

The European Grid Infrastructure EGEE Project

Fabrizio Gagliardi¹

CERN IT Division, 1211 Geneva 23, Switzerland

Abstract. This paper describes the motivation for the EGEE (Enabling Grids for e-Science in Europe) proposal recently made to the EU and the background activity behind this initiative

1. Progress of DataGrid

The EU DataGrid project, the major grid development European effort, will complete its activity at the beginning of 2004. The results of its intense three years of activity show in a testbed that comprises 12 sites in 6 countries and which provides significant computing and storage resources to a community of approximately 500 users from thirteen different virtual organizations. The latest release of the DataGrid software has been successfully validated on a large set of applications, ranging from High Energy Physics to Bio-Informatics and Earth Observation. This software is now the basis of the current CERN Large Hadron Collider Grid Project first production infrastructure, the facility that is being setup for the analysis of data that will be produced by the new CERN accelerator (LHC). Although a considerable amount of work remains to be done, EDG, with its achievements, has proved the validity of the Grid concept and paved the way for the next generation Grid production infrastructure for a much wider multi-science international community.

2. The EGEE Vision

EGEE (Enabling Grids for E-Science in Europe) aims to integrate current national, regional and thematic Grid efforts, in order to create a seamless European Grid infrastructure for the support of the European Research Area. This infrastructure will be built on the EU Research Network GEANT and exploit Grid expertise that has been generated by projects such as the EU DataGrid project, other EU supported Grid projects and the national Grid initiatives such as UK e-Science, INFN Grid, Nordugrid and the US Trillium (cluster of projects).

The EGEE vision is that this Grid infrastructure will provide European researchers in academia and industry with a common market of computing resources, enabling round-the-clock access to major computing resources, independent of geographic location. This infrastructure will support distributed

¹EU DataGrid Project Leader, EGEE designated Project Director,
on behalf of the EU DataGrid project and the EGEE Collaboration

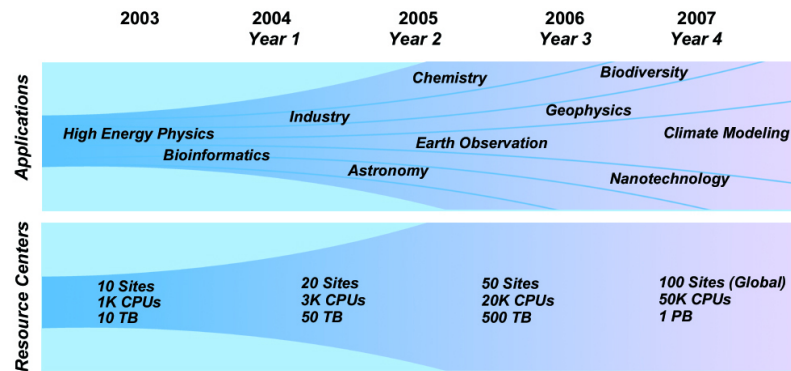


Figure 1. Schema of the evolution of the European Grid infrastructure from two pilot applications in high energy physics and biomedical Grids, to an infrastructure serving multiple scientific and technological communities, with enormous computer resources. The applications and resource figures are purely illustrative. The EGEE project covers Year 1 and 2 of a planned four-year programme.

research communities, including relevant Networks of Excellence, which share common Grid computing needs and are prepared to integrate their own distributed computing infrastructures and agree common access policies. The resulting infrastructure will surpass the capabilities of local clusters and individual supercomputing centres in many respects, providing a unique tool for collaborative compute-intensive science (“e-Science”) in the European Research Area. Finally, the infrastructure will provide interoperability with other Grids around the globe, including the US NSF Cyberinfrastructure, contributing to efforts to establish a worldwide Grid infrastructure. The scope of the project is illustrated in Figure 1.

EGEE has been proposed by experts in Grid technologies representing the leading Grid activities in Europe. The process of developing this project has led to a structuring of the European Grid community into ten partner regions or “federations” (Figure 2). A significant structuring effect due to EGEE is already apparent, as several of these partners have begun integrating regional Grid efforts in order to provide coordinated resources to the EGEE project. In addition, US representatives are participating as EU unfunded partners in the project, and are considering establishing a US EGEE federation. Participation of Japan and the Asia-Pacific region is considered desirable and will be pursued.

EGEE is a two-year project conceived as part of a four-year programme. Major implementation milestones after two years will provide the basis for assessing subsequent objectives and funding needs. Given the service-oriented nature of this project, two pilot application areas have been selected to guide the implementation and certify the performance and functionality of the evolving European Grid infrastructure. One is the Large Hadron Collider Computing Grid, which relies on a Grid infrastructure in order to store and analyse petabytes of real and simulated data from high-energy physics experiments at CERN. The

