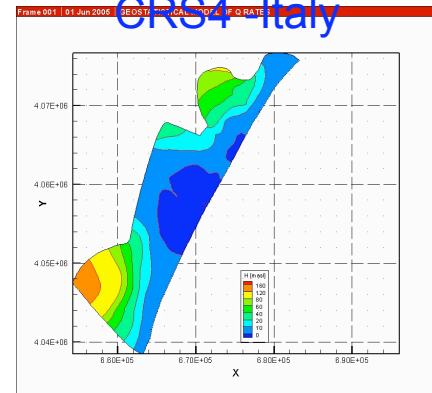


Modèles
physiques –
mécanique des
fluides, transfert
de chaleur,
acoustique...

VO- ESR

CODESA-3D

GRS4 - Italy

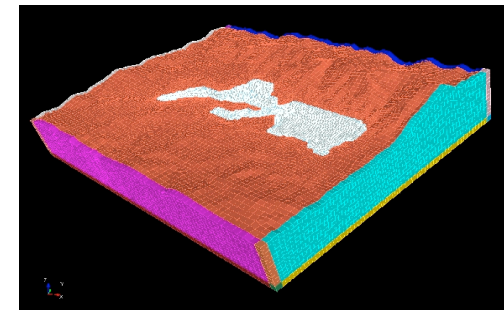


Modélisation
Intrusion d'eau
de mer dans
les aquifères
côtiers

EUMEDGRID

3DSEM_UNSTRUCT

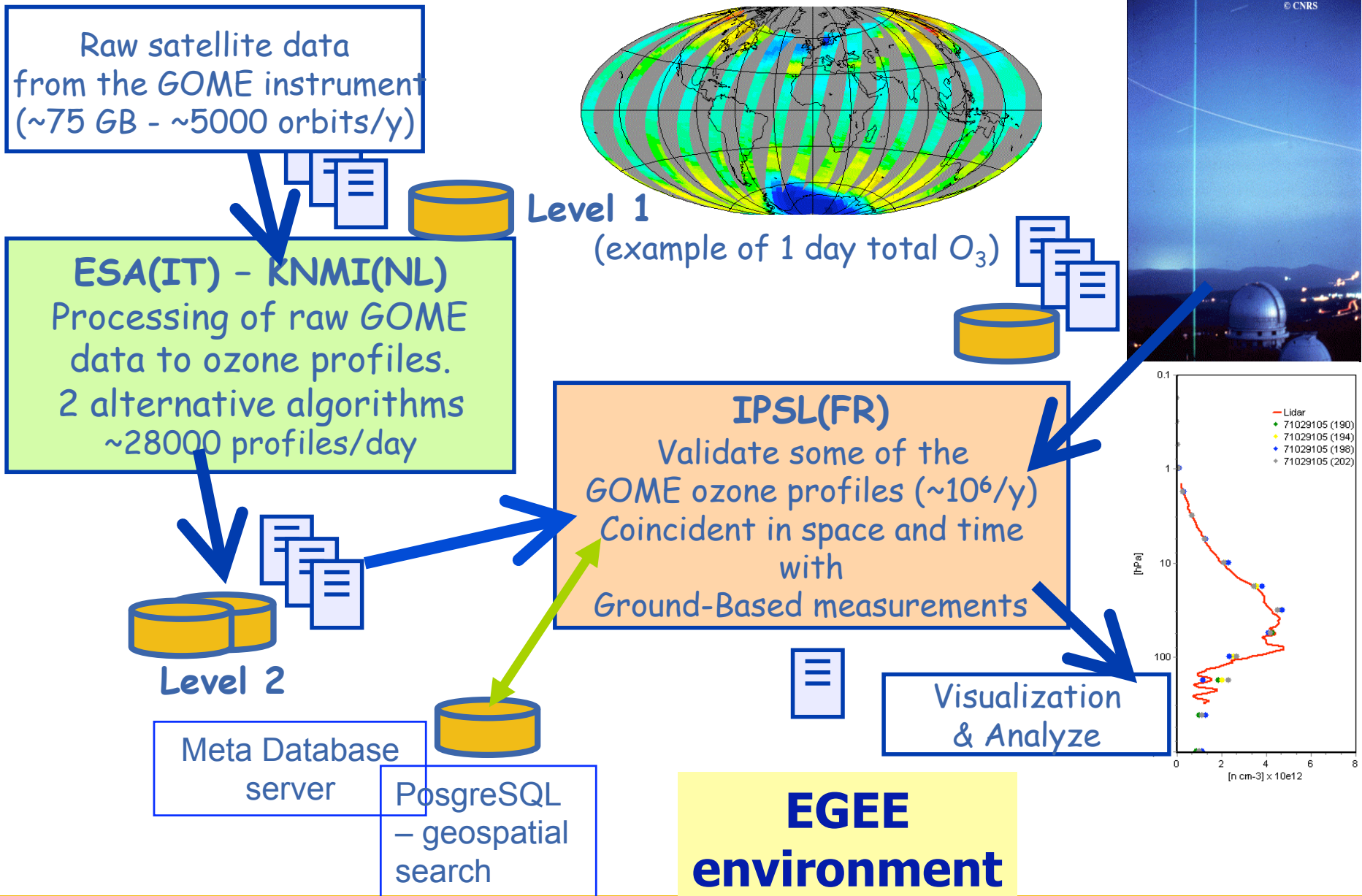
IPGP- France



Modélisation
Propagation 3D
des ondes
sismiques dans un
milieu géologique
complexe

EELA

- **Les données sont le coeur des applications en science de la planète**
