

L'Institut des Grilles

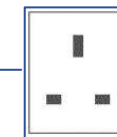
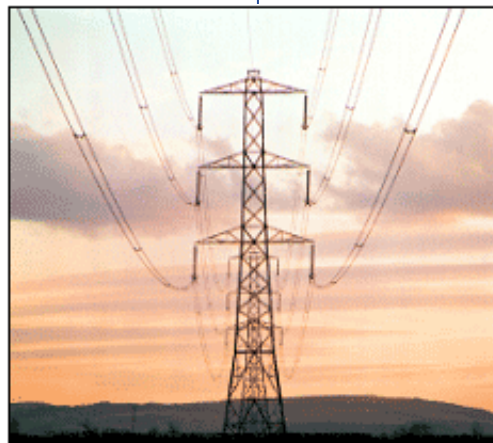
Guy Wormser
Directeur
Journée Mésocentres
13 Février 2008

Analogy with the Electricity Power Grid



Power Stations

Distribution Infrastructure



'Standard Interface'

EGEE : Enabling Grids for E-science

Goal

create a general European Grid
production quality infrastructure on top of
present and future EU RN infrastructure

Build on

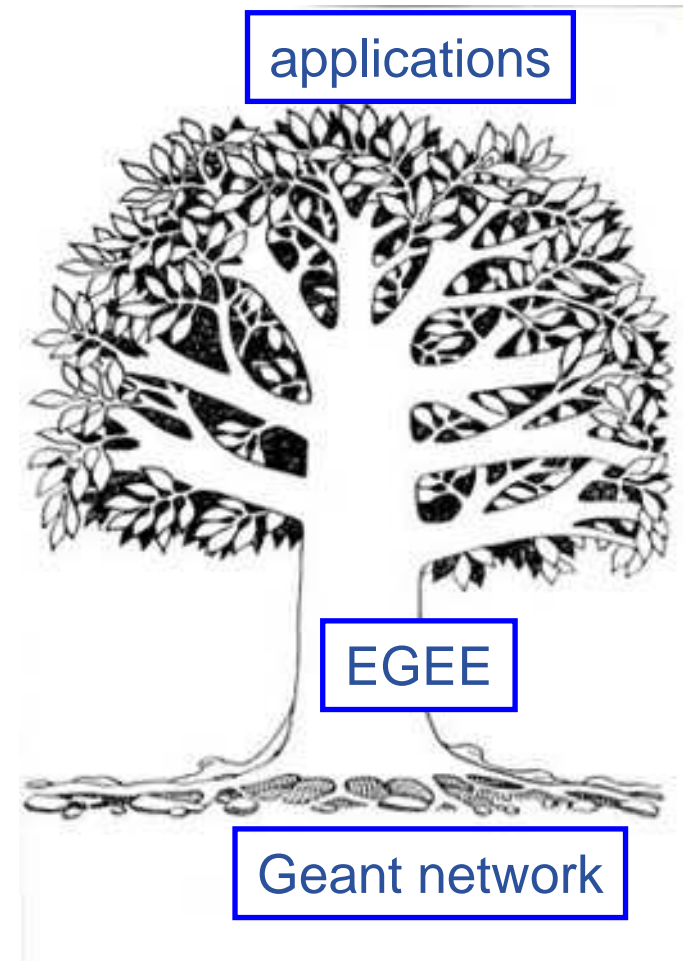
EU and EU member states major
investment in Grid Technology
Several pioneering prototype results
Largest Grid development team in the
world

Goal can be achieved for about €100m/4 years on top
of the national and regional initiatives

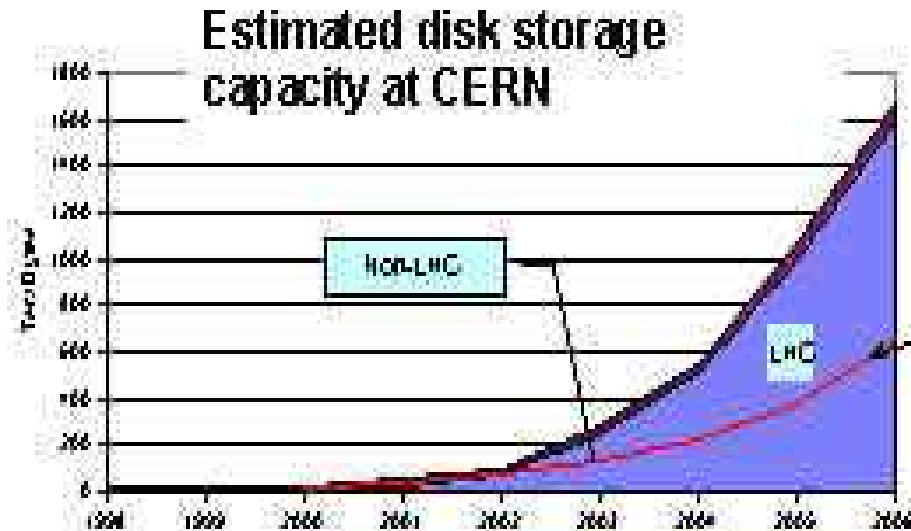
Approach

Leverage current and planned national
and regional Grid programmes (e.g.
LCG)

