Earth Sciences Applications for EGEE

IPSL: **M. Petitdidier**, C. Boonne, O. Marti, **CGG**: D. Thomas,

DKRZ: J. Biercamp, **ESA-ESRIN**: L. Fusco, J. Linford,

IPGP: JP Vilotte, G. Molguiny, **KNMI**: W. Som de Cerff

Outline

- Introduction
- Earth Sciences: Research
  - Community
  - Grid Awareness
  - Applications
  - Material means devoted to Grid activity
  - Added value for EGEE and ES communities
- Earth Sciences: Industry
  - Community
  - Grid Awareness
  - Applications
  - Material means devoted to Grid activity
  - Added value for EGEE and ES communities
- VO coordination and interaction among ES partners
- Requirements
- Conclusion

Introduction

- Earth Sciences:

  Solid Earth, Atmosphere and Ocean, Interface between them

- Applications proposed for EGEE:
  - Earth Observations by satellite
  - Climate
  - Solid Earth Physics
  - Geophysics : industrial applications
- User perspective: research and industry
  - application objectives
  - SE and CE security and access: EGEE MW
  - Requirements : data and metadata policy and access
- As a starting point in EGEE :
  - two groups or VOs proposed: ES Research, ES Industry (Geophysics)

EARTH SCIENCES: Research

- COMMUNITY

Earth Observations (EO): ESA(IT), IPSL(FR), KNMI(NL), UTV(IT), SRON(NL), RIVM(NL)

Climate: DKRZ (DE), IPSL(FR), KNMI(NL) (for information)

Solid Earth Physics : IPGP(FR)

- In the same Institute or location, teams with different applications but sharing the same human and material means for EGEE.
- Scattered resources for a given project, many copies of data…..

Ozone and EO: increase of the number of teams due to demonstration in DataGrid and extension to other satellite data and products

Climate: well defined community has worked for many years via EC projects

Solid Earth Physics: several communities centred on different topics and projects

Grid Awareness

DataGrid: ESA, KNMI, IPSL, IPGP (joined later)

DKRZ: German Grid Initiative (« D-Grid »)