- **Partage des données**
 - Jeu de données utilisé par plusieurs équipes
- **Production de grands jeux de données**
- **Exploration de grand jeu de données**
- **Développement d'outils communs: base de metadonnées, accès aux données, portail...**



Etude du bruit sismique dans un large spectre de fréquences, et pour différentes durées (jour, horaire...)

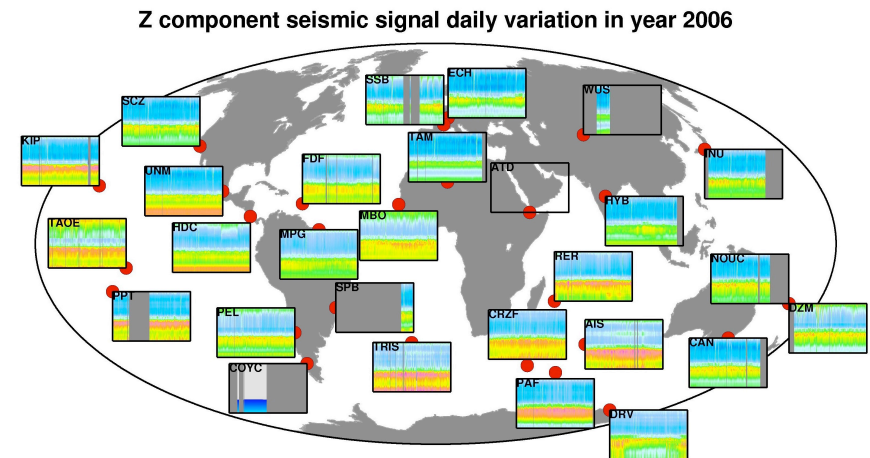
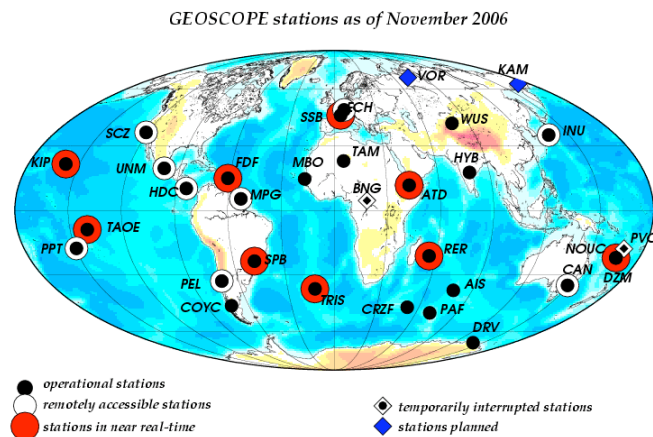
IPGP: E. Stitzmann, G. Patau, IPSL/EGEE: D.Weissenbach