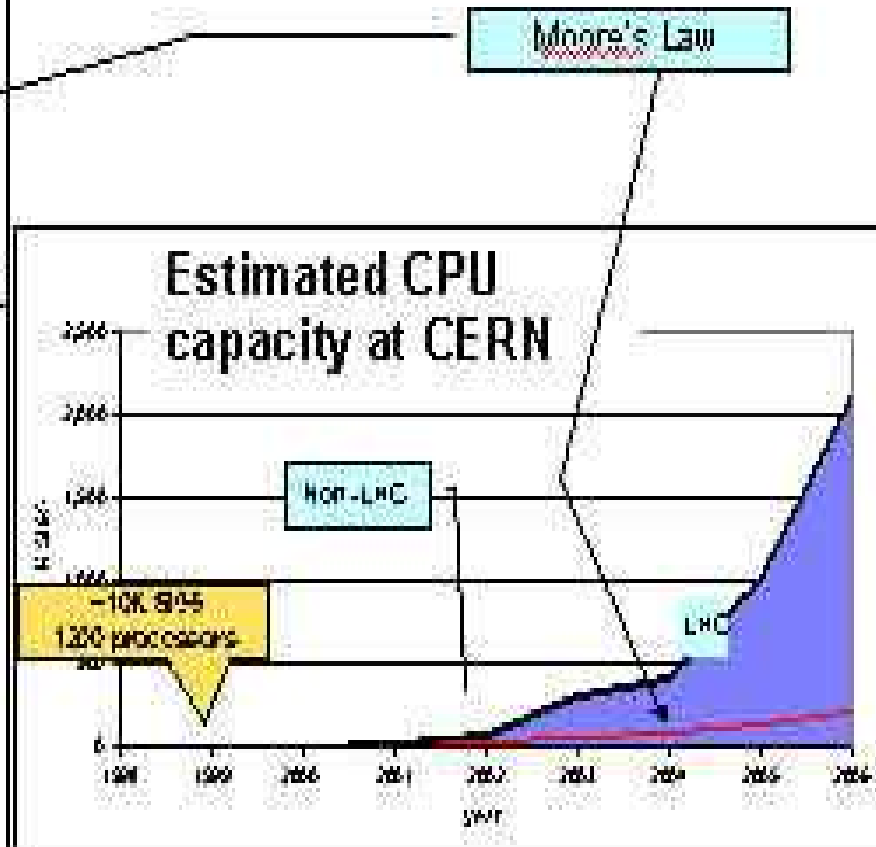
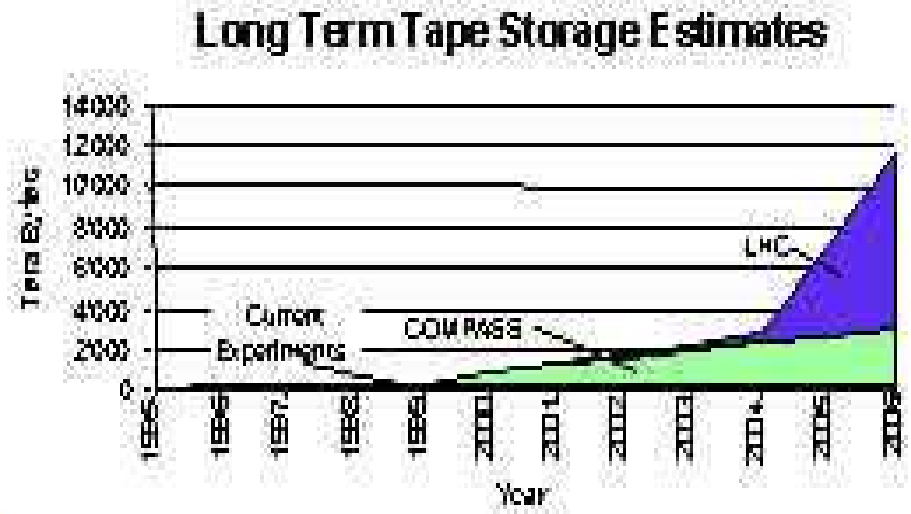
Work closely with relevant industrial Grid
developers, NRNs and US



New solutions are necessary!



