CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

IN2P3
Centre de Calcul

Infrastructures de Grille

Quelques exemples

Fabio Hernandez
Responsable Grid Computing
Centre de Calcul de l'IN2P3
fabio@in2p3.fr
Metz, 21 juin 2004

Table des Matières

- Présentation de l'IN2P3
- Objectif
- Projets
  - USA
  - Europe
  - Asie
- Conclusions
- Questions

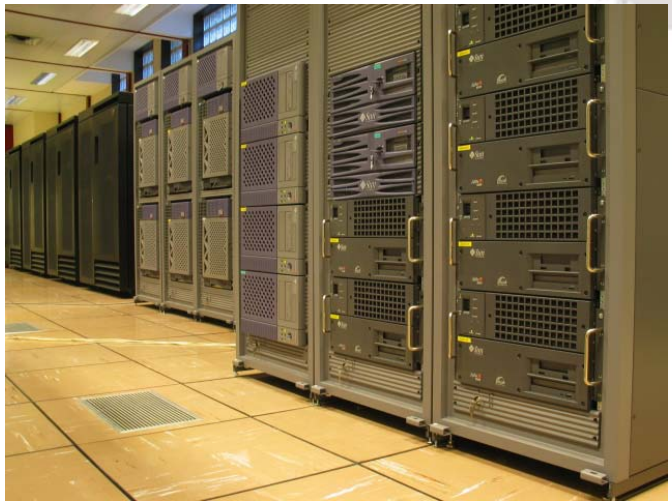
IN2P3

- IN2P3: Institut National de Physique Nucléaire et de Physique des Particules
  - Un institut du CNRS
- Mission: promouvoir et fédérer la recherche fondamentale en physique
  - But: identifier les constituants élémentaires de la matière, leur interaction et leur comportement
- En chiffres
  - 3300+ personnes (dont 1700+ chercheurs)
  - 18 laboratoires
  - 1 Centre de Calcul situé à Lyon

Centre de Calcul de l'IN2P3

- Services de traitement et de stockage massif de données pour 1800 utilisateurs (IN2P3 + CEA) 24h/24 7j/7
- Services de travail collaboratif
  - Visio-conférence, hébergement de sites web, mail, news, …
- 35+ collaborations internationales en physique de particules, physique nucléaire et astrophysique
- Depuis 2002, ouverture aux applications bio-médicales
- 44 ingénieurs

Centre de Calcul de l'IN2P3



Objectif

- Sélection des projets de construction des infrastructures de grille
  - Très orientée applications scientifiques
  - But commun des projets: construire des infrastructures opérationnelles de grille pour la recherche
    - Ne couvre pas les projets visant à créer les technologies de grille
  - Plusieurs disciplines scientifiques
  - Trois zones géographiques: Amérique du Nord, Asie-Pacifique, Europe (sauf France)
    - Les projets Français seront abordés dans la présentation suivante

