egee

Enabling Grids for E-sciencE

EGEE-II Related projects: Infrastructure & Application

Ian Bird
*CERN IT-GD*

IT Seminar

13th March 2006

Information Society

eGee

Enabling Grids for E-sciencE

- **EGEE infrastructure (middleware distribution and operations) was built up during 18 months prior to the start of EGEE by the LCG project**
  - The LCG work formed the basic infrastructure of EGEE
  - The middleware distribution retained this name (LCG-2.x) as it was expected to be replaced by gLite
  - Now the middleware distribution will *evolve* with additional or replacement services coming from gLite or elsewhere

- **EGEE started in April 2004 with a running grid infrastructure**
  - 40 sites, 3000 CPU
  - Basic operations
  - Developed certification and deployment process

- **Now expanded to:**
  - 200 sites, >20 000 CPU, 40 countries
  - Managed operations – stability of sites
  - >10 000 jobs / day sustained over the last year

Sites

CPU

Jobs/day
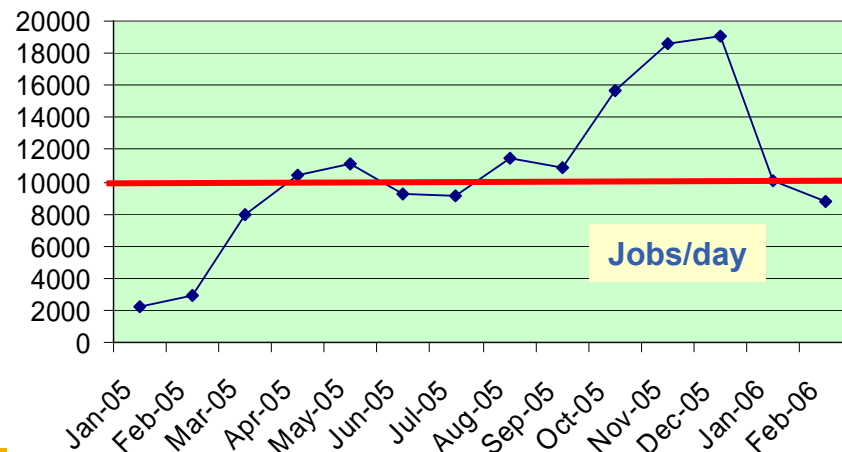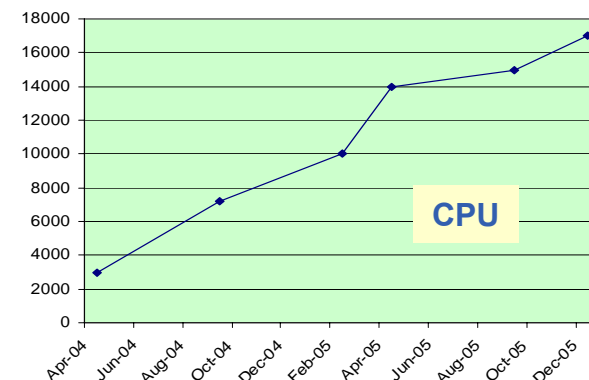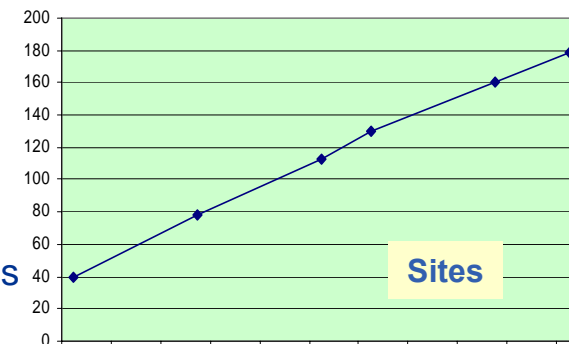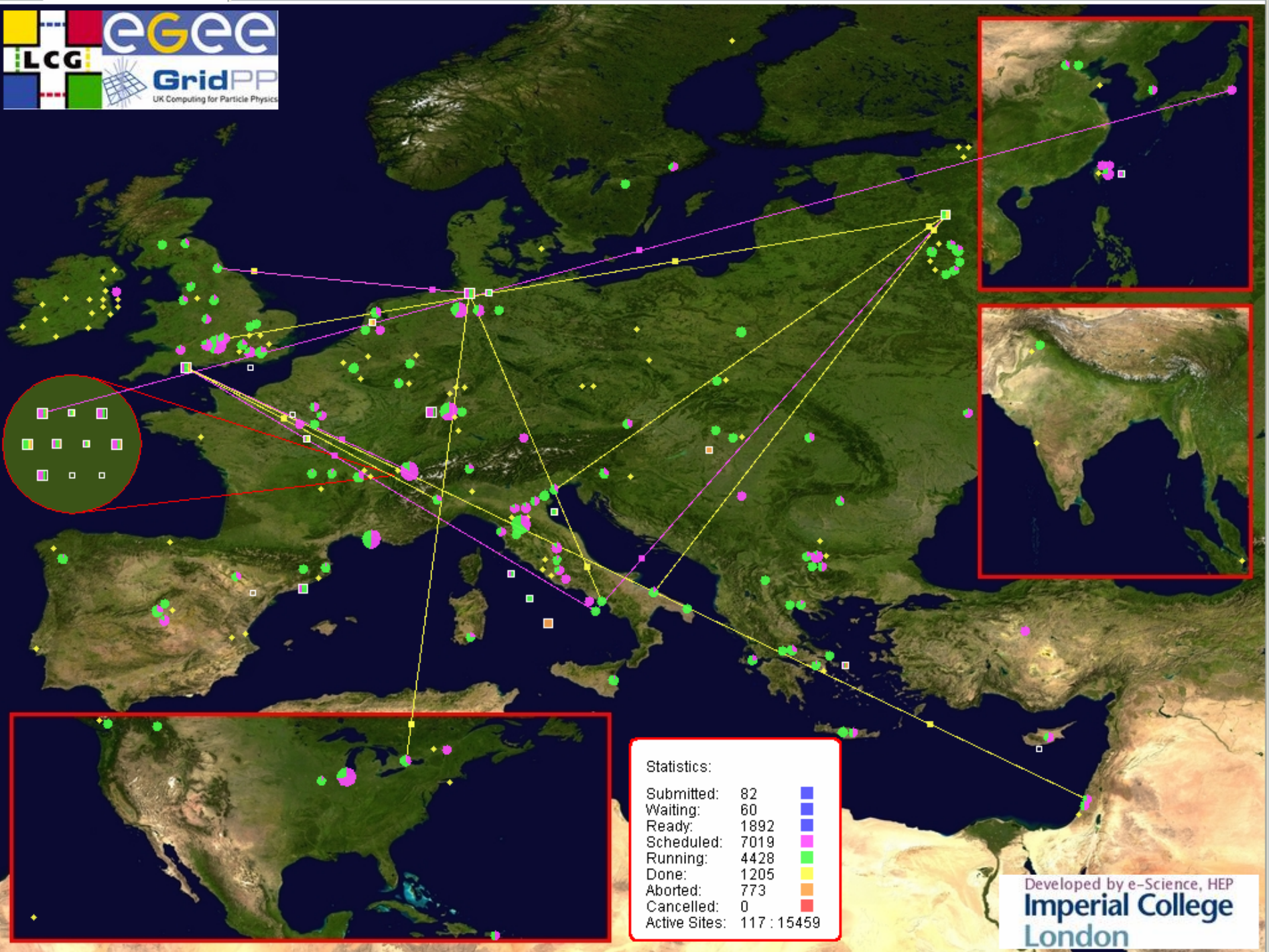
Statistics:

Submitted:	82	
Waiting:	60	
Ready:	1892	
Scheduled:	7019	
Running:	4428	
Done:	1205	
Aborted:	773	
Cancelled:	0	
Active Sites:	117 : 15459	

Developed by e-Science, HEP
Imperial College
London

- **Grid Deployment Group**
  - EGEE SA1: Operations and management
    - Includes testing, certification, release preparation, deployment, and support
    - IN EGEE-II this is a separate activity: SA3
  - IT-GD-OPS: SL Maite Barroso – Grid Operations (SA1)
  - IT-GD-ITR: SL Markus Schulz – Integration, Certification, deployment (SA3)
  - SA1: Activity Leader: Ian Bird
  - SA3: Activity Leader: Markus Schulz
  - ETICS: Leader Alberto Di Meglio (within ITR section)
    - ETICS provides build and test infrastructure to be used in EGEE-II
  - IT-GD-SC: SL Jamie Shiers – Coordination of the LCG service
- **PSS Group:**
  - NA4: Applications
- **EGE Group:**
  - Project Office, Project management, dissemination.

Enabling Grids for E-sciencE

- **Infrastructure Projects**
  - SEEGRID(-2), BalticGrid, EELA, EUChinaGrid, EUMedGrid
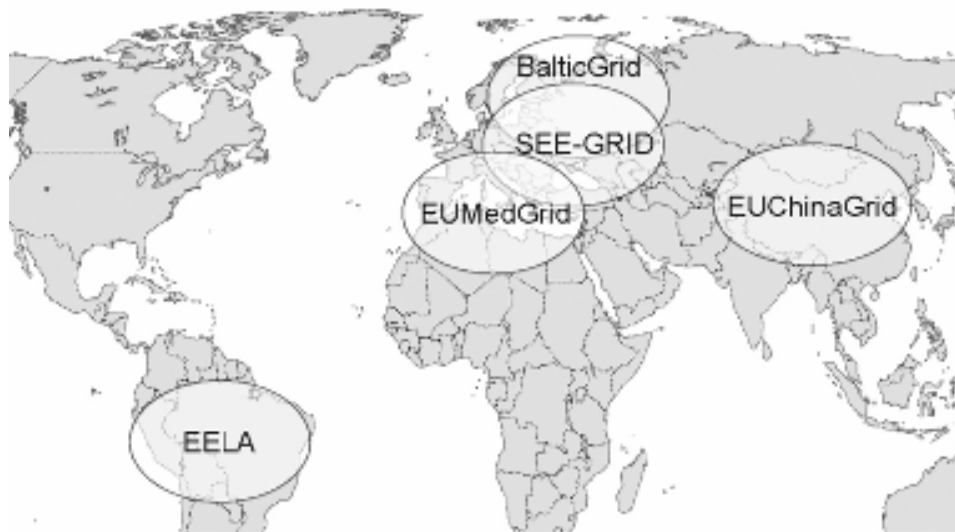
- **Application Project**
  - Health-e-Child

Enabling Grids for E-sciencE

- **As coordinating partner for EGEE Infrastructure and Operations (SA1)**
  - CERN (IT-GD) is asked to provide technical coordination and support to the related infrastructure projects
    - We want them to appear as much a part of the EGEE infrastructure as possible
  - This was very successfully prototyped with SEE-Grid
    - PJAS assigned to GD, worked as part of operations team
    - GD ensures that project gets 1 FTE of support, help with technical issues, planning etc.
- **Continue with this model:**
  - Each project contributes funding for 5 PJAS + 1 LD to work in GD as part of the operations, certification, deployment teams providing support to these projects
    - The LD will provide overall technical coordination of these and work on some of the planning/reporting tasks

Infrastructure Project: SEEGRID

Enabling Grids for E-sciencE

Aim: *To provide specific support actions to assist the participation of South Eastern European states to the pan-European Grid initiatives.*

Applications: *HEP, Biomedical, Earth Science, Regional: 3D imaging, On-demand web crawling*

Coordinator: *GRNET, Greece*

Participating countries: *Albania, Bosnia-Herzegovina, Bulgaria, Croatia, FYR of Macedonia, Greece, Hungary, Romania, Serbia-Montenegro, Switzerland, Turkey*

Start date: *May 2004 for 24 months*

http://www.see-grid.org/

Project partners: The regional dimension

Contractors

GRNET	Greece
CERN	Switzerland
SZTAKI	Hungary
IPP-BAS	Bulgaria
ICI	Romania
TUBITAK	Turkey
INIMA	Albania
BIHARNET	Bosnia-Herzegovina
UKIM	FYR of Macedonia
UOB	Serbia-Montenegro
RBI	Croatia

Third Parties

18 universities / research centres identifie
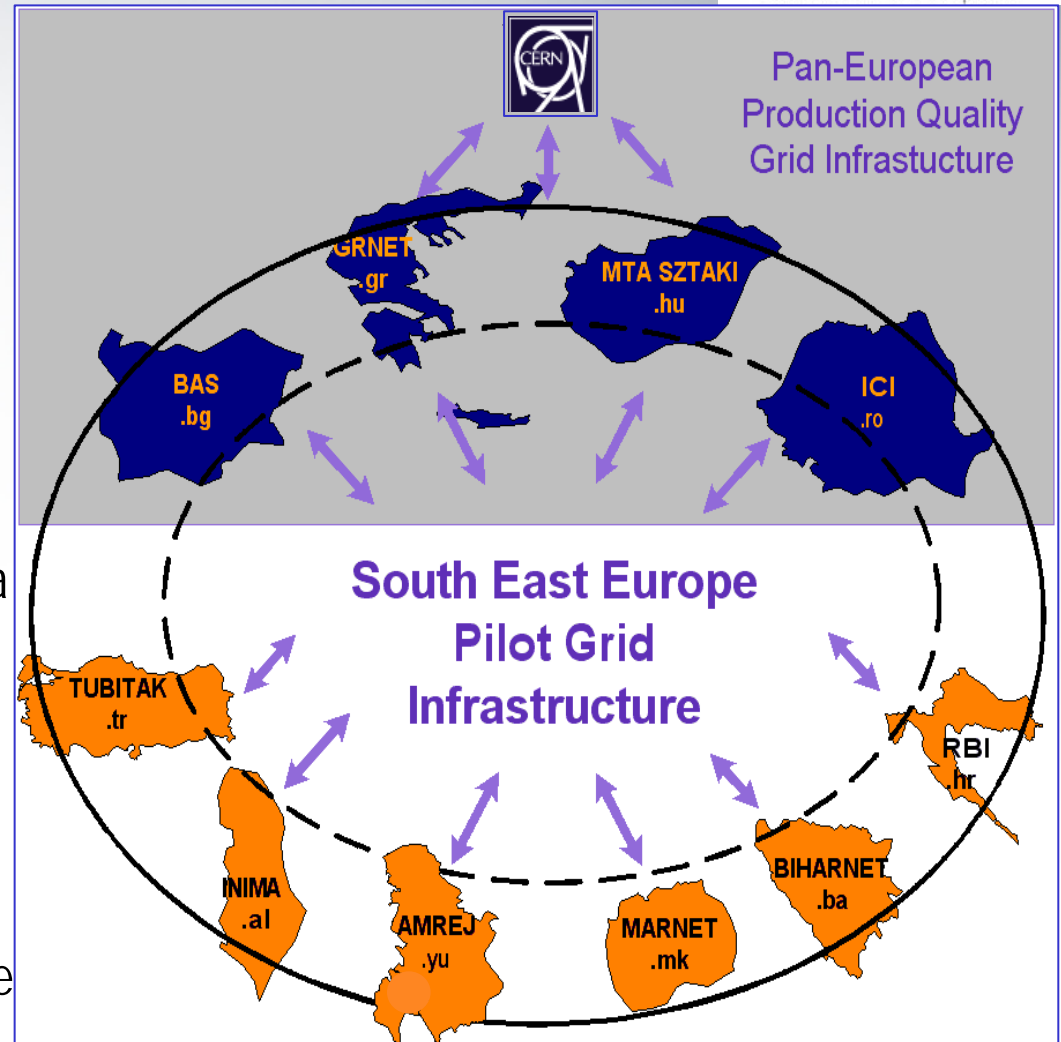
Objectives: Infrastructure