Small fraction of the main analysis capacity will be at CERN



L'intérêt scientifique des grilles

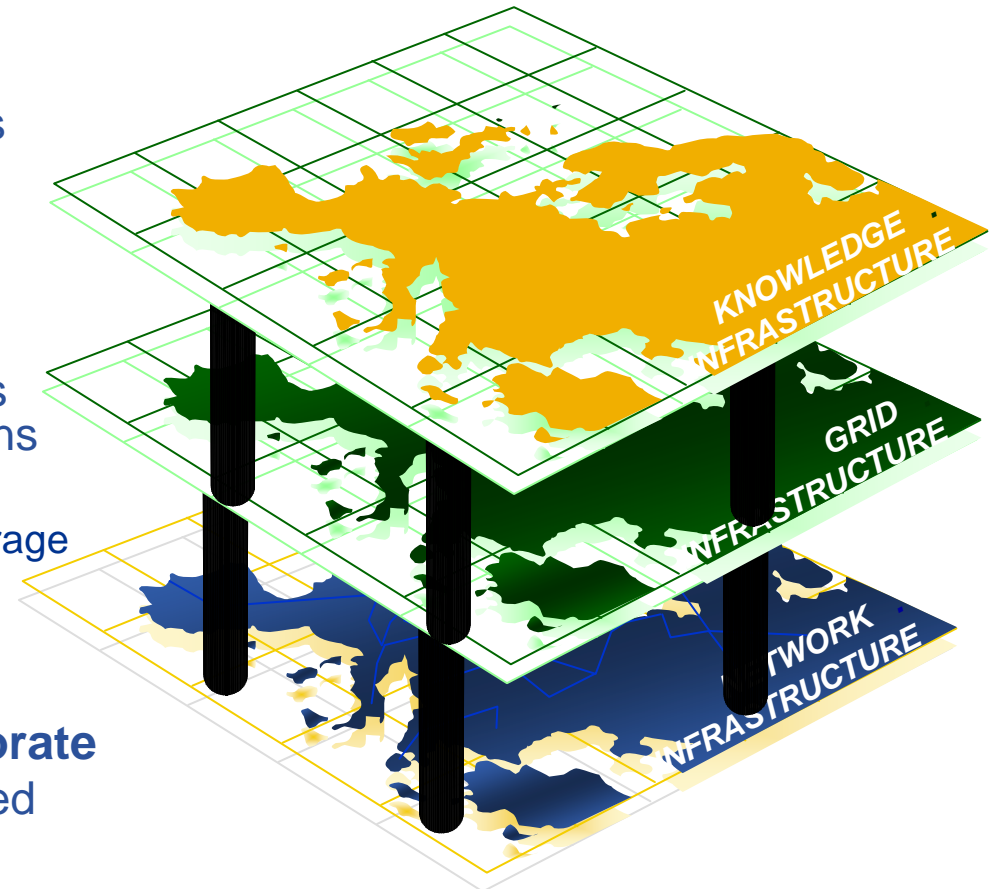
- **Accès transparent à des données distribuées**
 - Exemples Sciences de la terre, sciences de la vie
- **Manipulation de très grands volumes de données**
 - Physique des particules, astrophysique, sciences humaines
- **Très grande flexibilité des ressources de calcul**
 - Gestion des catastrophes
 - Challenge grippe aviaire , malaria

How e-Infrastructures help e-Science

- **e-Infrastructures provide easier access for**
 - Small research groups
 - Scientists from many different fields
 - Remote and still developing countries

- **To new technologies**
 - Produce and store massive amounts of data
 - Transparent access to millions of files across different administrative domains
 - Low cost access to resources
 - Mobilise large amounts of CPU & storage on short notice (PC clusters)
 - High-end facilities (supercomputers)

- **And help to find new ways to collaborate**
 - Develops applications using distributed complex workflows
 - Eases distributed collaborations
 - Provides new ways of community building
 - Gives easier access to higher education

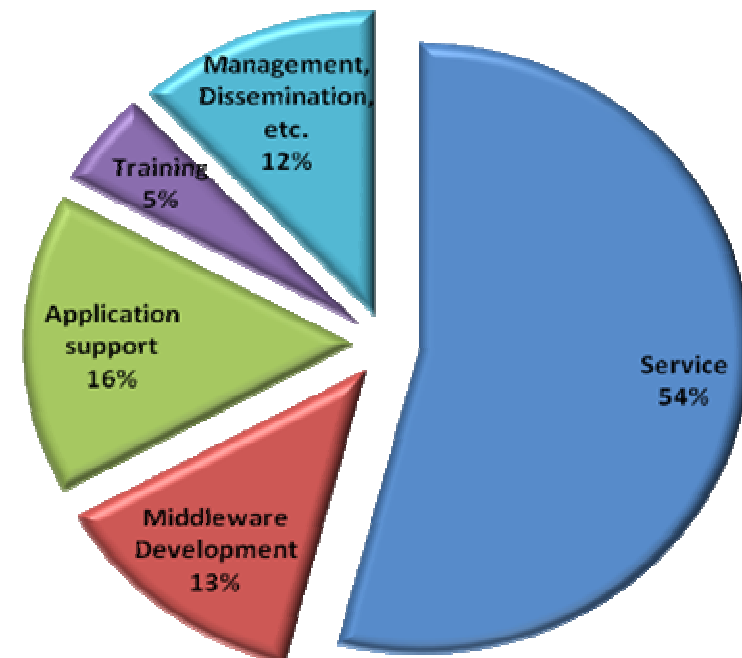




Main Objectives

- Operate a large-scale, production quality grid infrastructure for e-Science
- Attract new resources and users from industry as well as sciences

EGEE Project Activities



- **Infrastructure operation**
 - Currently includes ~250 sites across 45 countries
 - Continuous monitoring of grid services & automated site configuration/management
 - Support many Virtual Organisations from diverse research disciplines

- **Middleware**
 - Production quality middleware distributed under business friendly open source licence
 - Implements a service-oriented architecture that virtualises resources
 - Adheres to recommendations on web service interoperability and evolving towards emerging standards