Geoscope: (<http://geoscope.ipgp.jussieu.fr>) IPGP-France

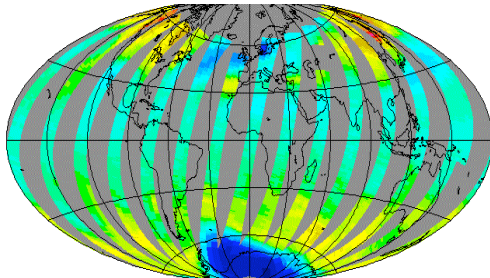
28 stations sismiques

20- 25 années de données

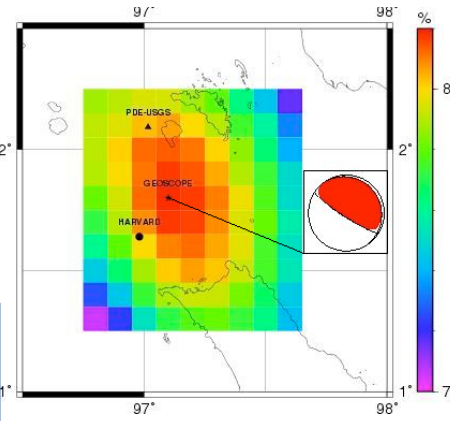
Calcul sur l'ensemble des données



- **Exemple précédent montre l'intérêt d'avoir accès à la grille à partir d'un centre de données – pas de téléchargement local de données à faire- donc possibilité d'explorer de grandes bases de données**
- **Interface d'accès aux sorties de modèles climatiques pour du post-traitement sur EGEE (DKRZ)**
- **Portails thématiques qui en plus des données donnent un accès à la grille:**
 - LEAD (Météorologie/Atmosphère) accès à TeraGrid
 - GEON (Geosciences Network) accès à TeraGrid
- **GPOD (Grid Processing On Demand) (ESRIN)**
 - Accès aux données satellitaires
 - Algorithme de calcul fourni ou non par l'utilisateur
- **Prototype pour les données météorologiques ECMWF, MétéoFrance, DWD et UK met office (projet SIMDAT)**



ESA, UTV(IT),
KNMI(NL), IPSL(FR)-
Production and
validation of 7 years of
Ozone profiles from
GOME

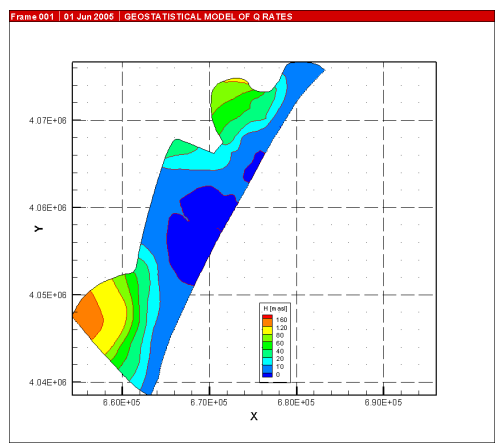


Rapid Earthquake
analysis
(mechanism and
epicenter)
50- 100CPUs
IPGP(FR)



Flood of a Danube river-
Cascade of models
(meteorology,hydraulic ,
hydrodynamic....)
UISAV(SK)-

Geocluster for
Academy and
industry CGG(FR)-



Data mining
Meteorology &
Space Weather
(GCRAS, RU)

DKRZ(DE)- Data access
studies, climate impacts on
agriculture

Specfem3D:
Seismic
application.
Benchmark for
MPI (2 to 2000
CPUs) (IPGP,FR)

Air Pollution
model- BAS(BG)

Modelling seawater
intrusion in costal
aquifer (SWIMED)
CRS4(IT),INAT(TU),
Univ.Neuchâtel(CH)-

Mars atmosphere
CETP(FR):

- **Nombreux succès –concrétisés par des publications dans des journaux scientifiques internationaux et étant inclus dans des thèses**
- **Grande variété d'applications**
- **Utilisation peu répandue dans la communauté des Sciences de la Planète**
 - Information insuffisante
 - Difficulté de porter des applications complexes
 -
- **DEGREE (Dissemination and Exploitation of GRids in Earth Science) www.eu-DEGREE.eu**
 - Dissémination d'information aux communautés Sciences de la terre et Grille
 - Analyse des besoins des Science de laTerre pour porter des applications complexes et utiliser les services déjà développés
 - Envisage le futur de la Grille pour cette communauté