- Establish <u>at least one</u> fully <u>operational</u> and <u>certified</u> grid site in <u>all</u> participating beneficiary SEE countries
  - based on LCG-2 M/W and SL operating system.
  - 5-50 nodes targeted in each site.
  - migration to gLite to be evaluated.
  - <u>all</u> sites to support <u>at minimum</u> UI, SE, CE, WNs.
  - each <u>contractor</u> to set up at least one site
  - <u>third-parties</u> to set up additional sites.
  - expand EGEE SEE ROC to cater for the operations in the SEE(-GRID) countries, with the extension of SEE-GRID-specific monitoring tools


SEE-GRID RESOURCES MAP

GEANT — BUDAPEST

Hungary
- No. of Cluster: 1
- No. of Nodes: 3
- LCG Version: 2.2.0

ZAGREB

Romania
- No. of Cluster: 3
- No. of Nodes: 45
- LCG Version: 2.3.0

BUCHAREST

Croatia
- No. of Cluster: 1
- No. of Nodes: 6
- LCG Version: 2.3.0

BELGRADE

SARAJEVO

Bosnia & Hercegovina
- No. of Cluster:
- No. of Nodes:
- LCG Version:

Serbia & Montenegro
- No. of Cluster: 2
- No. of Nodes: 16
- LCG Version: 2.3.0

SOFIA

Bulgaria
- No. of Cluster: 4
- No. of Nodes: 33
- LCG Version: 2.3.0

SKOPJE

TIRANA

FYR of Macedonia
- No. of Cluster: 3
- No. of Nodes: 9
- LCG Version: 2.2.0

ISTANBUL

Albania
- No. of Cluster: 1
- No. of Nodes: 20
- LCG Version: 2.3.0

Greece
- No. of Cluster: 1
- No. of Nodes: 5
- LCG Version: 2.2.0

Turkey
- No. of Cluster: 3
- No. of Nodes: 135
- LCG Version: 2.3.0

ATHENS

GEANT

Objectives: Applications

- Deploy at least two Grid applications developed by EGEE (one application from the HEP VO, and one application from the Biomed VO) in the regional infrastructure and demonstrate their usage over the regional infrastructure.
  - Taken up by participating countries having expressed a preference for either one or both.
  - Preference must be well-justified

- Deploy at least two Grid applications developed by SEE-GRID partners in the regional infrastructure.
  - Volumetric Image Visualization Environment (VIVE) for medical images and other static or time-dependent scalar and vector 3D fields
  - Search Engine for South-East Europe (SE4SEE) for Grid-aided web-crawling & data indexing
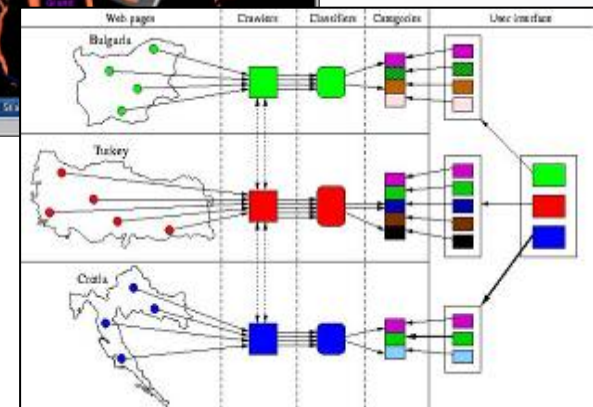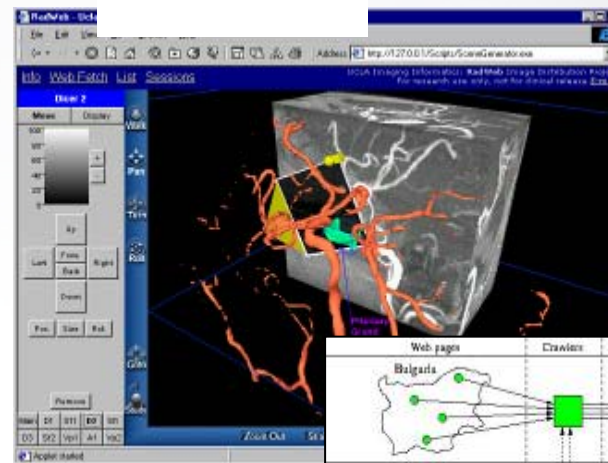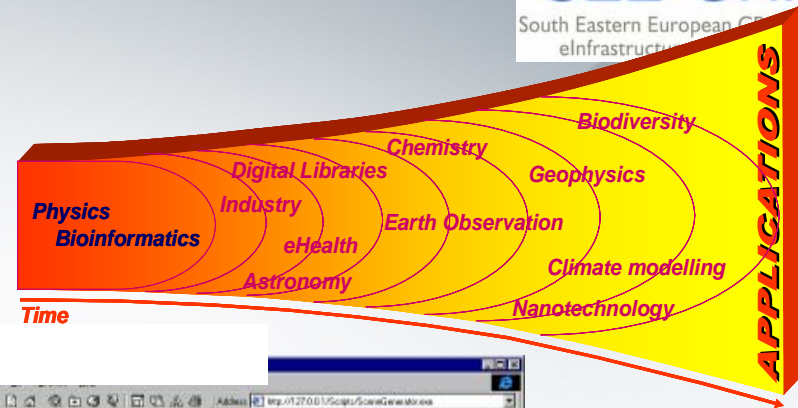
Objectives: Human Network