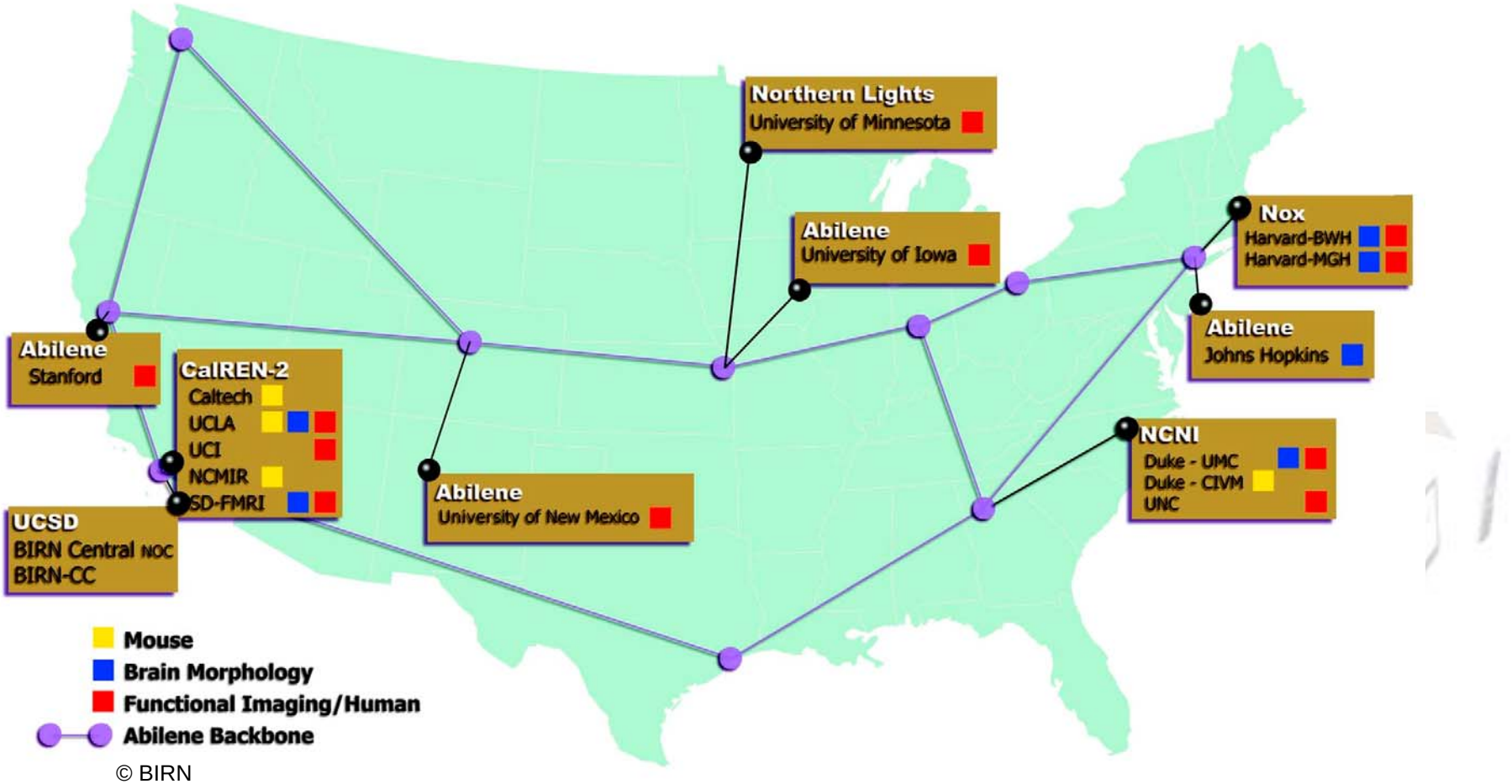
Biomedical Informatics Research Network (BIRN)

- Financé par le *National Institutes of Health* (NIH)
  - 20+ M$
  - Démarrage en 2001
- Objectif
  - Créer un environnement pour faciliter le **partage**, **l'analyse**, la **visualisation** et la **comparaison** des données entre chercheurs en sciences de la vie de plusieurs laboratoires
  - Recherche fondamentale et recherche clinique
- Partenaires
  - 14 universités/hôpitaux, 22 sites
    - Stanford, UCLA, Caltech, Johns Hopkins, …
  - Coordination par Univ. California, San Diego

BIRN (suite)

- Valeur ajoutée
  - Faciliter l'accès à des ressources de calcul et de stockage par des chercheurs et des étudiants à travers des outils intégrés de visualisation, de gestion et d'analyse de données
  - Créer les moyens de gérer et disséminer de grandes quantités de données biomédicales
  - Accélérer la transition entre la découverte scientifique et l'utilisation clinique
  - Créer un environnement de dissémination efficace des données et du logiciel crées avec des fonds publics
  - Encourager la collaboration entre scientifiques de laboratoires et disciplines diverses