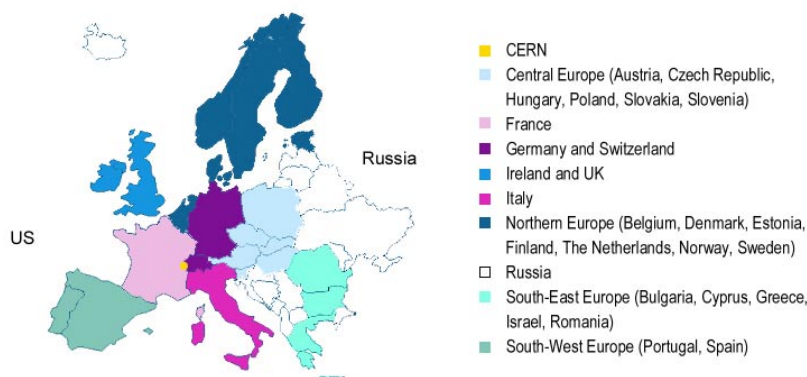


Figure 2. The EGEE federations

other is Biomedical Grids, where several communities are facing equally daunting challenges to cope with the flood of bioinformatics and healthcare data.

Given the rapidly growing scientific needs for a Grid infrastructure, it is deemed essential for the EGEE project to “hit the ground running”, by deploying basic services, and initiating joint research and networking activities before the formal start of the project. The LCG project will provide basic resources and infrastructure already during 2003, and Biomedical Grid applications will be planned at this stage. The available resources and user groups will then rapidly expand during the course of the project. To ensure that the project ramps up rapidly, project partners have agreed to begin providing their unfunded contribution prior to the official start of the project.

3. The EGEE Mission

In order to achieve the vision outlined above, EGEE has a three-fold mission:

1. To deliver production level Grid services, the essential elements of which are manageability, robustness, resilience to failure, and a consistent security model, as well as the scalability needed to rapidly absorb new resources as these become available, while ensuring the long-term viability of the infrastructure.
2. To carry out a professional Grid middleware re-engineering activity in support of the production services. This will support and continuously upgrade a suite of software tools capable of providing production level Grid services to a base of users which is anticipated to rapidly grow and diversify.
3. To ensure an outreach and training effort, which can proactively market Grid services to new research communities in academia and industry, capture new e-Science requirements for the middleware and service activities, and provide the necessary education to enable new users to benefit from the Grid infrastructure.

4. The Stakeholder Perspective

The key types of EGEE stakeholders are users, resource providers, and industrial partners.

4.1. EGEE Users

Once the EGEE infrastructure is fully operational, users will perceive it as one unified large scale computational resource. From the user perspective, the complexity of the service organisation and the underlying computational fabric will remain invisible. The benefits of EGEE from the user perspective include: simplified access, on demand computing, pervasive access, large scale resources, sharing of software and data and improved support.

A potential user community will typically come into contact with EGEE through one of the many outreach events supported by the Dissemination and Outreach activity, and will be able to express their specific user requirements via the Applications Identification and Support Activity. After negotiating access terms, which will depend, amongst other things, on the resources the community can contribute to the Grid infrastructure, users in the community will receive training from the User Training and Induction activity. From the user perspective, the success of the EGEE infrastructure will be measured in the scientific output that is generated by the user communities it is supporting.

4.2. Resource Providers

EGEE resources will include national GRID initiatives, computer centres supporting one specific application area, or general computer centres supporting all fields of science in a region. The motivation for providing resources to the EGEE infrastructure will reflect the funding situation for each resource provider. EGEE will develop policies that are tailored to the needs of different kinds of partners. Among the most important benefits for resource providers are large scale operations, specialist competence, user contacts and collaborations with resource partners. These benefits motivate the many partners that support the EGEE proposal already, representing aggregate resources of over 17000 cluster nodes.

EGEE builds on the integration of existing infrastructures in the participating countries, in the form of national GRID initiatives, computer centres supporting one specific application area, or general computer centres supporting all fields of science in a region. The motivation for providing resources to the EGEE infrastructure depends on the mission and funding situation for each of the resource partners. A new resource provider will typically approach EGEE through contact with the Regional Operations Centres. Specific policy and contractual issues for a given resource provider will be dealt with by dedicated staff in the Operations Management Centre, based on general guidelines defined and regularly reviewed by the Project Executive Board, with advice from the Project Management Board, and reviewed regularly.

4.3. Industrial Partners

The driving force for EGEE is scientific applications, and the current partners represent publicly funded research institutions and computer resource providers

from across Europe. Nevertheless, it is envisaged that industry will benefit from EGEE in several ways, as it can play the different roles of user, partner and provider.

The EGEE vision also has inspiring long-term implications for the IT industry. By pioneering the sort of comprehensive production Grid services which are envisioned by experts — but which at present are beyond the scope of national Grid initiatives — EGEE will have to develop solutions to issues such as scalability and security that go substantially beyond current Grid R&D projects. This process will lead to the spin off of innovative IT technologies, which will have benefits for industry, commerce and society going well beyond scientific computing. Major initiatives launched by several IT industry leaders in the area of Grids and Utility computing emphasize the economic potential of this emerging field.