- Create a Human Network on eInfrastructures and raise awareness in the wide R&E community per country.
  - Committed end-user communities is key
  - Actively disseminate Grids in the SEE R&E community with national and regional events
  - And most important of all:

Training, training, training

Objectives: Sustainability

- Achieve sustainability of the grid infrastructure.
  - Network sustainability is critical and must be demonstrated to be reliable and future-proof
  - Support and promote research and scientific endeavors in the area of eInfrastructures
    - Committed end-user communities is critical
    - Exploit interest for Grid applications and S/W development by the R&E communities
  - Establish National eInfrastructure Policies
    - a <u>formally</u> established NGI per SEE country
    - a national grid Certification Authority per SEE country
  - Implement the SEE-GRID policy workshop recommendations
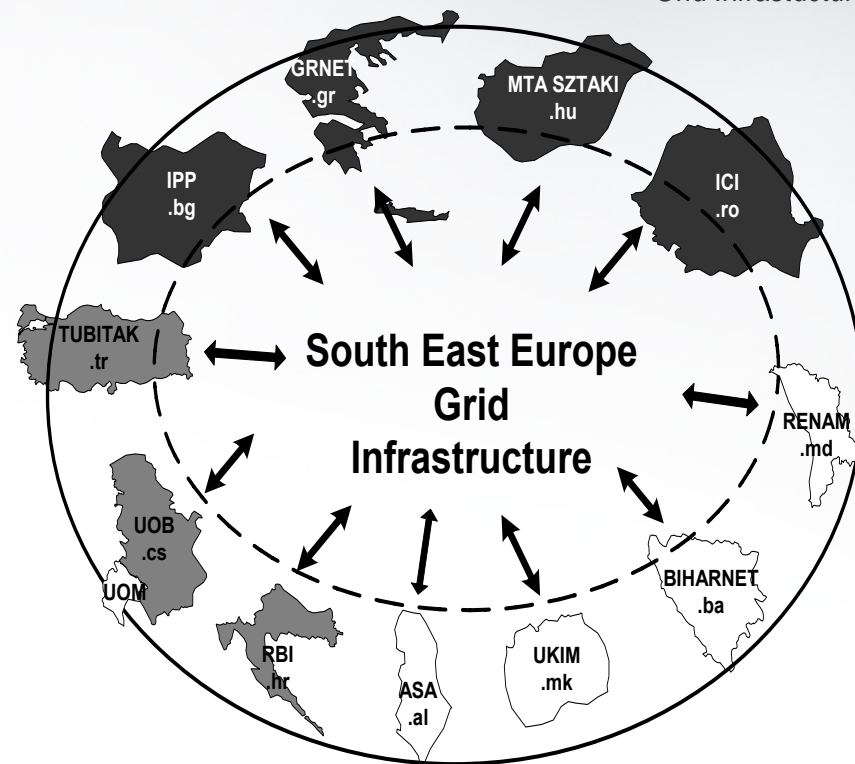
SEE-GRID-2 partnership

- SEE-GRID-2 partnership consists of 13 contractors representing 11 SEE countries
- Partnership includes
  - EU member-states (Greece, Hungary)
  - Acceding Countries (Bulgaria, Romania)
  - Candidate Countries (Croatia, Turkey)
  - Third Countries - Western Balkans (Albania, Bosnia-Herzegovina, Former Yugoslav Republic of Macedonia, Serbia-Montenegro)
  - European Neighborhood Policy countries (Moldova)
- By using participation in EGEE as reference for a partner's maturity, three layers can be identified:
  - Bulgaria, Greece, Hungary, and Romania were members of EGEE and will carry on in EGEE-II
  - Croatia, Serbia, and Turkey advanced within the course of SEE-GRID and have joined EGEE-II
  - Albania, Bosnia-Herzegovina, Former Yugoslav Republic of Macedonia, Moldova, and Montenegro will work towards integration with European eInfrastructures within the course of SEE-GRID-2.

Planned Start date:	01/05/2006
Planned Duration:	24 months
Planned Total Budget:	2,002,691 €

Pan-European Production-level Grid Infrastucture

GRNET .gr — MTA SZTAKI .hu — IPP .bg — ICI .ro — TUBITAK .tr — **South East Europe Grid Infrastructure** — RENAM .md — UOB .cs — UOM — BIHARNET .ba — RBI .hr — ASA .al — UKIM .mk

SEE-GRID-2 Objectives

- Upgrade SEE grid infrastructue
  - Upgrade the capacity of the regional pilot infrastructure
  - Guarantee stability and interoperability of the infrastructure
  - Support the accreditation of national Grid CAs and educate on formal CA/RA procedures.
  - Draw upon deployment experience/results of other grid projects (EGEE/EGEE-II, EUMEDGRID, BalticGrid, EELA, etc)

- Strengthen Human network
  - Liaise with and beyond SEE user communities
  - Training events
  - Dissemination events
  - Regional eInfra projects Policy Workshop
  - Participation in SEE Education and Research

- Ensure sustainable development
  - National commitment and support for incubating NGIs
  - Engage regional and national user communities

Aim: *To develop and integrate the research and education computing and communication infrastructure in the Baltic States into the emerging European Grid infrastructure.*

Applications: *HEP, Regional: Material Science, Bioinformatics*

Coordinator: *KTH, Sweden*

Participating countries: *Estonia (2 partners), Latvia (2 partners), Lithuania (2 partners), Poland (2 partners), Sweden, Switzerland*

**Start date:** *November 2005 for 30 months*

- **Objectives**

  – Sustainable integration of the research and educational computing and communication infrastructure in the Baltic States into the European Grid infrastructure.

  – Enable the formation of effective research collaborations in the Baltic States, within Europe and beyond.

  – Enabling an efficient sharing of unique instruments and data, as for instance environmental data related to the Baltic Sea.

- **Approach**
  – Maximize use of human resources for Grid development and deployment by implementing the Baltic Grid as an extension of EGEE
    - Grid Operations coordinated with the EGEE North European Regional Operating Centre

  – Assure manageable co-existence with other European Grids
    - DEISA, CrossGrid, NorduGrid…