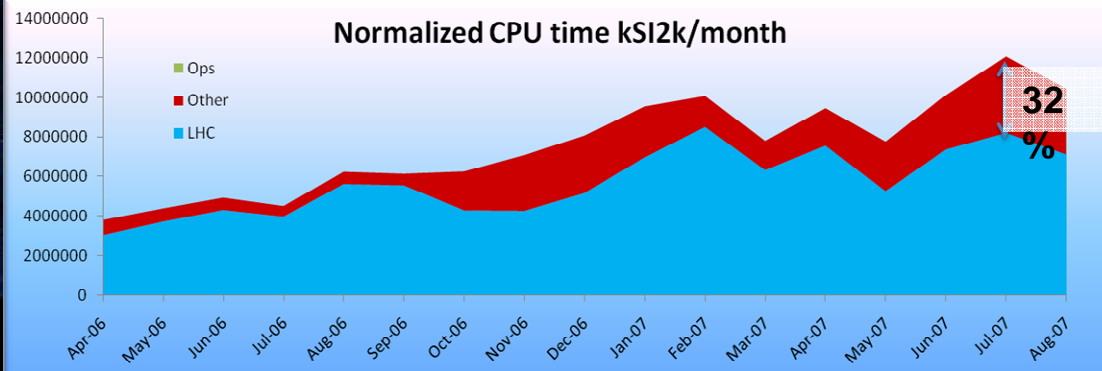
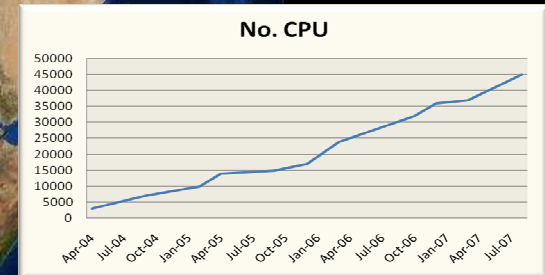
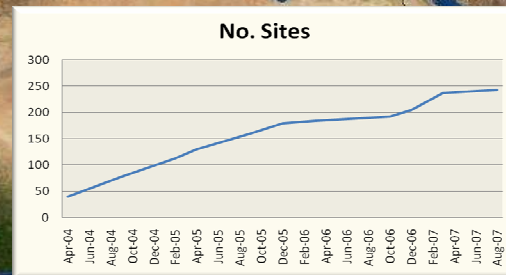
- **User Support - *Managed process from first contact through to production usage***
 - Training
 - Expertise in grid-enabling applications
 - Online helpdesk
 - Networking events (User Forum, Conferences etc.)



240 sites
45 countries
41,000 CPUs
5 PetaBytes
>5000 users
>100 VOs
>100,000 jobs/day

Archeology
Astronomy
Astrophysics
Civil Protection
Comp. Chemistry
Earth Sciences
Finance
Fusion
Geophysics
High Energy Physics
Life Sciences
Multimedia
Material Sciences

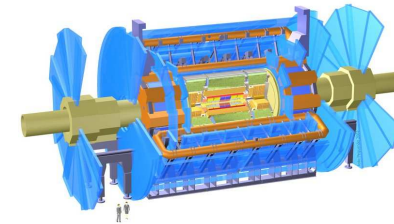
...



Types of applications

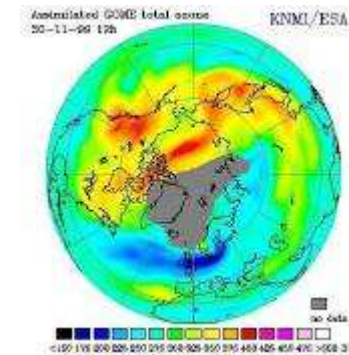
- **Simulation**

- LHC Monte Carlo simulations; Fusion; WISDOM
- Jobs needing significant processing power; Large number of independent jobs; limited input data; significant output data



- **Bulk Processing**

- HEP ; Processing of satellite data
- Distributed input data; Large amount of input and output data; Job management (WMS); Metadata services; complex data structures



- **Parallel Jobs**

- Climate models, computational chemistry
- Large number of independent but communicating jobs; Need for simultaneous access to large number of CPUs; MPI libraries

- **Short-response delays**

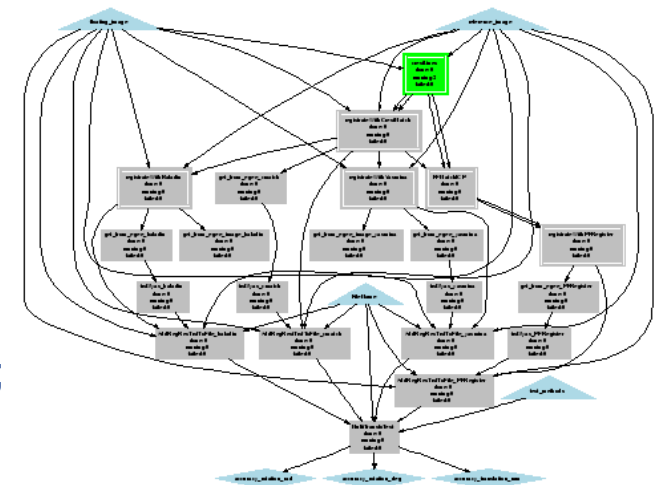
- Prototyping new applications; grid Monitoring grid; Interactivity
- Limited input & output data; processing needs but fast response and quality of service