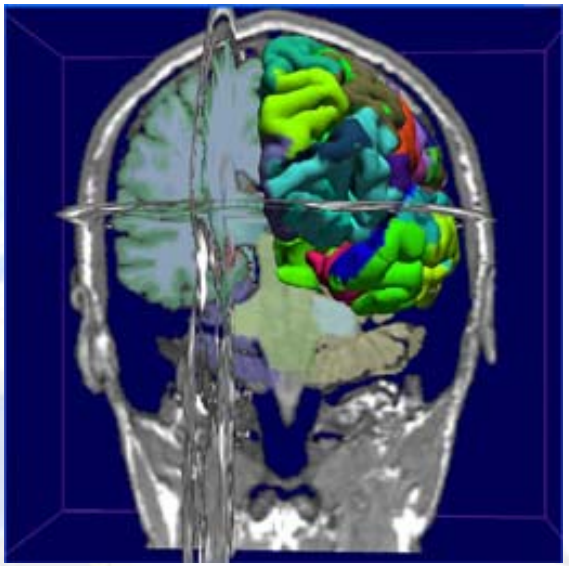
BIRN (suite)



© BIRN

BIRN (suite)

- Une plate-forme de test par sujet de recherche
  - *Morphometry BIRN*
    - corrélations de données patient des plusieurs maladies neuropsychiatriques
    - corrélations entre les différences structurelles du cerveau avec des symptômes tels les dysfonctionnements de la mémoire ou la dépression



© BIRN

BIRN (suite)

- Plates-formes de tests (suite)
  - *Function BIRN*
    - Comprendre les causes de la schizophrénie: déterminer le rôle des lobes frontal et temporel du cerveau dans la maladie
    - Développer de nouveaux traitements
  - *Mouse BIRN*
    - Études des maladies du cerveau chez la souris
    - Schizophrénia, Parkinson, cancer du cervau, abus de substances, sclérose multiple

BIRN (suite)

- Utilisation de l'infrastructure via un portail web
- Technologies utilisées
  - Globus (sécurité, système d'information, gestion des ressources, gestion de données)
  - SRB/MCAT pour la fédération des sources des données et la couche de médiation
  - GridPort (environnement personnalisé: ressources, applications, contrôle d'accès, …)

NPACI

- *National Partnership for Advanced Computational Infrastructure*
- Objectif
  - construire une infrastructure nationale de calcul pour la recherche scientifique
- Vision
  - utilisation de la grille pour la collecte, l'analyse, la visualisation, le partage des données via le web et la publication de résultats via des bibliothèques électroniques
- Financé par la National Science Foundation et coordonné par Univ. California San Diego/San Diego Supercomputer Center

NPACI (suite)

- Partenaires avec différents rôles
  - Fournisseurs des ressources de calcul et stockage
  - Porteurs des applications
  - Partenaires associés
- Exemples d'applications utilisant l'infrastructure
  - Simulation des fluides des systèmes biologiques comme le cœur
  - Structure des protéines
  - Modélisation de puits de pétrole
  - …

NEESgrid

- *Network for Earthquake Engineering Simulation*
- Objectif: utilisation de la grille pour le développement des nouveaux modèles de simulations numériques des tremblements de terre
- Financé par la *National Science Foundation*
- Opérationnelle à partir d'octobre 2004 pendant 10 ans

NEESgrid (suite)

- Partenaires
  - NCSA
  - Univ. Illinois
  - Argonne National Laboratory
  - Univ. California, Berkeley
  - ….

Autres Projets

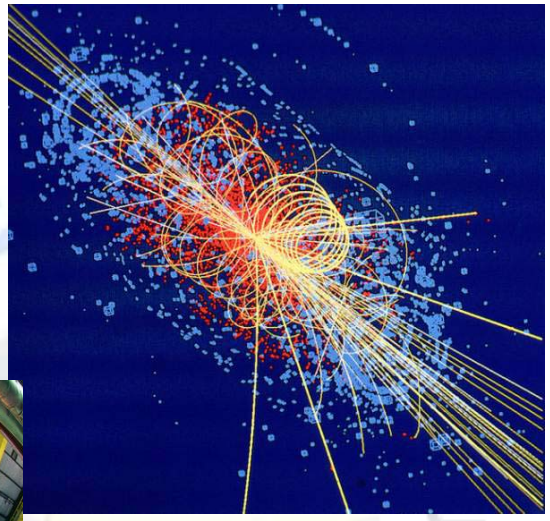
- NASA Information Power Grid
  - Intra-grid, simulation, visualisation, dynamique de fluides, aéronautique
- Grid 2003
  - Physique de hautes énergies, astrophysique, astronomie, biochimie
- TeraGrid
  - Multi-disciplinaire, calcul intensif, stockage de masse, réseau très haut débit
- Telescience
  - Environnement collaboratif de télé microscopie
- AccessGrid
  - Environnement multi-média sur la grille pour faciliter la communication interactive entre groupes de travail