  – Engage the Baltic States in Grid related policy and standards activities

- **Estonia**
  - Tallinn, Tartu
- **Lithuania**
  - Vilnius
- **Latvia**
  - Riga
- **Poland**
  - Kraków, Poznan
- **Switzerland**
  - Geneva
- **Sweden**
  - Stockholm
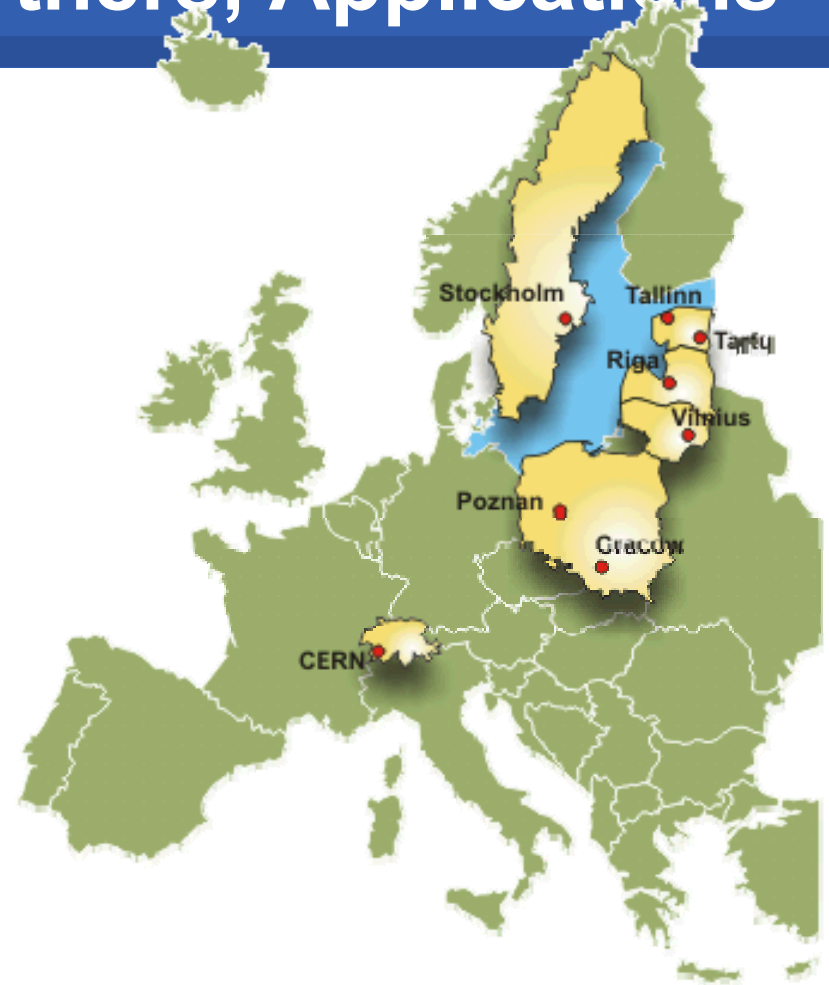
- **High-Energy Physics**
  - statistical data analysis, Monte Carlo simulation
- **Material Sciences**
  - atomic and molecular structures, solid state surfaces, photon and electron interactions
- **Bioinformatics**
  - sequence pattern discovery, modeling of biosensors
- **Special interest groups**
  - Baltic Sea eco-system, Baltic Sea Marine Research, E-Health, etc.

eGee

Enabling Grids for E-sciencE

Aim: *To provide specific support actions to assist in the participation of the states of the Mediterranean region in the pan-European and worldwide Grid initiatives.*

Applications: *HEP, Biomedicine, Earth Science, Regional: Water Control*

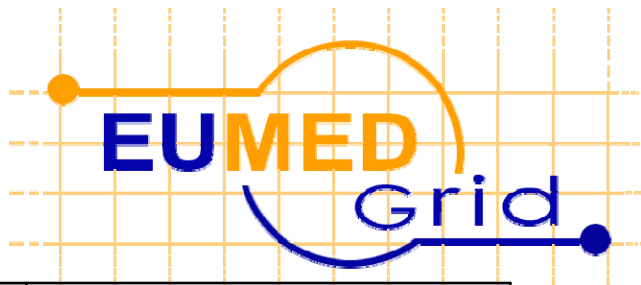**Coordinator:** *INFN, Italy*

Participating countries: *Algeria, Cyprus, Egypt, Greece, Italy, Malta, Morocco, Spain, Switzerland, Syria, Tunisia, Turkey, UK*

**Start date:** *January 2006 for 24 months*

EUMEDGRID

Main Objectives

▸ **EUMEDGRID aims to provide specific support actions to assist the participation of the states of the Mediterranean region in the pan-European and worldwide Grid initiatives, thus expanding and supporting the European Research Area (ERA) in the region.**

▸ **The core of the EUMEDGRID approach is to establish a human network in the eScience area, enlarge and train this community, and establish a pilot Grid infrastructure supporting proof of concept regional applications.**

▸ **The reference GRID Infrastructure in Europe will be EGEE.**

▸ **EUMEDGRID will build upon and exploit the infrastructure provided by the Gigabit Pan-European Research & Education Network (GEANT) and the Mediterranean Research and Education Networking (EUMEDCONNECT) initiative in the region.**
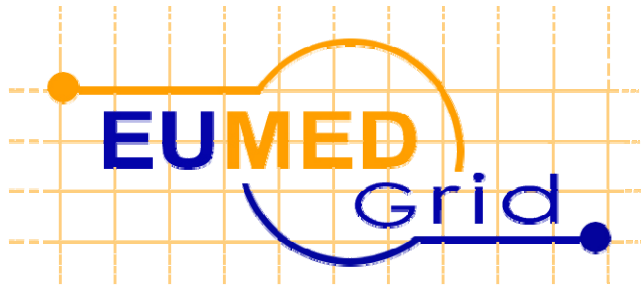
EUMEDGRID Infrastructure

	Participant name
1	INFN (Italy) Coordinator
2	CERN (Switzerland)
3	CYNET (Cyprus)
4	DANTE (UK)
5	GARR (Italy)
6	GRNET (Greece)
7	RED.ES (Spain)
8	University of Malta (Malta)
9	CERIST (Algeria)
10	CNRST (Morocco)
11	EUN (Egypt)
12	HIAST (Syria)
13	RNRST (Tunisia)
14	TUBITAK - ULAKBIM (Turkey)