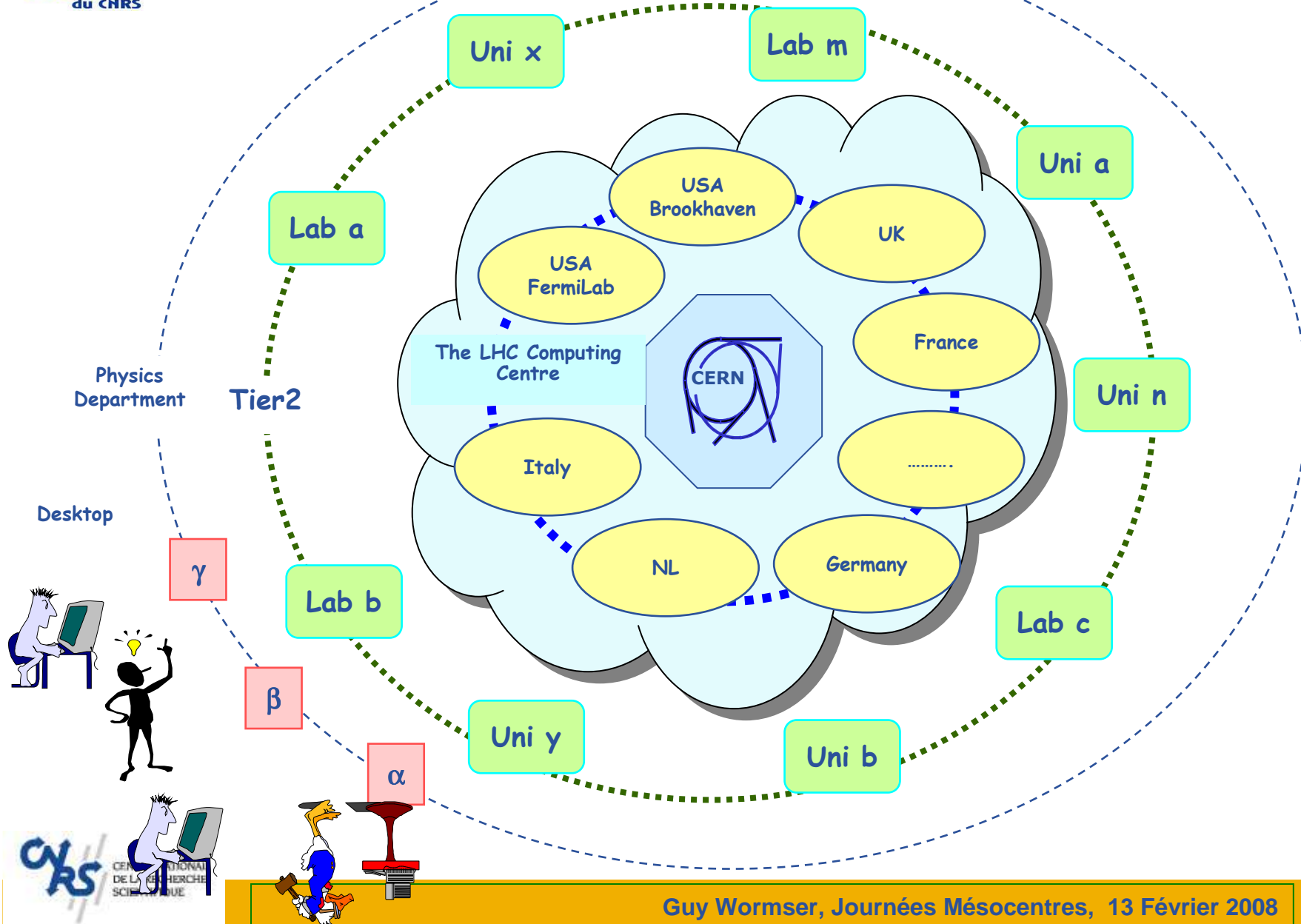
- **Workflow**

- Medical imaging; flood analysis
- Complex analysis algorithms; complex dependencies between jobs

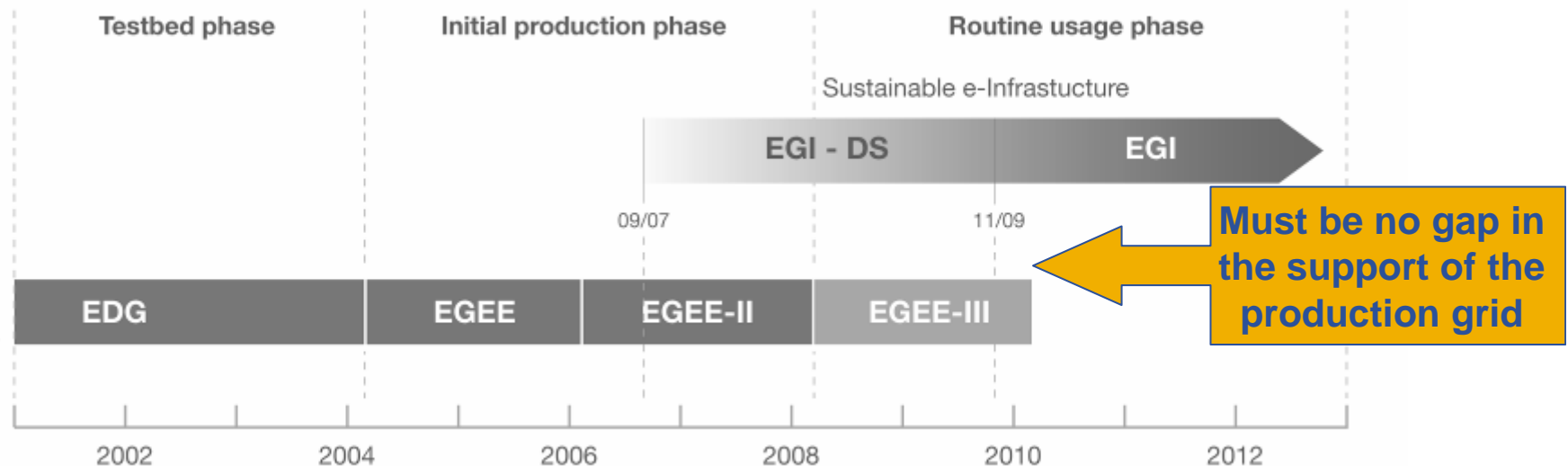
- **Commercial Applications**

- Non-open source software; Geocluster (seismic platform); FlexX (molecular docking); Matlab, Mathematics; Idl, ...
- License server associated to an application deployment