LHC Computing Grid (LCG)

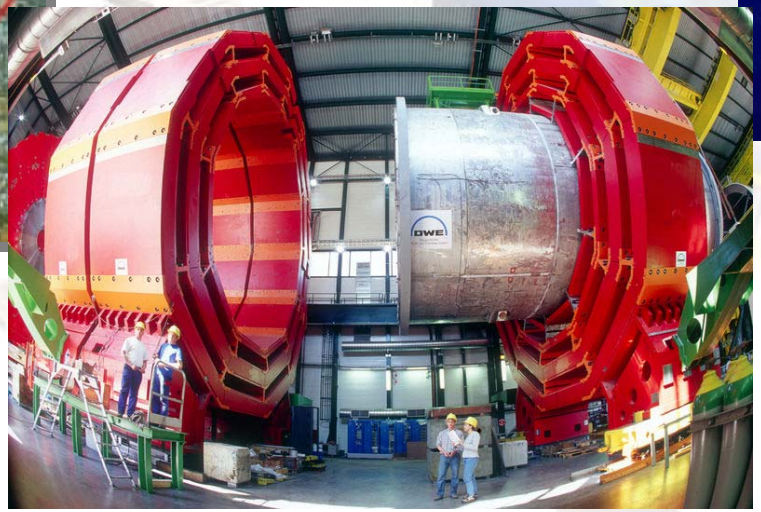
- Objectif: construire l'infrastructure de calcul planétaire pour la simulation et le traitement des données de l'accélérateur de particules *Large Hadron Collider* (LHC)
  - En cours de construction au CERN (Genève)
  - Acquisition des données à partir de 2007
- Besoins
  - 12-14 PétaOctets/an
  - 70.000 CPUs seraient nécessaires actuellement pour traiter ces données

LCG (suite)