20
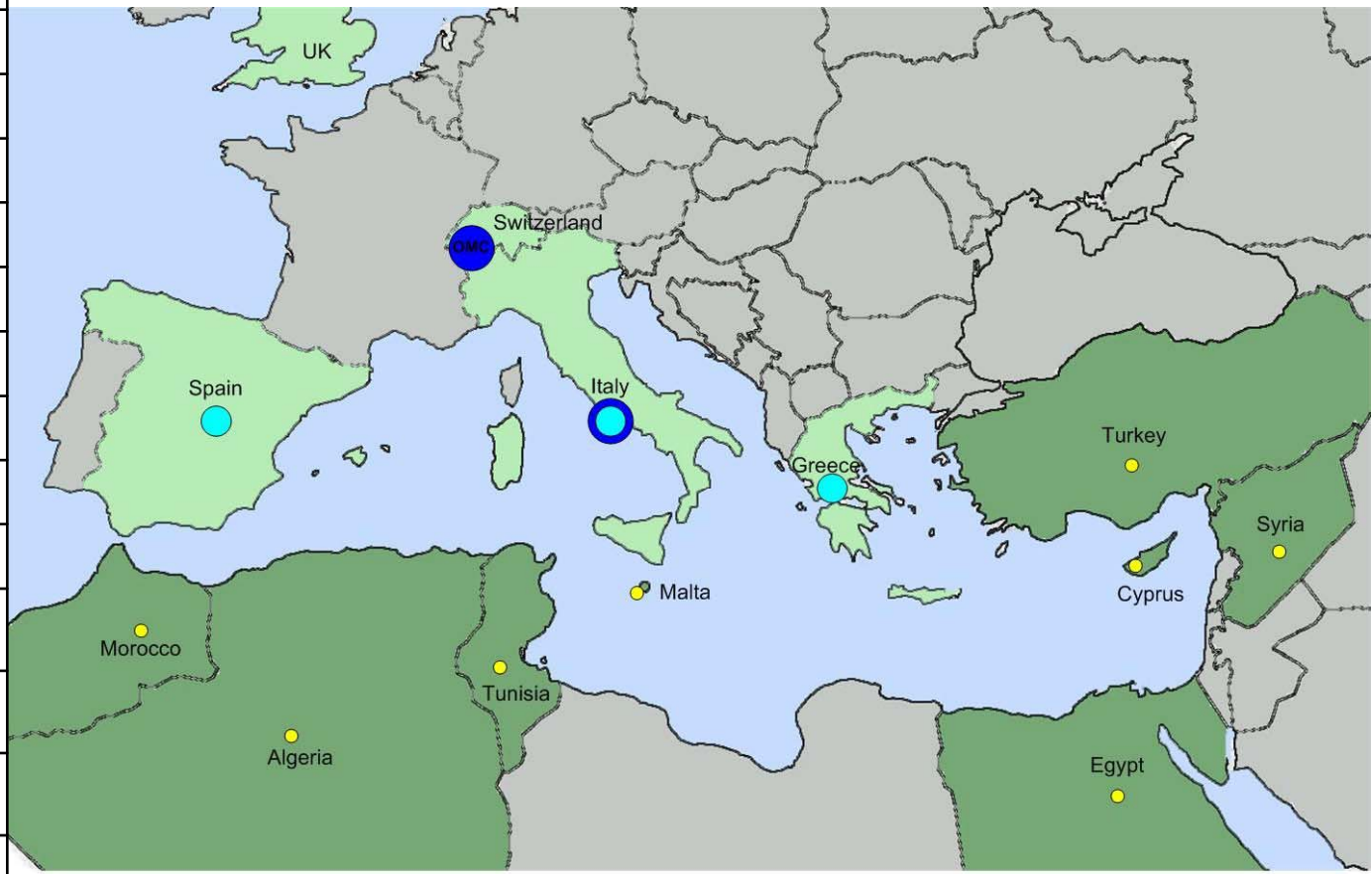
Grid Applications

▸ **Biomed applications (already EGEE pilot applications)**

▸ **HEP applications (already EGEE pilot applications)**

▸ **New candidates:**

- Earth science (hydrology)
- e-Learning (multi-media content on-demand)

▸ **Others to be selected**

Hydrology

▸ **Aim: estimate sustainable extraction scheme - improve management**

▸ **CODESA-3D: Density-dependent 3D coupled groundwater flow and transport simulations**

▸ **Data requirement**

- Geology
- Topography
- Meteorology
- Water extraction by the farmer
- Aquifer properties
- Soil maps
- Land use
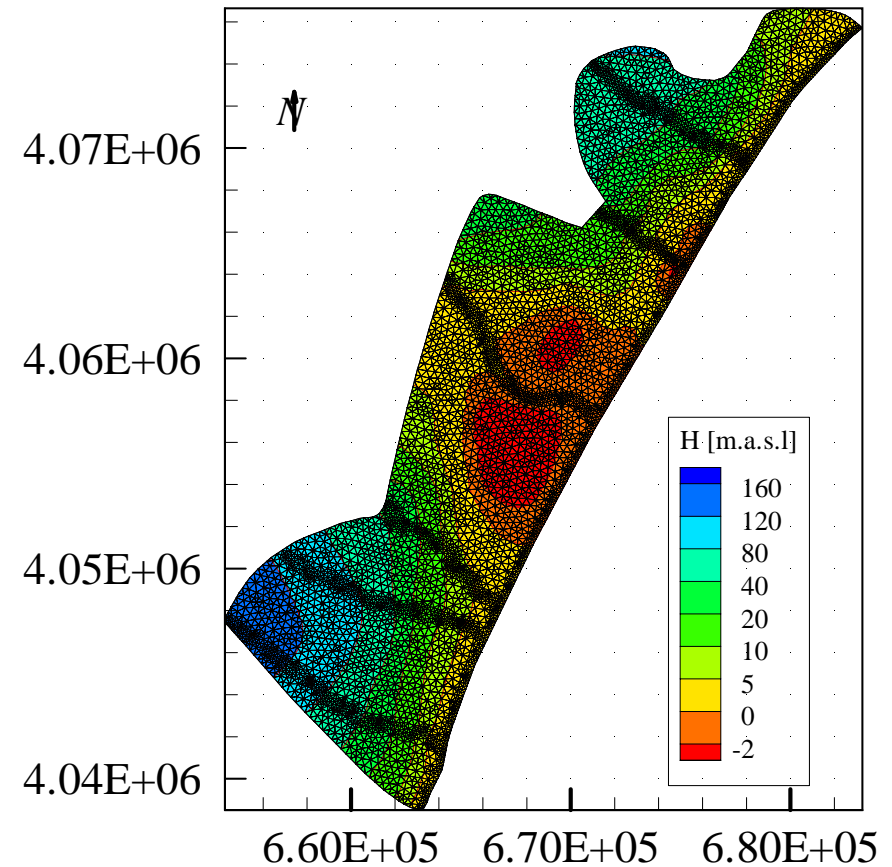
One simulated map of water levels

Enabling Grids for E-sciencE

Aim: *To build a bridge between consolidated e-Infrastructure initiatives in Europe and emerging ones in Latin America.*

Applications: *HEP, Regional: Bioinformatics, Education, Climate*

Coordinator: *CIEMAT, Spain*

Participating countries: *Argentina, Brazil (5 partners), Chile (3 partners), Cuba, Italy, Mexico, Portugal, Peru, Spain (5 partners), Switzerland, Venezuela plus CLARA*