UTV: work with ESA

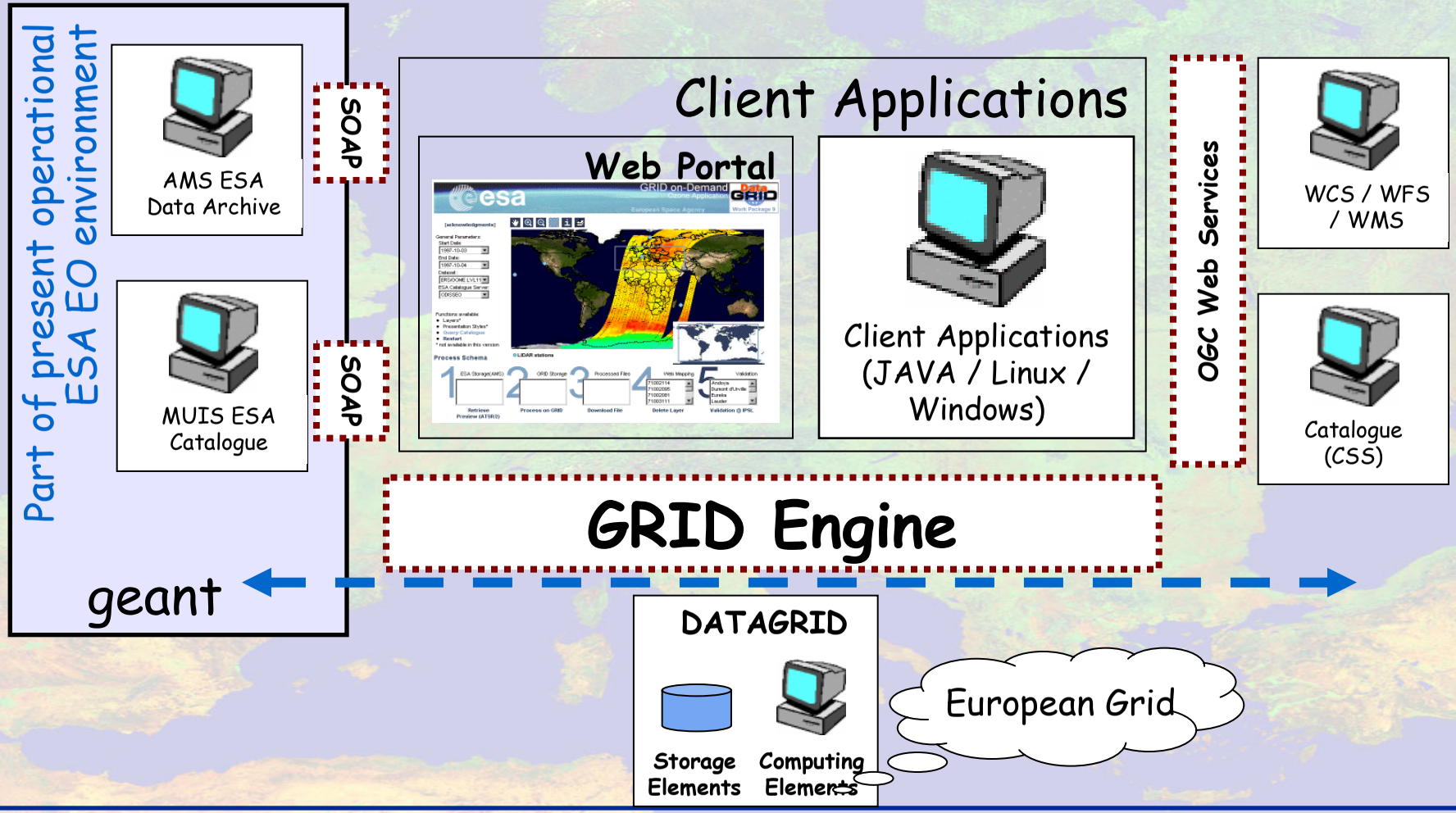
DataGrid experience

- Demonstrate how Grid infrastructure can respond to the complexity and constraints imposed by applications in the EO domain
  - By deploying EO applications on the EDG middleware (test and evaluate)
  - Development of EO Grid application interfaces and tools
- Dissemination & promotion of this new technology to other scientific teams, satellite operational services and private companies
  - Participation to non-grid oriented workshops, demonstrating Grid
  - Publication of results obtained on Grid in scientific papers

Requirements - Achievements

- Handling of large number and large volume of files from different satellite instruments (Gome, Gomos) and ground measurements
  - Production of about 70,000,000 of Ozone profiles on the Grid stored by orbit (2000 profiles/orbit) and a smaller part by individual files. Not all of them are registered.
- Processing with complex algorithms (e.g. Neural approach, inversion approach) using IDL(Interactive Data Language)-runtime
- Creation and secure access to metadata catalogues and data
  - Satellite and metadata catalogue
  - Access with security and role to data and metadata
- Interface Grid MW with the already developed operational satellite tools and infrastructure e.g. ESA proprietary product catalogue and Archive system

INTERFACE WITH OPS TOOLS:
MUIS (Multimission User Information Service) and AMS (Archive Management System)



EO Applications in EGEE

Extend the Ozone experience:

Partners: KNMI, IPSL, SRON, RIVM, Univ. TorVergata, ESA, ...

Goals: Use of "operational Grid" by the Ozone community

⇒ To produce and/or store the retrieved Ozone profiles or columns on the Grid

⇒ To store on the Grid the Ozone databases, old and new ones

⇒ To extend the processing capabilities (assimilation, new parameters, short term prediction, climatology, process studies...), and obtained new scientific results

⇒ To facilitate the collaboration and communication among the community

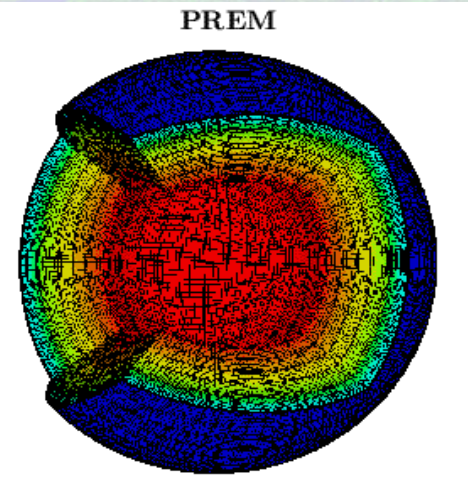
⇒ Important for validation

⇒ to exchange data in the emerging large scale European projects (GMES)