Dissemination and Exploitation of GRIDs in Earth science

- **Strategic objectives**
 - Bridge the ES and GRID communities throughout Europe
 - Ensure that ES requirements are satisfied in next Grid generation
 - Ensure the integration of emerging technologies for managing ES knowledge

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

The DEGREE team:

IISAS, Slovakia (Coordinator)

CNRS, France

KNMI, The Netherlands

UNINE, Switzerland

CRS4, Italy

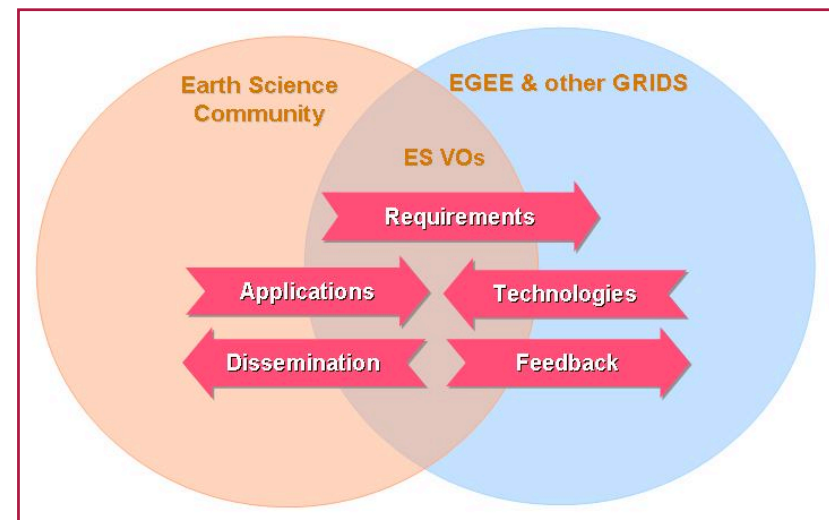
SCAI, Germany

GCRAS, Russia

ESA-ESRIN, Italy

CGG, France

Dutch Space, The Netherlands

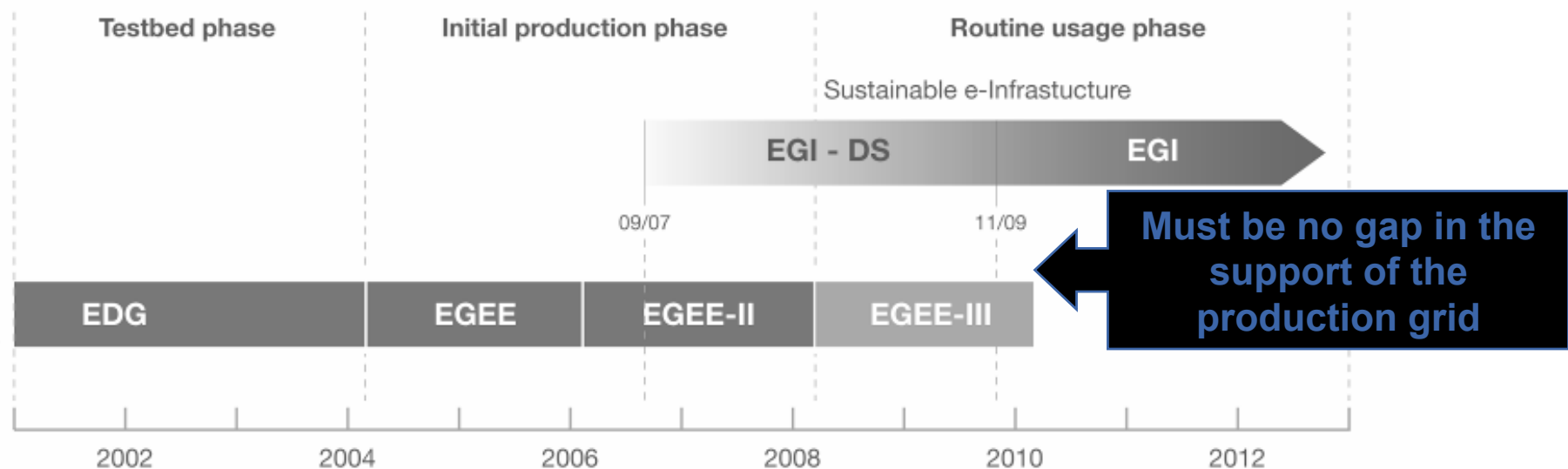


Project Vision

Build a bridge linking the ES and Grid communities

- **Nouveaux outils à développer pour utiliser tout le potentiel de la grille**
 - Du aux changements d'échelle – nombre de CPUs disponible, stockage..)
 - Lié aux besoins d'explorer de vastes jeux de données