© CERN



© CERN

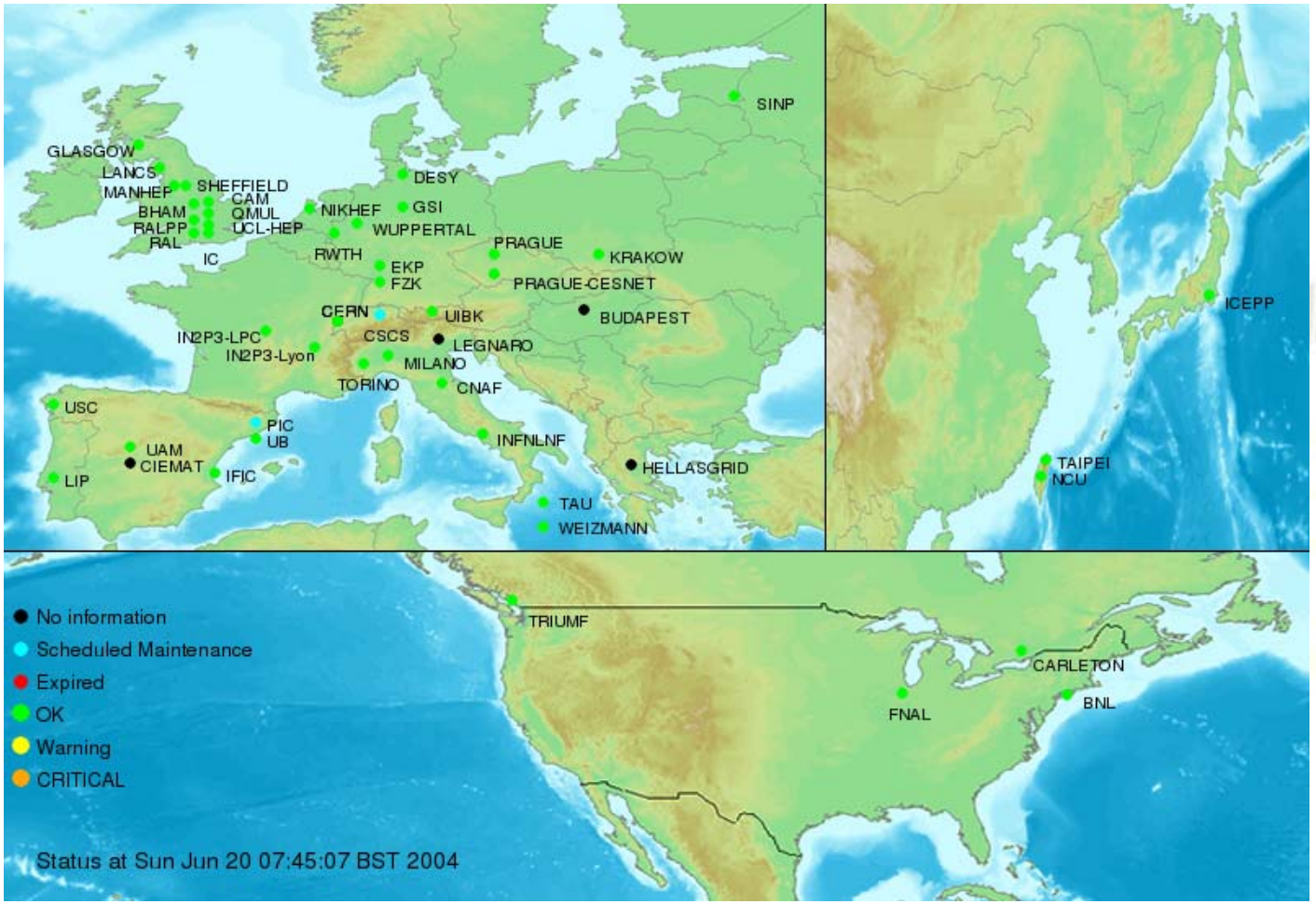


© CERN

LCG (Suite)

- Actuellement en phase I (2002-2005)
  - Développement de l'environnement commun aux applications
  - Développement de l'intergiciel
  - Prototypage des services
  - Déploiement d'une infrastructure de test
    - 40+ sites à travers le monde
    - 3000+ CPUs
  - Plusieurs *Data Challenges* en cours

LCG (suite)


SINP
GLASGOW
LANCS
MANHEP
SHEFFIELD
CAM
QMUL
UCL-HEP
BHAM
RALPP
RAL
IC
NIKHEF
GSI
DESY
WUPPERTAL
RWTH
EKP
FZK
PRAGUE
KRAKOW
PRAGUE-CESNET
CERN
UIBK
BUDAPEST
IN2P3-LPC
IN2P3-Lyon
CSCS
LEGNARO
MILANO
TORINO
CNAF
USC
PIC
UB
UAM
CIEMAT
IFIC
LIP
INFNLNF
HELLASGRID
TAU
WEIZMANN
ICEPP
TAIPEI
NCU
TRIUMF
CARLETON
BNL
FNAL
No information
Scheduled Maintenance
Expired
OK
Warning
CRITICAL
Status at Sun Jun 20 07:45:07 BST 2004

LCG (Suite)

