



Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

MINISTÈRE
DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE



Les e-Infrastructures en France

A. Lichniewsky
DGRI/A3

Ministère de l'Enseignement Supérieur et de la Recherche



Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

MINISTÈRE
DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE

PLAN

- Introduction
- Les e-Infrastructures: un peu de typologie
- Les e-Infrastructures en France
- Les infrastructures consacrées aux données scientifiques



Le programme Capacités



Budget 2007-2013: 4,2Md €

Capacités de recherche:

- Infrastructures de recherche
- Recherche au profit des PME
- Régions de la connaissance
- Potentiel de recherche
- La science dans la société
- Soutien au développement cohérent des politiques de recherche
- Activités spécifiques de coopération internationale



Des services de haut niveau pour des scientifiques de haut niveau

Capacités
Budget: 4,2 milliards EUR

Le programme «Capacités» vise à optimiser l'utilisation et le développement des infrastructures de recherche, tout en renforçant les capacités d'innovation des PME et leur aptitude à tirer profit de la recherche. Le programme consistera à favoriser le développement de groupements régionaux axés sur la recherche et à libérer le potentiel de recherche dans les régions de «convergence» et les régions ultrapériphériques de l'UE. Il consistera également à lancer des actions et des mesures horizontales en faveur de la coopération internationale. Enfin, les initiatives du programme «Capacités» devraient permettre de rapprocher la science et la société européenne.

Le programme «Capacités» interviendra dans les sept domaines suivants:

- Infrastructures de recherche
- Recherche en faveur des PME
- Régions de la connaissance
- Potentiel de recherche des régions de convergence
- La science dans la société
- Soutien au développement cohérent des politiques de recherche
- Coopération internationale.

‘Capacités’ dans le cadre du 7^e PC

www.ec.europa.eu/research



Les Infrastructures de recherche



Budget 2007-2013: 1,7Md €

- Rôle de plus en plus important dans les progrès et l'exploitation de la connaissance et de la technologie.
- Demandent un large éventail de compétences
- Demandent à être mutualisées entre scientifiques et avec les entreprises
- Exemples: sources de rayonnement, banques de données en génomique et en sciences sociales, observatoires des sciences de l'environnement et de l'espace, les systèmes d'imagerie, les salles blanches, les infrastructures électroniques de calcul et de communication.



Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

MINISTÈRE
DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE



Les e-Infrastructures

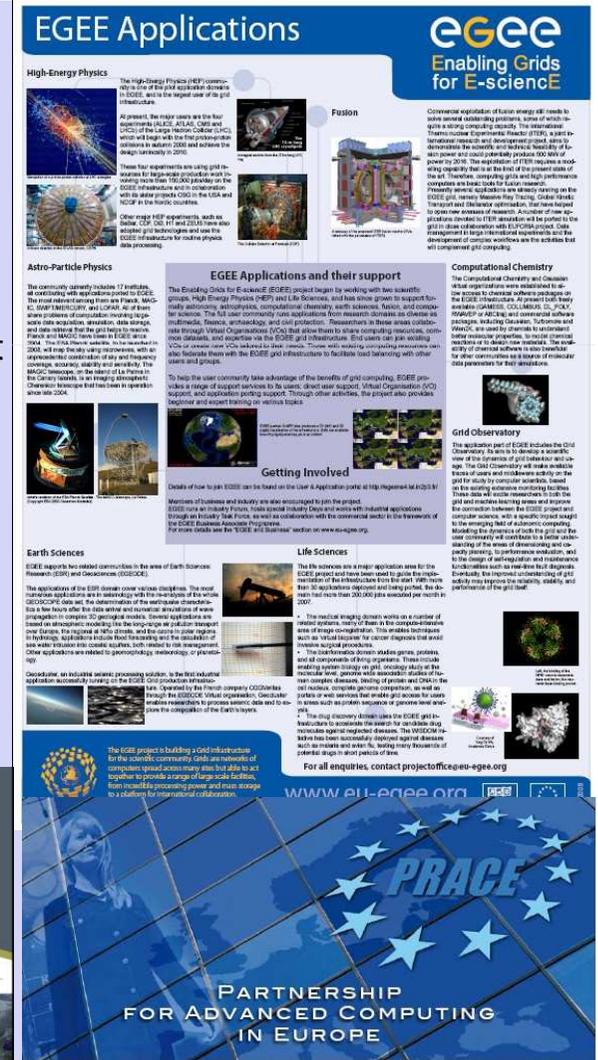
- Terme utilisé depuis 2003 pour décrire une vision **d'infrastructures de recherche distribuées reposant sur les technologies de l'information et de la communication**
- Donnent accès aux chercheurs à des ressources partagées pour la science et l'ingénierie:
 - Sûr et contrôlé
 - Facile et sans raccords
 - Économique