 - Creation de plateforme intégrant des services web, des ressources de calcul, des systèmes d'information, des bases de connaissance
- **Nouvelle approche du travail dans les Sciences de la Terre**
 - Role des scientifiques
 - § collaboration interactive -> moins de duplication de développement et/ou d'adaptation
 - § Plus de temps pour des nouvelles idées, des recherches nouvelles
 - § Besoin de confidentialité des recherches
 - Conception de nouvelles applications (accès à de grands jeux de données, plus de CPUs...)

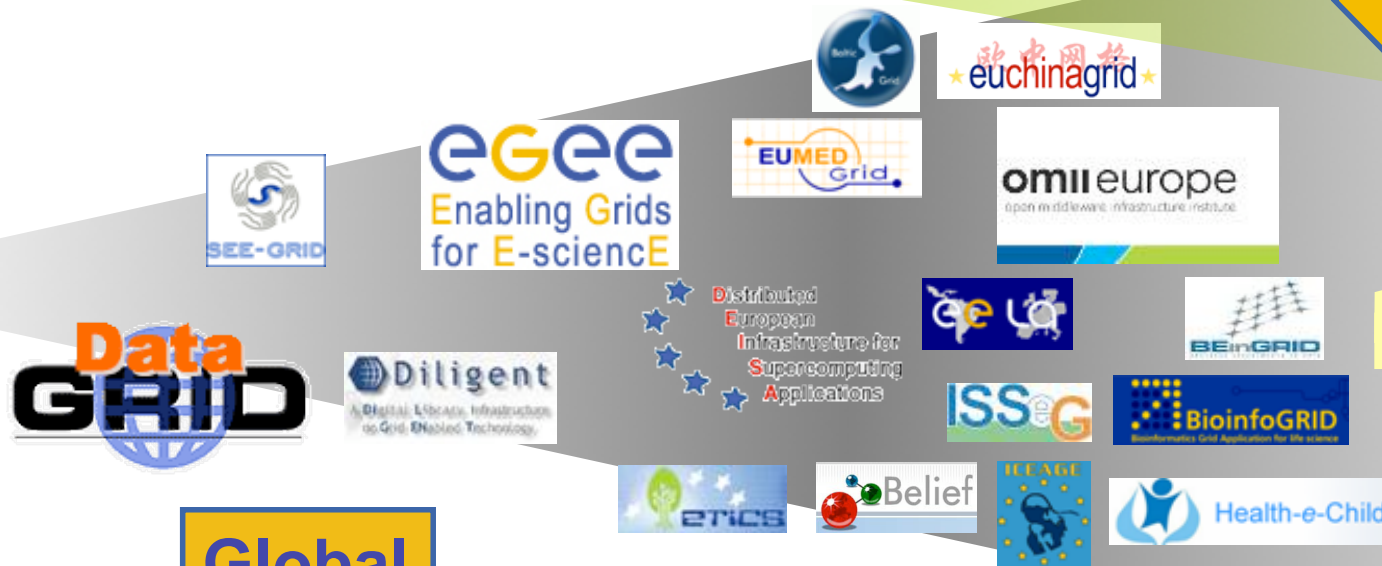
- Need to prepare permanent, common **Grid infrastructure**
- Ensure the long-term sustainability of the European e-Infrastructure independent of short project funding cycles
- Coordinate the integration and interaction between National Grid Infrastructures (NGIs)
- Operate the production Grid infrastructure on a European level for a wide range of scientific disciplines