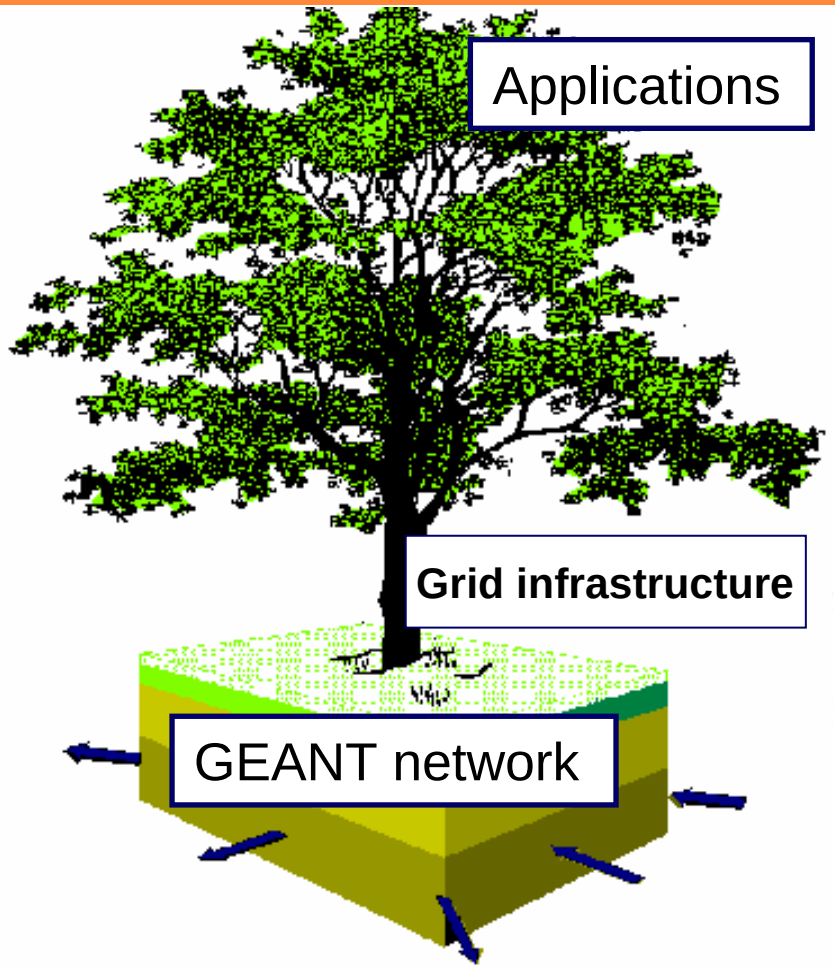
- Technologies utilisées
  - EU DataGRID : R-GMA, Resource Broker, système d'information, infrastructure de certification, services pour les organisations virtuelles, infrastructure de sécurité (authentification, autorisation), …
  - Globus Toolkit
  - Storage Resource Manager (SRM)
  - Surveillance: GridIce, Monalisa,
  - Portail: Genius

EGEE

- *Enabling Grids for E-Science in Europe*
- Objectif
  - intégrer des ressources des grilles nationales, régionales ou européennes afin de construire une infrastructure européenne pour la recherche
  - plate-forme opérationnelle 24j/24 7j/7
- Financé par l'Union Européenne (FP6) à hauteur de 30M sur 2 ans à partir d'avril 2004 (inscrit dans un programme de 4 ans)
  - 70 partenaires, 20 pays
  - France: CNRS, C-S, CEA, CGG, Centrale Paris, …

EGEE (suite)

- Basée sur les infrastructures nationales et européennes
  - Réseau d'interconnexion GEANT
  - Projets de grille (EU DataGRID, Crossgrid, …)



EGEE Organisation (suite)