Les Infrastructures de Données Scientifiques

- **Concernent l'appel INFRA-2008-1.2.2 Scientific Data Infrastructure**
- Faire usage des réseaux (GEANT, RENATER) et des Grilles pour développer des domaines comme e-santé, e-learning,....
- Adresse:
 - Exploitation et gestion des données scientifiques: extraire l'information, archiver, conserver
 - Utilisation du contenu numérique en recherche et pour la dissémination des connaissances scientifiques et techniques
 - Le nouvel équilibre à trouver entre les sources de données primaires sous forme numérique et les sources secondaires dont les publications.

Les e-Infrastructures en France: Les infrastructures physiques

- RENATER: le réseau national et la participation dans GEANT
- Les Grilles
 - EGEE: accès aux données du LHC, mutualisation, applications en physique, biologie,...
 - DEISA: grille de supercalculateurs
- La recherche sur les grilles:
 - Grid 5000: la grille de recherche, avec le soutien de l'ACI Grid et du MESR
- Le calcul intensif
 - GENCI: maîtrise d'ouvrage pour les supercalculateurs d'intérêt national. Associés: CEA, CNRS, Universités, Etat (MESR). Maîtrises d'œuvre: CNRS/IDRIS, CEA/CCRT, CINES
 - PRACE: prépare une infrastructure de calcul intensif européenne de « capacité » pour les grands challenge scientifiques et industriels
- Infrastructures régionales et thématiques



EGEE Applications
Enabling Grids for E-science

High-Energy Physics
The High-Energy Physics (HEP) community is one of the most active domains in EGEE, and is the largest user of the grid infrastructure.

At present, the major users are the four experiments ATLAS, CMS, LHCb and ALICE of the Large Hadron Collider (LHC) which will begin with the production of collisions in autumn 2009 and achieve the design luminosity in 2010.

These four experiments are using grid resources for large-scale production and processing of data. The ATLAS and CMS experiments are also participating in the EGEE infrastructure with the production of an atlas project (ATLAS) in the USA and HEP in Europe.

Other major HEP experiments, such as Belle, CDF, D0, H1 and ZEUS have also adopted grid technologies and use the EGEE infrastructure for routine physics data processing.

Astro-Particle Physics
The community already includes 17 institutes all contributing with applications in EGEE. The most relevant ones are from France: MGC, ICGF, IMCCP, and ICGF. It also includes other institutes contributing to large scale data acquisition, data analysis and simulation. The EGEE infrastructure is used for the production of an atlas project (ATLAS) in the USA and HEP in Europe.

EGEE Applications and their support
The Enabling Grids for E-science (EGEE) project began by working with two scientific domains: High Energy Physics (HEP) and Life Sciences, and has since grown to support many other scientific domains, including astronomy, astrophysics, computational chemistry, earth sciences, fusion, and computer science. The EGEE community also supports from research domains as diverse as multimedia, finance, homeland, and civil protection. Researchers in these areas collaborate through Virtual Organizations (VO) that allow them to share computing resources, data, tools, and expertise via the EGEE grid infrastructure. EGEE users can join existing VOs or create new ones. EGEE also provides a range of support services to its users: direct user support, Virtual Organization (VO) support, and application profiling support. Through other activities, the project also provides beginner and expert training on various topics.

Computational Chemistry
The Computational Chemistry and Chemistry user organizations were established in 2004. The EGEE infrastructure is used for many scientific applications, including molecular dynamics, quantum chemistry, and materials science. The EGEE infrastructure is used for many scientific applications, including molecular dynamics, quantum chemistry, and materials science.