- 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



Cyprus



Israel



- EGI Design Study proposal approved to the European Commission (started 1st September'07)
- Supported by 30+ National Grid Initiatives (NGIs)
- 2 year project to prepare the setup and operation of a new organizational model for a sustainable pan-European grid infrastructure
- Federated model bringing together NGIs to build a European organisation
- Well defined, complimentary responsibilities between NGIs and EGI

Each NGI

- recognized national body with a single point-of-contact
- mobilise national funding and resources
- operate the national e-Infrastructure
- support user communities (application independent, and open to new user communities and resource providers)
- contribute and adhere to international standards and policies

Responsibilities between NGIs and EGI are clearly separated and complementary

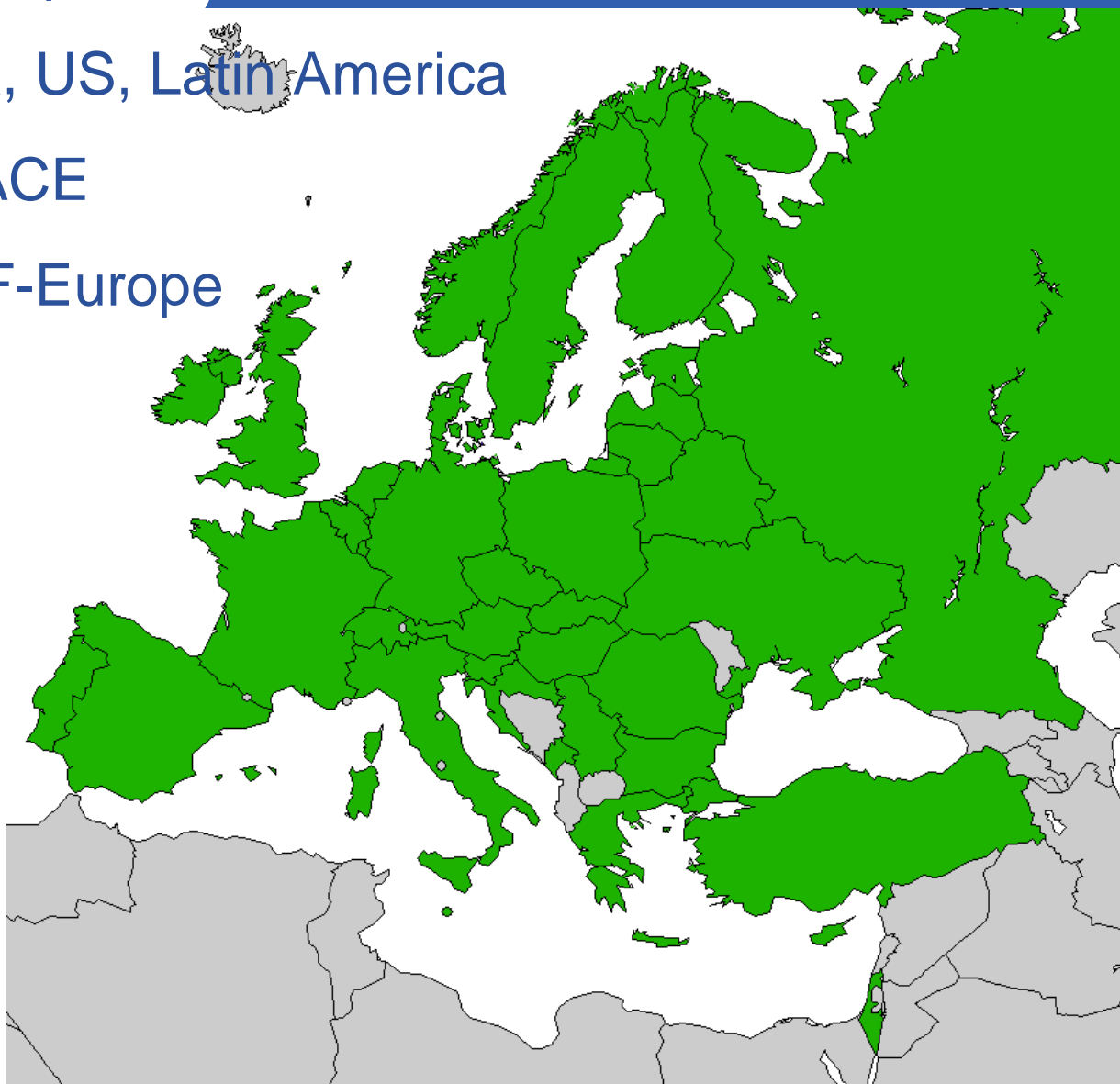
European NGIs

+ 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
- Montenegro
- Netherlands
- Norway
- Poland
- Portugal
- Romania
- Russia
- Serbia
- Slovakia
- Slovenia
- Spain
- Sweden
- Switzerland
- Turkey
- UK
- Ukraine

- Etudes de nouveaux langages, de nouveaux protocoles
- Action Concertée GRID en 2001
- Lancement de Grid5000, outil original au service de la recherche sur les grilles
- Projets européens Réseaux d'excellence COREGRID
- Projet d'Observatoire de la Grille