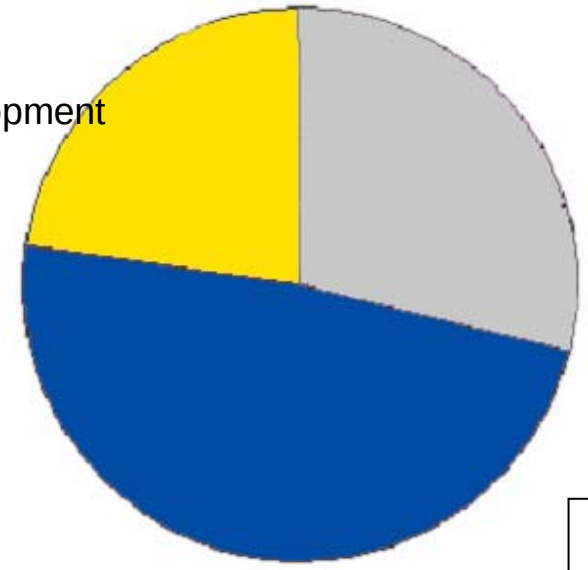
- Partenaires organisés en fédérations

- CERN
- Central Europe (Austria, Czech Republic, Hungary, Poland, Slovakia, Slovenia)
- France
- Germany and Switzerland
- Ireland and UK
- Italy
- Northern Europe (Belgium, Denmark, Estonia, Finland, The Netherlands, Norway, Sweden)
- Russia
- South-East Europe (Bulgaria, Cyprus, Greece, Israel, Romania)
- South-West Europe (Portugal, Spain)

EGEE Activités (suite)

24% Joint Research

JRA1: Middleware Engineering and Integration
JRA2: Quality Assurance
JRA3: Security
JRA4: Network Services Development



28% Networking

NA1: Management
NA2: Dissemination and Outreach
NA3: User Training and Education
NA4: Application Identification and Support
NA5: Policy and International Cooperation

48% Services

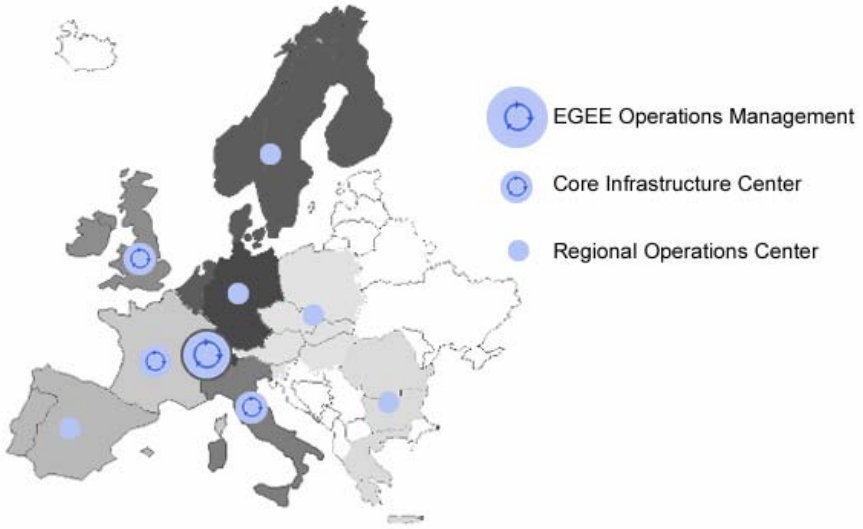
SA1: Grid Operations, Support and Management
SA2: Network Resource Provision

Emphasis in EGEE is on operating a production grid and supporting the end-users

Bob Jones (CERN)

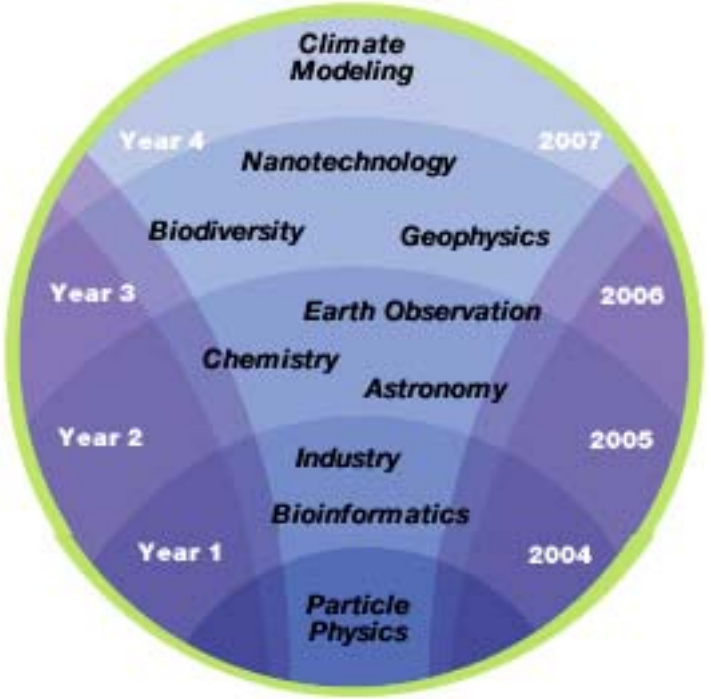
EGEE Activités (suite)