Start date: *January 2006 for 24 months*

http://www.eela-grid.org/

Enabling Grids for E-sciencE

- **Goals:**
  - Build a bridge between consolidated e-Infrastructure initiatives in Europe and emerging ones in Latin America
  - Reinforce collaboration between Latin America and Europe
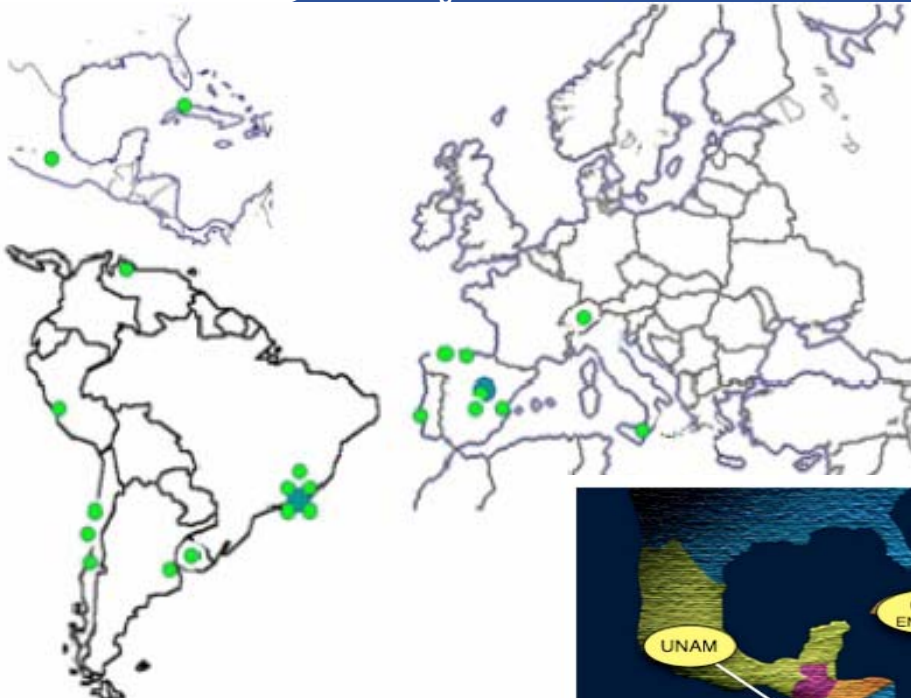
- **Objectives:**
  - Establish a human collaboration network between Europe and Latin America:
    - Setting up the structure of the collaboration network
    - Establishing adequate support mechanisms
    - Adopting policies regarding the shared use of e-Infrastructure
    - Evaluating new areas of collaboration and relevant partners, both in Europe and Latin America

Enabling Grids for E-sciencE

- Build a pilot e-Infrastructure in Latin America:
  - Implementing basic mechanisms for an interoperable e-Infrastructure, adopting a security policy, establishing Certification Authorities and defining basic tools middleware
  - Setting up a Pilot Testbed, establishing Virtual Organizations and supporting application developers and users
  - Supporting advanced network services

- Promote a sustainable framework for e-Science:
  - Identifying research communities and applications
  - Supporting dissemination efforts
  - Coordinating participation in possible new projects
  - Defining a map for a future consolidated e-Infrastructure in LA

- **EUROPE**
  - Italy: INFN
  - Portugal: LIP
  - Spain: CIEMAT (coordinator),
         CSIC, RED.ES, UC, UPV

- **INTERNATIONAL ORGANIZATIONS**
  - CERN
  - CLARA

- **LATIN AMERICA**
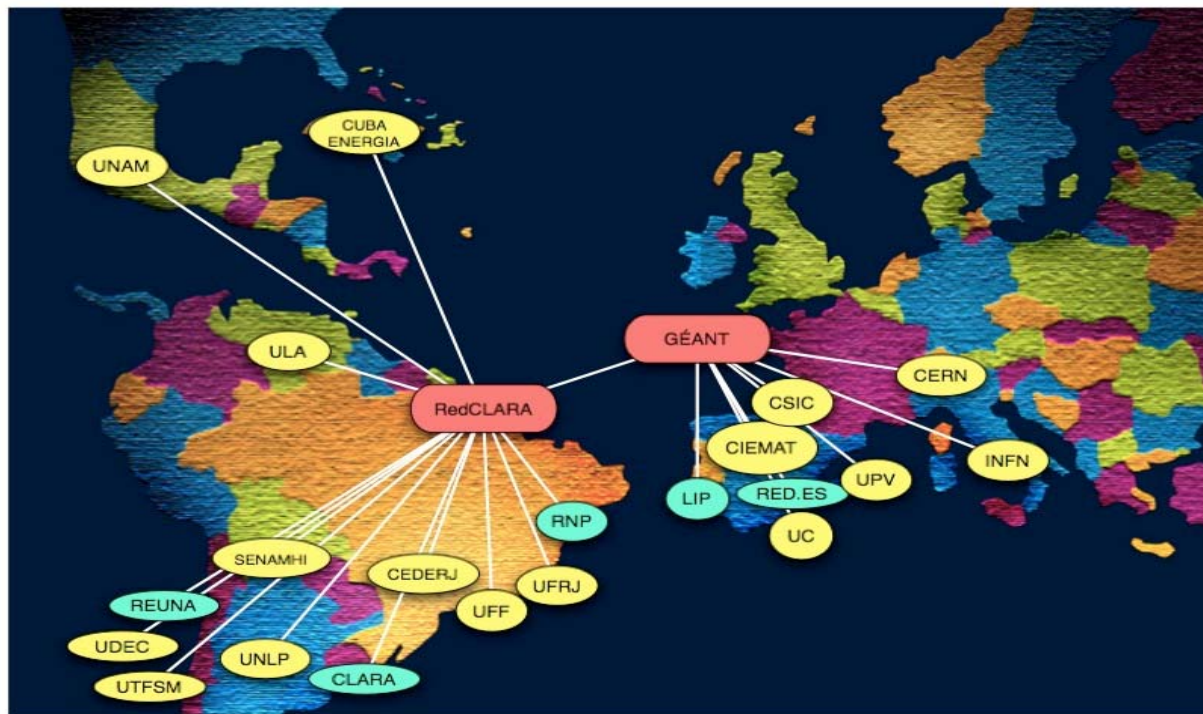 - Argentine: UNLP
  - Brazil: CECIERJ/CEDERJ,
         RNP, UFF, UFRJ
  - Chile: REUNA, UDEC, UTFSM
  - Cuba: CUBAENERGIA
  - Mexico: UNAM
  - Peru: SENAMHI
  - Venezuela: ULA

Enabling Grids for E-sciencE

- **Biomed**
  - GATE
  - WISDOM
  - CECALC web portal (ULA - http://www.cecalc.ula.ve/)
  - Protein Dynamics (UFRJ)
- **HEP**
  - LHC Exps.
- **Additional (EELA specific)**
  - Climate in the Grid Environment
  - Education in the Grid Environment
  - ... And others to come

Aim: *To interconnect the Grid infrastructure in Europe and China for the benefit of e-Science applications.*

Applications: *HEP, Regional: Cosmic Ray Experiment, Biology*

Coordinator: *INFN, Italy*

Participating countries: *China (4 partners) Greece, Italy (3 partners), Poland, Switzerland*

Start date: *January 2006 for 24 months*

- CNGrid - A pilot GRID infrastructure is already in place in China supported by the China National High-tech Research and Development Program.
  - CNGrid aims to provide a platform for nation-wide resource sharing and cooperative work in China, which will support applications in different disciplines and domains.

Introduction

- ▶ **Status, Timescale & Budget**
  - The official start of the project is the 1 January 2006.
  - 24 Months duration
  - EU Contribution of 1,300,000 €.
  - A total of 495 Person Months (325 Funded).
- ▶ **Targets**
  - To foster the creation of a intercontinental eScience community
    - – Training people
    - – Supporting existing and new applications
  - To support interoperable infrastructure for grid operations between Europe and China

Main Objectives

O1 Contribute to the creation of an human network promoting international collaboration.

O2 Disseminate European experience in China and confront with Chinese experience.

O3 Support interoperability of EGEE with the corresponding infrastructure in China, fronting the multi-protocol (IPv4/IPv6) network infrastructure.

O4 Use existing network infrastructure and foster its development

O5 Foster interoperability of solutions across different disciplines to achieve broader scale uptake of Grid technology.