- La grille pour les supercalculateurs, le réseau européen DEISA

Pourquoi un Institut des Grilles

- L'activité au CNRS sur les grilles a atteint en 2007 une importance considérable en volume et en impact
- **Fédérer l'ensemble des activités du CNRS dans le domaine des grilles de recherche et des grilles de production**
 - Meilleure visibilité
 - Meilleure efficacité
 - Renforcer l'interaction entre ces deux domaines
- **Point de contact pour les partenariats nationaux et internationaux**
 - Représentant du CNRS pour les contrats européens, pour les discussions auprès du Ministère
 - Noyau central pour la « National Grid Initiative »
 - Partenariat à nouer avec l'INRIA et les autres organismes pour la recherche sur les grilles de calcul
 - Partenaire des initiatives régionales
- « **Evangelisation** » auprès de nouvelles communautés scientifiques
- **Actions d'animation, de formation et de dissémination**
- **Partenaire fort d'une future structure de type « Institut Français des Grilles »**

Ce qu'est l'Institut des Grilles

- **Formellement Unité Propre de Service UPS3107, créée le 1^{er} Septembre 2007**
- **Possède en tant que telle**
 - Un directeur (GW), deux sous-directeurs (D. Boutigny, V. Donzeau-Gouge)
 - Budget d'environ 400 k€
 - Soutien opérationnel projet EGEE
 - Actions nouvelles
 - Personnel : 3 CDD attendus en 2008, post-docs
 - Rattachée au département MPPU
- **Délégation pour les contrats européens (EGEE-III, EGI, EDGES, EELA2)**
- **Liste des Membres de l'Institut**

Les laboratoires participants

- **Les 30 laboratoires:**

APC, CC_IN2P3, CPPM, **CREATIS**, **LIP**, **I3S**, **IBCP**, **IN2P3_adm**,
IPGP, IPHC, IPNL, IPNO, **IRISA**, **IRIT**, **LABRI**, LAL, LAPP, **LIFL**,
LIG, **LIP6**, LLR, **LORIA**, LPC Clermont, LPNHE, **LRI**, **IPSL**,
LPSC, **LSIT**, Subatech, **UREC**

- 13 laboratoires IN2P3 liés à EGEE/LCG

- 11 laboratoires d'informatique

- 5 laboratoires applicatifs liés science de la vie, science de la Terre,...

- Support administratif

- **GDR Architecture Systèmes et Réseaux (ASR)**

- **Liste de 305 personnes !**

- **Comité de pilotage**
 - Présidé par le Président du CNRS (ou le Directeur Général)
 - Comprend les directeurs des départements scientifiques et instituts nationaux
 - Un représentant de chaque partenaire
- **Conseil scientifique**
 - Conseille le comité de pilotage
 - Formé de hautes personnalités scientifiques étrangères provenant pour moitié du monde de la recherche sur les grilles et pour moitié des infrastructures de production
- **Comité de direction**
 - Formé des directeurs des unités CNRS (ou de leurs représentants) impliqués dans les travaux de l'Institut
 - Première réunion plénière le 4 Décembre à Orsay

- **Série de réunions au Ministère sous l'impulsion de la DGRI**
 - Grilles de recherche, recherche sur les grilles (20/09)
 - Infrastructures digitales en SHS (21/09)
 - Réunion préliminaire sur les grilles de production (25/09) puis réunion plénière du groupe de travail (12/10)
- **Rôle central reconnu pour l'Institut des Grilles**
- **Vers un partenariat étroit entre IdG-CNRS, ADT-INRIA (Aladdin), GET, Universités pour les grilles de recherche**
- **De même, nécessité d'un partenariat fort pour les grilles de production dans le cadre du futur programme EGEE-III et de la création d'une National Grid Initiative**
- **MoU signé d'ici fin Février 2008 sous l'égide du Comité de Pilotage National (Ministère DGRI, CNRS, CEA, Universités, INSERM, INRA, INRIA, RENATER)**
 - Comité de pilotage national
 - Exercice de prospective nationale dont la maîtrise d'œuvre est confiée à l'Institut des Grilles
 - Institut des Grilles représentant des signataires pour les projets internationaux en commun dans le domaine des grilles de production

L'Institut des grilles principal maître d'œuvre de l'exercice de prospective nationale

- **Formation des groupes de travail avec les différents partenaires (RENATER, CEA,..)**
- **Organisation générale, nomination des animateurs**
- **Collecte des informations, préparation du colloque final**
- **Interaction avec le Conseil Scientifique**
- **Invitations orateurs étrangers, logistique,...**