- Activité Service
  - Mission: mettre en place, exploiter et supporter une infrastructure grille de production
  - Organisation
    - Operations Management Center (CERN/CH)
    - Core Infrastructure Center (IN2P3 Lyon/FR, RAL/UK, CNAF/IT, CERN/CH)
    - Regional Operations Center (un par fédération, IN2P3 Lyon en France)



EGEE Applications (suite)

- Cible: les applications scientifiques du monde académique
  - Ouverte aussi au monde industriel
- Mesure du succès du projet
  - combien d'utilisateurs satisfaits de combien de domaines scientifiques différents?
- Applications pilotes: physique de hautes énergies et applications bio-médicales
  - Guider l'implémentation des services
  - Certifier les performances et les fonctionnalités de l'infrastructure
- 500 utilisateurs (3000 à la fin de la 2ème année) d'au moins 5 disciplines différentes



EGEE Applications (suite)

- EGEE met à disposition des applications
  - Une infrastructure à large échelle
    - Des milliers de CPUs et une capacité de stockage en ligne dans la zone des PétaOctets
  - Intergiciel amélioré de façon continue
    - Plus de 3 ans d'expérience avec des tests grandeur nature
  - Equipe d'experts avec expérience sur le déploiement d'applications sur une infrastructure de grille
  - Support utilisateur
  - Formation

EGEE Intergiciel (suite)

- Stratégie: développer un ensemble de services génériques légers, basés sur des composants existants
- Prototypage et cycle de développement rapide
  - Applications participent activement dans le processus
- Approche orientée service
  - Basée sur des standards de Web Services
    - OGSA n'est plus d'actualité
  - WSRF pas encore finalisé
    - Migration vers WSRF à terme

Autres Projets

- DEISA
  - Grille de super-calculateurs en Europe
  - Coordinateur IDRIS (CNRS)
- UK e-Science
  - Multi-disciplinaire: physique, bio-informatique, imagerie médicale, climatologie, …
- IT INFN Grid
  - Programme national, multi-disciplinaire, coordonnée par des laboratoires de physique

Asia-Pacific Grid (ApGrid)

- 15 pays, 49 institutions
- Testbed international déployé
- Application pilote: climatologie

NAREGI

- *National Reseach Grid Initiative*
- Projet de collaboration: secteur industriel, académique et gouvernemental
- But: construire un superordinateur virtuel à 100 Teraflops à l'horizon 2007
  - Plusieurs sites reliés par un réseau à haut débit
  - EarthSimulator: 36 Teraflops

Conclusions

- Plusieurs initiatives de mise en place des infrastructures de grille partout dans le monde
  - Efforts nationaux et internationaux considérables
- Technologie sous-jacente non stabilisée
  - Beaucoup de travail reste à faire pour construire des plates-formes de production, mais beaucoup de progrès a déjà été accompli
- Le modèle de grille a montré ses avantages pour l'intégration des ressources de calcul, instruments et groupes de travail

Pour en savoir plus…

- CC-IN2P3 http://cc.in2p3.fr
- EGEE http://www.eu-egee.org
- LCG http://www.cern.ch/lcg
- BIRN http://www.nbirn.net
- NPACI http://www.npaci.edu
- NEESgrid http://www.neesgrid.org
- Grid2003 http://www.ivdgl.org/grid2003
- TeraGRID http://www.teragrid.org
- Telescience https://telescience.ucsd.edu
- AccessGrid http://www.accessgrid.org
- DEISA http://www.deisa.org
- e-Science http://www.rcuk.ac.uk/escience
- INFN Grid http://grid-it.cnaf.infn.it
- ApGrid http://www.apgrid.org
- NAREGI http://www.naregi.org

Questions