National



Global



Testbeds

Routine Usage

Utility Service

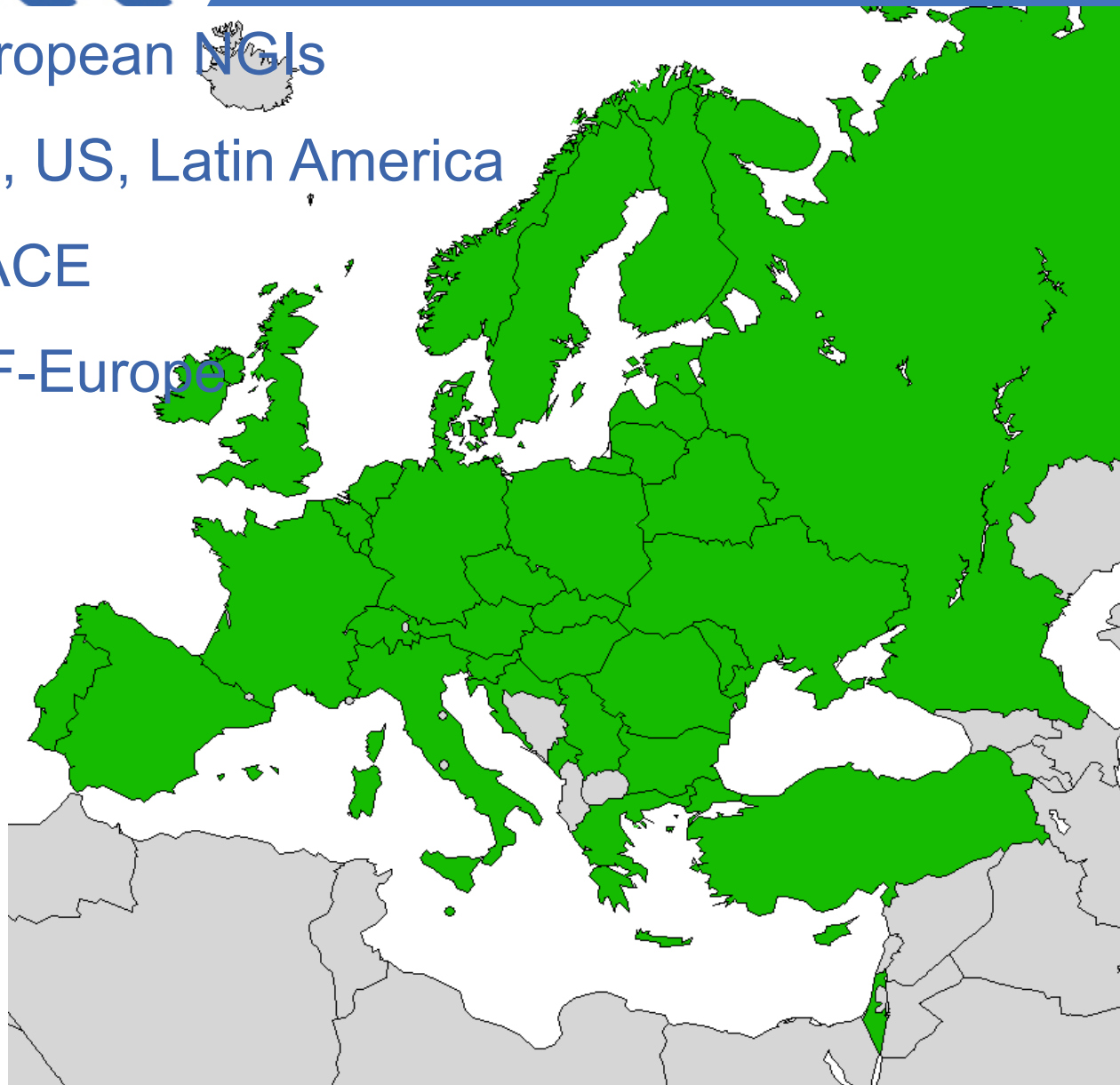
38 European NGIs

+ Asia, US, Latin America

+ PRACE

+ OGF-Europe

+ ...



- Austria
- Belarus
- Belgium
- Bulgaria
- Croatia
- Cyprus
- Czech_Republic
- Denmark
- Estonia
- Finland
- France
- Germany
- Greece
- Hungary
- Ireland
- Israel
- Italy
- Latvia
- Lithuania
- Luxembourg
- Malta
- Moldova
- Montenegro
- Netherlands
- Norway
- Poland
- Portugal
- Romania
- Russia
- Serbia
- Slovakia
- Slovenia
- Spain
- Sweden
- Switzerland
- Turkey
- UK
- Ukraine

- **Pushing frontiers of scientific discovery by exploiting advanced computational methods.**