Additional Ozone and Other data sets

- Maintain/extend data availability on the Grid
  - all the level1/2 data and reference data for level 2 (Lidar, Radiosounding)
- Generate level 3 data on the Grid
  - Assimilation of Ozone data into models (UK), Study of Ozone holes….
- Extend GOME experience to other datasets
  - Envisat Sciamachy, Gomos….
  - Reprocessing with new algorithm and comparison
- Training of neural network algorithms and validation
  - Ozone, Oil spill detection
- Other Satellite data: Envisat ASAR, MERIS, …
  - build on ongoing ESA funded projects
  - Integration of science tools and application following community interests
- ESA will provide data according to ESA policy and infrastructure for the Research VO

Solid Earth Physics Applications in EGEE



- Partner: IPGP, France
- Objectives : demonstration to drive the community, and production of scientific results
- GPS data:
  - Access to external database
  - Integration of existing tools working under Unix
  - final goal: workflow with data storage, processing, analysis and visualisation
- Synthetic seismograms
  - Simulation: propagation of waves in the Solid Earth
  - Numerous data and computations, access to databases
- Earth Core dynamo
  - Test started in DataGrid by accessing a cluster

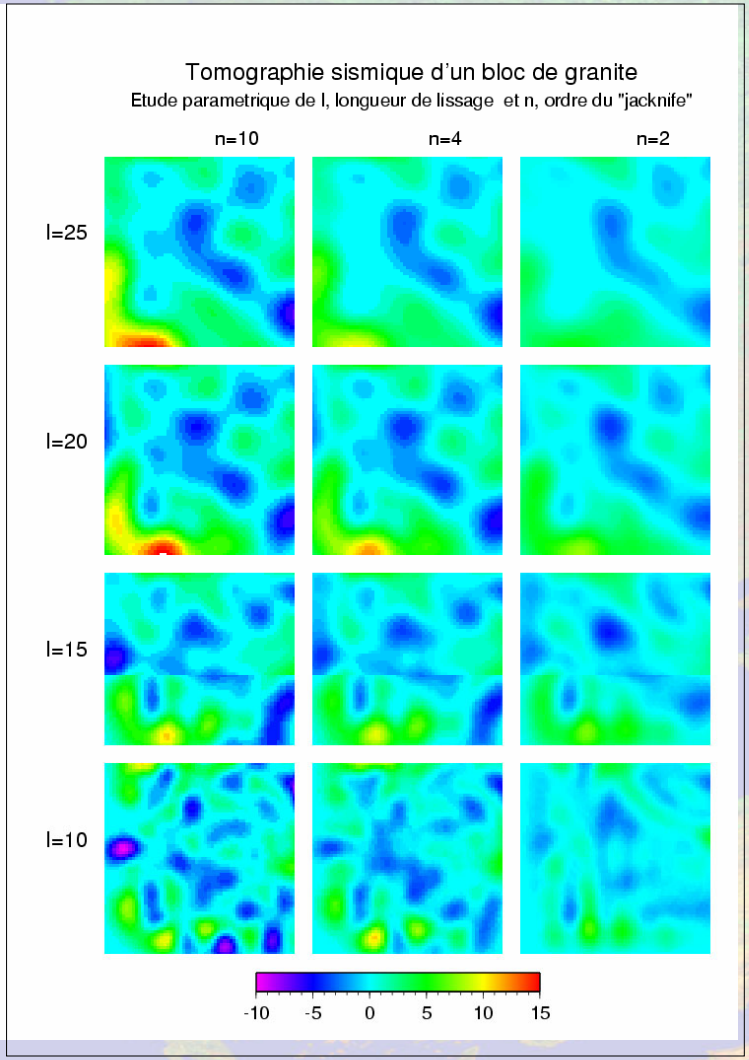
Solid Earth Physics Application

- Example obtained with DataGrid (A. Nercessian, G. Moguilny)

General tomography example (numerical integration + solving of a linear symmetric system by Choleski method)

110 jobs:

Minimum time: 0.22 s
Maximum time: 2 h 37 m 58 s
Total : 77 h 21 m 19 s
Average time value : 0 h 42 m 34 s



- Strategy
  - Demonstrate the secure and restricted access to database
  - Propose tests inside EU project like SPICE
  - Obtain scientific results to constitute databases and propose to the concerned community access via the Grid