O6 Harmonize European, and Chinese eScience user and infrastructure requirements in terms of resources needed, Grid services, and application software. Provide recommendations for adapting the present best practices and tools (Virtual Organizations, Certification Authority, Policies, etc.).

euchinagrid
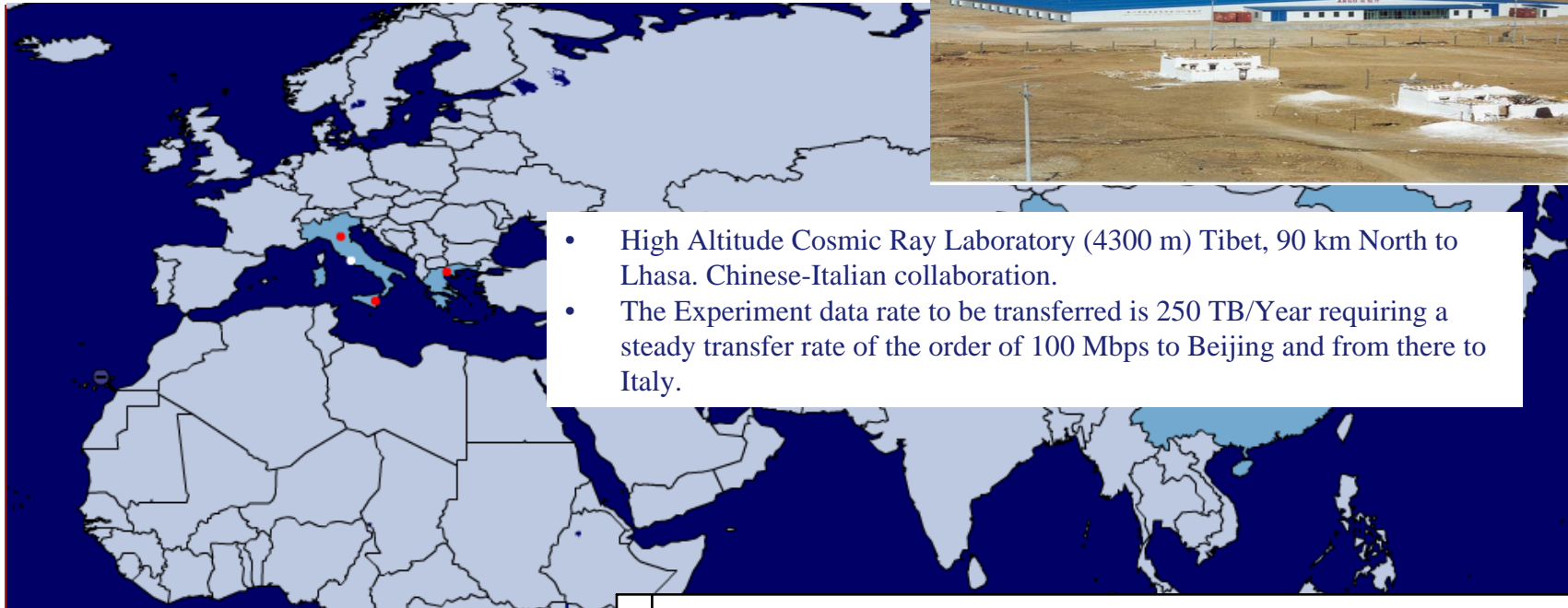
- High Altitude Cosmic Ray Laboratory (4300 m) Tibet, 90 km North to Lhasa. Chinese-Italian collaboration.
- The Experiment data rate to be transferred is 250 TB/Year requiring a steady transfer rate of the order of 100 Mbps to Beijing and from there to Italy.

Applications:

- EGEE applications (LHC, Bio, etc)
- ARGO-YBJ and Gamma ray bursts
- Never Born Proteins

1	Istituto Nazionale di Fisica Nucleare (IT) (coordinator)
2	European Organisation for Nuclear Research CERN (CH)
3	Dipartimento di Biologia - Università di Roma Tre (IT)
4	Consortium GARR (IT)
5	Greek Research & Technology Network (GR)
6	Jagiellonian University – Medical College, Cracow (PL)
7	School of Computer Science and Engineering – Beihang University Beijing (CN)
8	Computer Network Information Center, Chinese Academy of Sciences – Beijing (CN)
9	Institute of High Energy Physics, Beijing (CN)
10	Peking University – Beijing (CN)

Giuseppe Andronico -

Aim: *To provide an integrated healthcare platform for European pediatrics.*

Coordinator: *Siemens, Germany*

Start date: *January 2006 for 48 months*

- **EGEE**
  - **provide the foundation on which the integrated platform will be based.**

- **Health-e-Child**
  - **provide feedback to EGEE on new medical requirements, security issues, and resource sharing and allocation, and will closely collaborate with the Biomed group of NA4.**

- Siemens AG, Germany (lead partner)
- Lynkeus SRL, Rome Italy
- Gaslini Institute, Genoa, Italy
- UCL Great Ormond St. Children's Hospital, London, UK
- Assistance Publique Hopiteaux de Paris, France
- CERN, Geneva
- Maat-G Knowledge, Toledo, Spain
- University of the West of England, Bristol, UK
- University of Athens, Greece
- University of Genoa, Italy
- INRIA, France
- European Genetics Foundation, Bologna, Italy
- Aktaselts Asper Biotech, Tartu, Estonia
- Gerolamo Gaslini Foundation, Genoa, Italy

eGee

- **Mission**
  - There is a compelling demand for the integration and exploitation of heterogeneous biomedical information for improved clinical practice, medical research, and personalised healthcare for the citizens of the EU.

- **Aims**
  - Develop an integrated healthcare platform for European paediatrics,
    - providing seamless integration of traditional and emerging sources of biomedical information.
    - The long-term goal of the project is to provide uninhibited access to universal biomedical knowledge repositories for personalised and preventive healthcare, large-scale information-based biomedical research and training, and informed policy making.

- **Focus**
  - Individualised disease prevention, screening, early diagnosis, therapy and follow-up of paediatric heart diseases, inflammatory diseases, and brain tumours.
  - Build a Grid-enabled European network of leading clinical centres to share and annotate biomedical data, validate systems clinically, and diffuse clinical excellence across Europe by setting up new technologies & clinical workflows

Enabling Grids for E-sciencE

- **Objectives**
  - To gain a comprehensive view of a child's health by vertically integrating biomedical data, information, and knowledge, that spans the entire spectrum from genetic to clinical to epidemiological;
  - To develop a biomedical information platform, supported by sophisticated and robust search, optimisation, and matching techniques for heterogeneous information, empowered by the Grid;
  - To build enabling tools and services on top of the Health-e-Child platform, that will lead to innovative and better healthcare solutions in Europe:

  - Integrated disease models exploiting all available information levels;
  - Database-guided biomedical decision support systems provisioning novel clinical practices and personalised healthcare for children;
  - Large-scale, cross-modality, and longitudinal information fusion and data mining for biomedical knowledge discovery.

Summary

Enabling Grids for E-sciencE

- **Extending the EGEE infrastructure worldwide**
  - CERN IT provides the infrastructure coordination point

INFSO-RI-508833 35