Grid Observatory
The Grid Observatory (GO) is a project of EGEE that aims to monitor and analyze the usage of the EGEE infrastructure. The GO community will make available data on the usage of the EGEE infrastructure for the grid for study by computer scientists. The GO community will make available data on the usage of the EGEE infrastructure for the grid for study by computer scientists.

Getting Involved
Details of how to join EGEE can be found on the user & application portal at <http://www.egee.org>.

Members of business and industry who are encouraged to join the project. EGEE offers an Industry Forum, holds special Industry Days and works with industrial applications through an Industry Task Force, as well as collaboration with the commercial sector in the framework of the EGEE Business Associates Programme. For more details see the "EGEE grid business" section on www.egee.org.

Earth Sciences
EGEE supports the related communities in the Earth Sciences: Research (ESR) and Observations (EGEO).

The applications of the ESR domain cover various disciplines. The most prominent applications are in geology with the production of an atlas project (ATLAS) in the USA and HEP in Europe.

Life Sciences
The life sciences are a major application area for the EGEE project and have been used to guide the implementation of the infrastructure from the start. EGEE has more than 30 applications domains and is being used by the community to support a wide range of scientific research.

The EGEE project is building a Grid infrastructure. For the scientific community, this is an essential of computer spread across many sites but able to act together to provide a range of large-scale facilities, from research processing power and data storage to a platform for international collaboration.

For all enquiries, contact project@ceae-egee.org

www.eu-egee.org

PRACE
PARTNERSHIP FOR ADVANCED COMPUTING IN EUROPE

Les nouveaux enjeux

- Il faut maintenant construire par-dessus les infrastructures physiques:
 - Des données à l'information
 - Valider, standardiser, calibrer
 - Gérer, conserver, partager
 - Donner un sens et permettre l'exploration
 - La capture des données
 - Les collaborations
 - Les services aux utilisateurs

Les e-Infrastructures nationales: la prospective

- Sous l'impulsion du MESR, lancement en 2008 de prospectives et de coordinations
 - **Grilles de production**
 - Prospective interorganismes en cours
 - Protocole d'accord en cours de finalisation
 - Objectif: assurer la coopération et envisager la représentation française dans les projets de e-Infrastructure européens.
 - **Grille de recherche**
 - Protocole d'accord inter-organismes

Infrastructures de données: l'implication française

- Cet exposé ne vise pas à l'exhaustivité, mais essaie de montrer quelques exemples significatifs

Données spatiales et astronomie		Univ. Strasbourg, CNRS
Climat		IPSL (CEA, CNES, CNRS, IRD, ...), CERFACS
Observation de la terre et ressources naturelles		ESA, ERCIM
		ESA, CNES, CS
Bio diversité	 	MNHN, IRD
Biologie		INRIA

Contenus: l'implication française



Le projet a réalisé des études sur:

- les archives en Europe,
- les technologies,
- la curation
- la préservation des données.

Actuellement, un service pilote est en construction pour démontrer les possibilités offertes par une infrastructure européenne des archives.

Il fournira un ensemble de recommandations et de services (Service Oriented Architecture) qui permettront de connecter les réseaux nationaux et de constituer une infrastructure européenne d'archives ouvertes.

Les données en biologie



IMproving Protein Annotation through Coordination and Technology

- **But:** permettre à l'utilisateur de rechercher efficacement les propriétés des protéine à travers les génomes et protéomes . Il s'agit d'utiliser les sites fonctionnels pour prédire les fonctions et les structures des protéines.
- Une difficulté est que les bases de données de protéines ne partagent pas les mêmes formats et la même nomenclature. Chacune a ses avantages et ses inconvénients.
- Outils
 - Base de données (InterPro)
 - Logiciel pour extraire des propriétés pertinentes pour la recherche et sciences de la vie
- La définition et l'adhésion à un standard commun permettant l'échange d'informations est au centre du projet IMPACT

Données et Bio diversité

MINISTÈRE DES AFFAIRES ÉTRANGÈRES
MINISTÈRE DE L'AGRICULTURE ET DE LA PÊCHE
MINISTÈRE DE L'ENVIRONNEMENT ET DU DÉVELOPPEMENT DURABLE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

MUSEUM
NATIONAL
D'HISTOIRE
NATURELLE



Systeme mondial d'information sur la biodiversité



- L'accès aux données sur la biodiversité est une nécessité scientifique et politique.
- Des points nodaux nationaux sont mis en place dans chacun des pays membres.
- Le GBIF France a vocation de rassembler toutes les données hébergées en France, que celles-ci concernent la biodiversité du territoire français ou du reste du monde.
- Le point nodal français joue un rôle de soutien aux pays francophones du Sud.



Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

MINISTÈRE
DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE

Données et Bio diversité



La stratégie du GBIF

- Accès aux données et interopérabilité des bases de données pour l'exploitation systématique des données
- Digitalisation des collections d'histoire naturelle, pour accroître les connaissances sur la biodiversité disponibles sur Internet
- Catalogue électronique des noms d'organismes connus, pour améliorer la recherche de données sur la biodiversité et permettant de combiner les données de différentes disciplines
- Sensibilisation et formation
- Banque d'espèces (SpeciesBank)
- Ressources électroniques en littérature sur la biodiversité (Digital Biodiversity Literature Resources)

Les données pour observer la Terre

*Distributed colLaboratories
Infrastructure on Grid ENabled
Technology 4 Science*



Le projet D4Science

- déploiera les infrastructures construites par les projets EGEE et DILIGENT
- Pour satisfaire les besoins des communautés de “Surveillance de l’Environnement” et de “Gestion des Ressources pour la Pêche et l’Aquaculture”

Les données pour l'astronomie



L'Observatoire virtuel est une initiative internationale de la communauté astronomique.

Son objectif est

- De donner l'accès électronique aux archives de données astronomiques, aux observatoires, et aux bases de données décrivant le ciel.
- De permettre l'analyse de données de sources multiples à travers une coordination qui mettra en place les standards, un réseau à grand débit et des outils d'analyse à l'état de l'art.

L'observatoire virtuel étend le champ d'investigation scientifique:

- Données de sources multiples (longueur d'onde)
- Etude d'un grand nombre d'objets

SHS et données



Objectif

- Accès sécurisés aux données de la statistique publique et aux grandes enquêtes socio-économiques

Organisation

- Une organisation en réseaux INSEE, CNRS, INED, EHESS, Sciences Po. dépendant d'un comité interministériel
- Appui à trois infrastructures européennes « ESFRI » : CESSDA, ESS, SHARE

La recherche sur les bibliothèques numériques



Le réseau DELOS travaille à intégrer et coordonner les recherches sur les Bibliothèques Numériques, pour développer la prochaine génération.

Les objectifs sont:

- Définir des théories et les cadres applicatifs supportant le cycle de vie de l'Information Numérique
- Construire des services multi-langues et multi-mode et des concepts de gestion intégrée des contenus.
- Animer le réseau de la recherche européenne sur les bibliothèques numériques



Sources d'information utile

- e-IRG:e-Infrastructure Reflection Group
 - <http://www.e-irg.org/>
- Standards
 - <http://www.beliefproject.org/e-infrastructures-as-standardisation-drivers/view>
 - <http://www.etsi.org/WebSite/NewsandEvents/ECConcertation/ECConcertationDaysHome.aspx>
- Conservation de données scientifiques
 - <http://hla.stsci.edu> Hubble Legacy Archive
- EGI Base de connaissances
 - http://knowledge.eu-egi.eu/index.php/European_Projects
- DCC: data curation center
 - http://www.dcc.ac.uk/events/events_archive

- Ressources naturelles
 - <http://www.d4science.eu/consortium>
- Rhone Alpes Bioinformatics Center
 - <http://www.prabi.fr/>
- DRIVER
 - <http://www.driver-repository.eu/>
- Climat
 - <http://www.metaforclimate.eu/> -- metadata for climate -- CERFACS IPSL
- Observatoires virtuels
 - <http://www.euro-vo.org/pub/>
 - <http://www.ivoa.net/>
 - <http://www.genesi-dr.eu/>
- GBIF
 - <http://www.gbif.org/>
 - <http://www.gbif.fr/>
- Bibliothèques numériques
 - http://www.delos.info/index.php?option=com_content&task=view&id=618&Itemid
- EGEE
 - <http://www.eu-egee.org/>
- DEISA
 - www.deisa.eu
- PRACE
 - [/www.prace-project.eu](http://www.prace-project.eu)
- Réseau Quetelet
 - <http://www.centre.quetelet.cnrs.fr/>