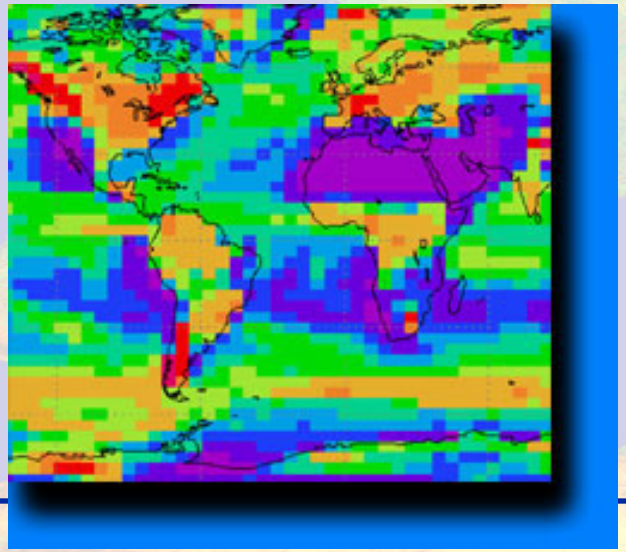
Climate Applications in EGEE

Model: Atmosphere, Ocean, Hydrology, Atmospheric and Marine chemistry….

European effort: PRISM, ENES, GEMS

Goal: Post treatment: Comparison of model outputs from different runs and/or institutes

- Characteristics of model outputs :
  - Large volume of data (TB) from different model outputs, and experimental data
  - Run made on super computer
- Post-treatments: visualisation, statistics, comparison….



EXAMPLE: For the IPCC Assessment reports many experiment are performed with different models (different spatial resolution, different time-step, different "physics" ..) and various sites.

The generated date need to be compared in a comprehensive and "unified" way.

Climate : Strategy and Perspective

- Strategy
  - Secure access to the mass storages where the data are stored (like the ESA portal) (prototype by IPSL+CEA in discussion)
  - Metadata Catalogues implemented on the GRID (Spitfire or RMC)
  - Access to models and data
  - Demonstration between two institutes (DKRZ, IPSL)
  - extend to the other institutes
- Perspective
  - Extension to other European laboratories linked via European projects(PRISM, ENES, GEMS) to develop coupled model, and to compare them.
  - Pilot studies to run models (e.g. ensembles) on the Grid
  - Make modelling frameworks (PRISM) grid aware
    => Link the EGEE infrastruture with super computer Grids (DEISA)

Material Means devoted to Grid activity

Starting point:

- ESA: UI, CE (15 nodes), SE (1.4 TB)
- IPSL+IPGP at Paris University Computer Center : 4PC, SE (500Gb), UI
- IPGP: UI
- DKRZ: UI, CE (2nodes), SE up to several TB as a function of the application
- KNMI: UI + possibility to use VO NIKHEF and Sara facilities for the Research ES

According to the applications ported new material will be devoted to Grid

Added Value for EGEE and ES Communities

- EGEE
  - Deployment of the Grid in Europe among research and operation (ESA)
  - Dissemination of Grid expertise
  - Interaction with DEISA via Climate applications
- ES Communities
  - Exchange and Sharing of large sets of data
  - Access to powerful computing and storage facilities
  - Collaboration, Coordination and Communication among the community partners

Earth Sciences: Industry

- Community:

Virtual Organization to share IT resources and best-practices. Opened to all Research centers in environmental geophysics from both Industrial and Academic world, working on two critical issues: Energy for today and tomorrow and CO2.

- Grid awareness

CGG: initial experience in Grid

Official partner of EGEE SA1 and Industry Forum

- Material Means

CE: 100 nodes, SE:100GB

Applications

Seismic processing Generic Platform:

- Based on Geocluster, an industrial application – to be a starter of the core member VO.