Industry will typically come in contact with EGEE via the Industry Forum organised by the Application Identification and Support activity, as well as more general dissemination events run by the Dissemination and Outreach activity. Interested companies will be able to consult about potential participation in the project with the Project Director and with regional representatives on the EGEE Project Management Board. As the scope of Grid services expands during the second two years of the programme, it is envisaged that established core services will be taken over by industrial providers with proven service capacity. This service would be provided on commercial terms, and selected by a competitive tender.

5. EGEE Activities

Reflecting the three-fold mission outlined in section 3, EGEE is structured in three main areas of activity: services, middleware re-engineering and networking.

5.1. Service Activities

The Service Activities will create, operate, support and manage a production quality European Grid infrastructure which will make resources at many resource centres across Europe accessible to user communities and virtual organisations in a consistent way according to agreed access management policies and service level agreements, while maintaining an overall secure environment. These activities will build on current national and regional initiatives such as the UK e-Science Grid, the Italian Grid, and NorduGrid, as well as infrastructures being established by specific user communities, such as LCG. The structure of the Grid services will comprise: EGEE Operations Management at CERN; EGEE Core Infrastructure Centres in the UK, France, Italy and at CERN, responsible for managing the overall Grid infrastructure; Regional Operations Centres, responsible for coordinating regional resources, regional deployment and support of services. The basic services that will be offered are: middleware deployment and installation; a software and documentation repository; Grid monitoring and problem tracking; Bug reporting and knowledge database; Virtual Organization (VO) Services; Grid Management Services. Continuous, stable Grid operation represents the most ambitious objective of EGEE, and requires the largest effort.

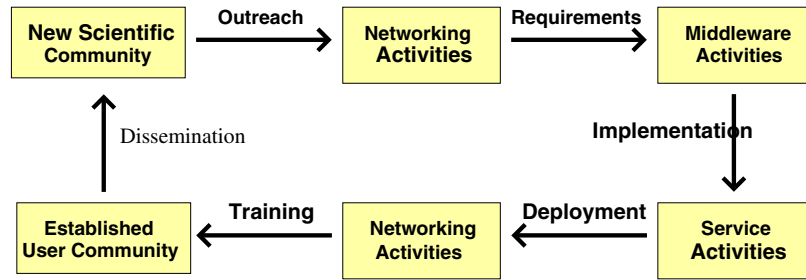


Figure 3. The “Virtuous Cycle” for EGEE development

5.2. Middleware Re-engineering Activities

The current state-of-the-art in Grid Computing is dominated by research Grid projects that aim to deliver test Grid infrastructures providing proofs of concept and opening opportunities for new ideas, developments and further research. Only recently there has been an effort to agree on a unified Open Grid Services Architecture (OGSA) and an initial set of specifications constituting the Open Grid Service Infrastructure that set some of the standards in defining and accessing Grid services. Building a European Grid infrastructure based on robust components is thus becoming feasible. However, this will still take a considerable integration effort in terms of making the existing components adhere to the new standards, adapting them to evolution in these standards, and deploying them in a production Grid environment. The middleware activities in EGEE focus primarily on re-engineering existing middleware functionality, leveraging the considerable experience of the partners with the current generation of middleware. Based on experience, geographic co-location of development staff is essential, and therefore these activities are based on tightly-knit teams concentrated in a few major centres with proven track records and expertise.

5.3. Networking Activities

The networking activities in EGEE aim to facilitate the induction of new users, new scientific communities and new virtual organisations into EGEE community. EGEE will develop and disseminate appropriate information to these groups proactively, and take into account their emerging Grid infrastructure needs. The goal is to ensure that all users of the EGEE infrastructure are well supported and to provide input to the requirements and planning activities of the project. Specific activities included in the EGEE proposal are: Dissemination and Outreach; User Training and Induction; Application Identification and Support; Policy and International Cooperation. The Application Identification and Support Activity has three components, two Pilot Application Interfaces — for high energy physics and biomedical Grids — and one more generic component dealing with the longer term recruitment of other communities.

It is essential to the success of EGEE that the three areas of activity should form a tightly integrated “Virtuous Cycle”, illustrated in Figure 3. In this way, the project as a whole can ensure rapid yet well-managed growth of the computing resources available to the Grid infrastructure as well as the number of scientific communities that use it. As a rule, new communities will contribute

new resources to the Grid infrastructure. This feedback loop is supplemented by an underlying cyclical review process covering overall strategy, middleware architecture, quality assurance and security status, and ensuring a careful filtering of requirements, a coordinated prioritization of efforts and maintenance of production-quality standards.

6. Conclusions

The EGEE project has successfully concluded the negotiation of the FP6 Research Infrastructure contract with the EU at the end of October and expects to start operations early Spring 2004. More than 160 researchers are gathering to participate in the various activities of the project and several new job positions have been opened by most of the seventy EGEE partners. If the first phase of the project will deliver by 2006, according to the expectations, a production quality Grid infrastructure for the European Research Area and the international scientific community, it is planned to propose a second phase to extend both the geographical coverage of this infrastructure and the number of supported end-user international scientific communities.

Acknowledgments. EGEE is proposed as a project funded by the European Union under contract IST-2003-508833