- **Thématiques**

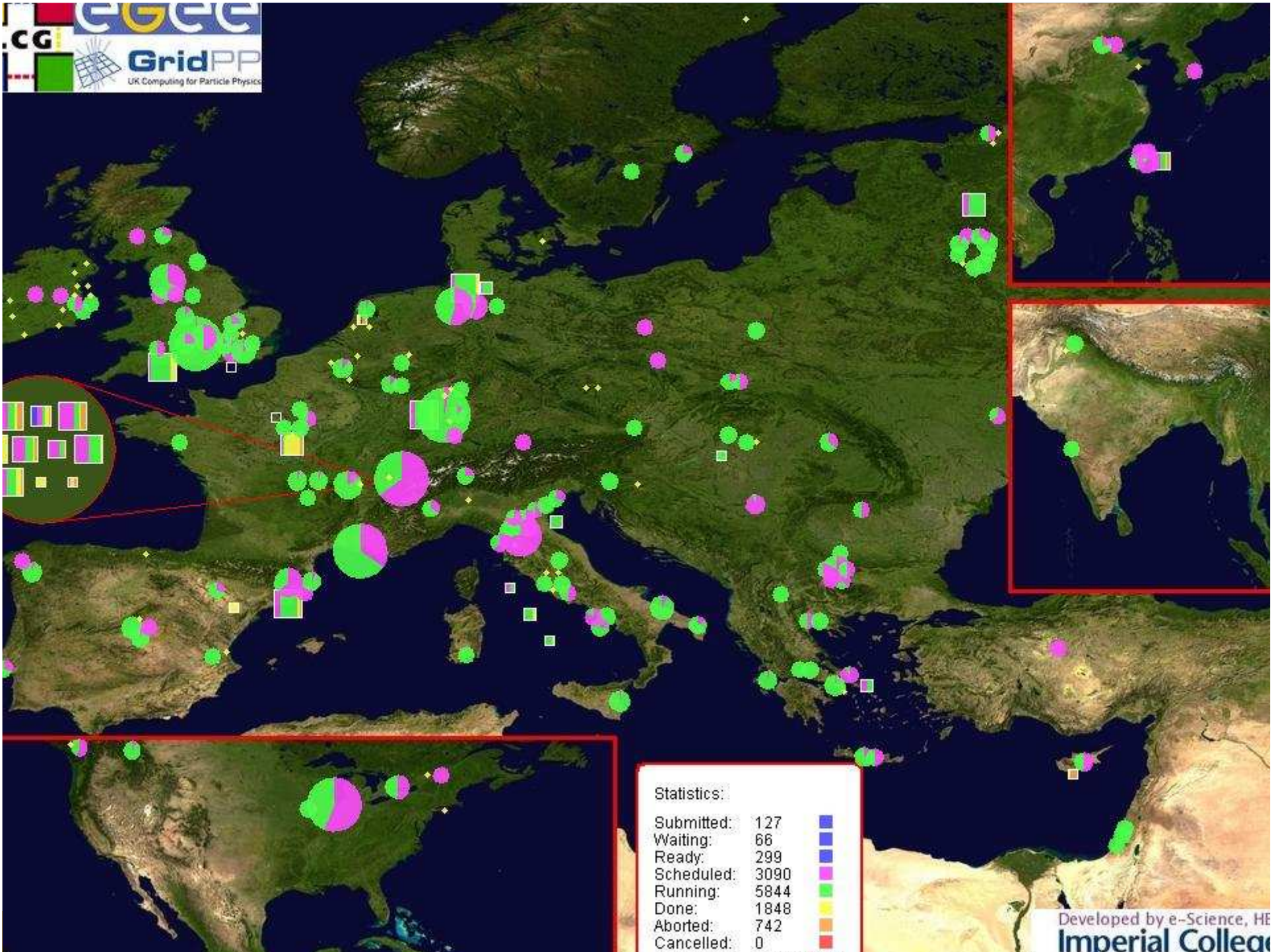
- Sciences de la Planète et environnement-Sciences de l'Univers
- Sciences du Vivant (avec d'éventuels sous groupes vu la taille du domaine: bioinformatique, imagerie, physiologie,...)
- Sciences Humaines
- Chimie
- Sciences de l'ingénieur et Informatique
- Physique (Nanosciences, Mathématiques, calcul scientifique, optique, fusion,...)
- Physique des particules, physique nucléaire, astroparticules

- **Transverses**

- Grilles de données
- **Grilles régionales; relation avec GRID5000**
- Relation supercalculateurs
- TGE/ESFRI
- Accès à la grille-Comité utilisateurs- expression des exigences
- Relations avec les Industriels

- **Partenariat INRIA**
- **Création NGI/EGEE-III**
- **Participation EGI Design Study**
- **Participation à des actions très visibles**
 - Atelier « Afrique et Internet » 10-12 Décembre Montpellier
 - Implantation de nœuds de grille en Afrique subsaharienne (Sénégal et Afrique du Sud) en 2008 , formation
 - Partenariat CNRS-Helmoltz (Allemagne)
 - EGEE-III Users Forum
- **Favoriser les contacts avec des disciplines encore peu actives sur la grille**
- **Visite en régions**
 - Rhône-Alpes 10 Janvier
 - Bordeaux Février
 - Lille Mars
- **Contact avec des industriels (IBM,..)**

- **NGI/EGI bien en place**
- **Grille nationale intégration verticale/horizontale**
 - Résultats scientifiques importants obtenus grâce aux grilles
- **Passerelles GRID5000/ Grille de production**
- **Diffusion des nouveaux langages/logiciels vers la production**
- **Observatoire de la grille**
- **Partenariats INRIA autres organismes**
- **Forte visibilité**
- **Liens avec la communauté supercalculateurs**
- **Publications –projets menés grâce à l’Institut**
- **Labellisation**
- **Actions de formation, dissémination, valorisation**



- Interconnecter des mésocentres sur la grille de production nationale
- Interopérer des grilles locales basées sur certains mésocentres avec la grille nationale de production
- Construire les passerelles grilles-supercalculateurs à partir des mésocentres
- Favoriser l'émergence de grilles locales et régionales autour de mésocentres

- **L'Institut des Grilles est né pour mieux faire face à des défis exaltants : apporter le meilleur outil possible aux diverses communautés applicatives pour permettre des découvertes scientifiques de première ampleur**
- **L'Institut des Grilles est un espace idéal pour encourager les collaborations entre les chercheurs en informatique travaillant sur les grilles, les chercheurs et ingénieurs en charge des grilles de production et les communautés scientifiques les utilisant**
- **Une des missions essentielles de l'Institut est de nouer des partenariats tant en France qu'à l'étranger**
- **Le bon outil au bon moment! A nous de nous en emparer pour servir au mieux la recherche !**