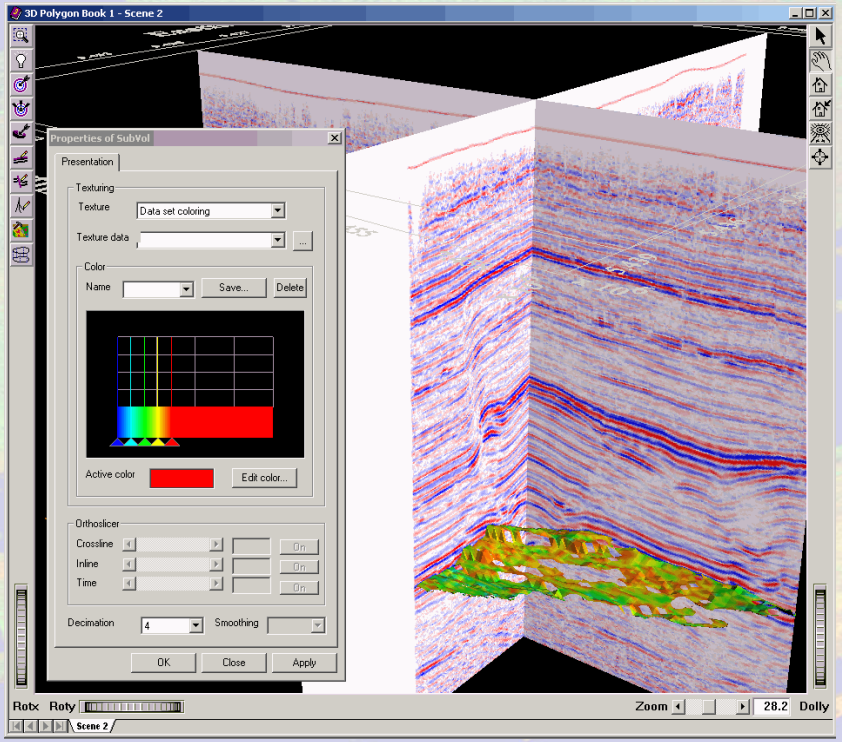
- Include several standard tools for signal processing, simulation and inversion.

- Opened: any user can write new algorithms in new modules (shared or not)

- Free for academic research

-Controlled by license keys (opportunity to explore license issue at a grid level)

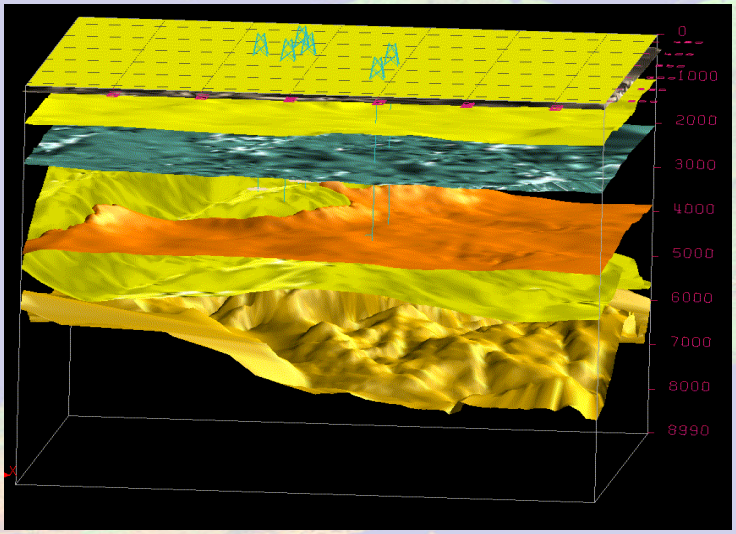
- initial partners F, CH, UK, Russia, Norway

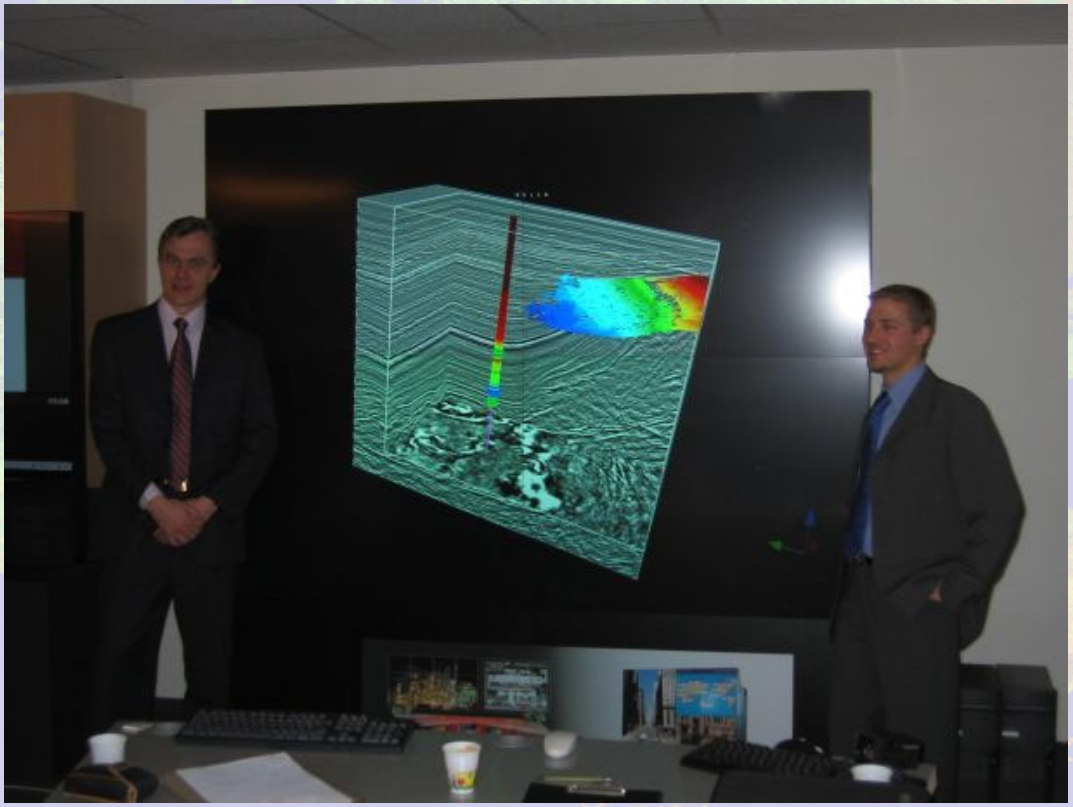


Applications

Imaging in geosciences: a typical cpu/data intensive application

- Possible collaboration with Russian organization (Novosibisrk)
- explore solutions for // systems, MPI, complex workflows





Added Value for for the EGEE and Geophysics Community

- EGEE
  - Geophysics is a key technology for earth sciences, it shares and complements requirements of actual EGEE applications. "Geophysics" community is large (thousands of researchers across Europe) but very scattered. EGEE will benefit in enabling such a community to collaborate and progress on critical issues Energy and CO2.
  - One of the applications is an industrial application which will be a reference for other industries and will help to develop and support credibility of EGEE infrastructure.
- Geophysics Community
  - Capability to solve complex problems and to validate innovative algorithms on real size data sets
  - Close the gap between Research and Industrial environment
  - Attract and keep brightest researchers

VO Coordination and Interaction Among ES partners

- VO Coordination

Research: Monique Petitdidier (IPSL)

  - NIKHEF and KNMI to maintain the research VO

Industry : Dominique Thomas (CGG)

- Interaction between ES partners :

Sharing of Scientific, technical and Grid Expertise, Requirements

Meetings in Paris with IPSL, IPGP and CGG + other partners

Phone Conference

Requirements

- Continuation of EO Virtual Organization by ES and Geophysics VOs
  - New Applications proposed in EGEE
- Improvements:
  - data access (restriction), the replica metadata catalogue
  - VOMS
- Emerging new requirements :
  - Porting and query of external metadata catalogue (compatibility)
  - Update and mirroring of database (security, integrity)
  - Start « operational use » of middleware components
  - Move algorithms and data analysis tools close to data repositories
  - Consider WEB services developments in Open GIS consortium
  - Integration of e-collaboration tools and technologies

Conclusion

- ES applications ready to be ported on EGEE
  - Some applications already ported on the Grid (EO, Solid Earth Physics)
  - Other applications on the way or in discussion among partners
- Deployment of ES on the Grid driven by goals to attain:
  - In EU projects, even if Grid not explicitly mentioned
- Meetings for dissemination
  - EOGEO (satellite operational and research Community) 23-25 June 2004 in London with the presence of Kyriakos Baxevanidis (EC) in order to define outlines for next EU call for tenders
  - Climate meeting about the application to be ported on the Grid organised by EGEE partners: DKRZ, IPSL
- Interaction with CrossGrid partners working on same topics (pollution, meteorology, flood prediction…)