The Future European Grid Infrastructure – Roadmap and Challenges

Dieter Kranzlmüller
Ludwig-Maximilians-Universität (LMU) München &
Leibniz Supercomputing Centre (LRZ), Germany
kranzlmueller@ifi.lmu.de

In many research domains, support from computing technology is seen as a major tool for obtaining scientific breakthroughs. This involves many different kinds of technologies from networking, supercomputing and storage technologies often provided within so-called e-Infrastructures.

At present, Europe is preparing the next step in consolidation of its e-Infrastructures. This concerns particularly the initiation of the European grid infrastructure as a permanent and sustainable grid for all sciences. It is expected that the provisioning of an always-available and always-on grid would be beneficial for many applications. These benefits have been demonstrated in many grid projects in Europe and abroad, and there is potential for grids emerging in many other application domains as well. An example is the EGEE Grid Infrastructure (Enabling Grids for e-SciencE), where (as of January 2009) 267 sites in 52 countries are interconnected with the gLite middleware. This infrastructure alone provides 114.000 CPUs and 20 PetaBytes of storage capacity, which are utilized by more than 16.000 users within more than 200 Virtual Organizations (VOs), who submit more than 140.000 Jobs per day. The range of applications served by EGEE is impressive and contains different areas, from high energy physics, earth sciences, and astrophysics to civil protection, finance, and multimedia.

Yet, the problems with today's grid infrastructures stem from a series of drawbacks, which users face when "moving" to the grid. The main issue is the long-term perspective, which cannot be ensured by today's grids. With typical funding cycles of 2 to 4 years, there is no guarantee that the same grid will be available and operating in 5 or 10 years. Yet, adapting applications to the grid, a process often called "gridifying", still requires a substantial and non-neglectable amount of effort, time, and thus money from the developers. This investment by the users as well as the previous funding by the various funding agencies must be protected. At the same time, some application domains (e.g. high energy physics) already depend on the grid today. Somebody utilizing the grid on a daily basis in a production environment needs to ensure that the infrastructure is available in the future.

As a consequence, the project-based funding has been identified as the major obstacle by the e-Infrastructure Reflection Group (e-IRG – www.e-irg.eu), an advisory committee for the European Commission (EC) and the national funding agencies, which provides recommendations on actions relevant for the evolution of e-Infrastructures in Europe. An example is the following statement from December 2005: ""The e-IRG recognizes that the current project-based financing model of grids (e.g., EGEE, DEISA) presents continuity and interoperability problems, and that new financing and governance models need to be explored – taking into account the role of national grid initiatives as recommended in the Luxembourg e-IRG meeting." Besides the identification of project-based funding as a major obstacle, this statement also provides an idea on how to overcome the situation. The idea, originally copied from the efforts in networking, applies the main characteristics of the National Research and Education Networks (NRENs) to the domain of grid computing by asking for the establishment of National Grid Initiatives (NGIs).

Following this advice from the e-IRG, interested communities assembled together in a series of meetings, finally leading to an endeavor called the "European Grid Initiative (EGI)". The vision of the EGI has been drafted in the EGI Vision Paper (www.eu-egi.eu/vision.pdf), which describes the expectations and characteristics of the future European Grid. The resulting EGI infrastructure will be composed of the grids provided by the NGIs and a coordinating body, the so-called EGI organization (EGI.eu).

Proceedings of the *ITI 2009 31st Int. Conf. on Information Technology Interfaces,* June 22-25, 2009, Cavtat, Croatia

After drafting the EGI vision, it has been circulated within the community with the goal to obtain support and to identify the NGIs and their representatives. In an impressive effort in 2007, up to 42 institutions provided strong and written statements of support for the EGI vision, and identified the representatives of their NGIs. The latter are summarized in the EGI Policy Board (EGI PB), which is the governance body for EGI and takes the corresponding decisions on the management level. In addition, the 42 institutions expressed their support for a small design project called EGI Design Study (EGI_DS) with the EU's 7th Framework Programme, which is exploited to drive the developments of EGI by drafting corresponding documents and obtaining consensus within the NGIs.

At this point in time, the EGI Policy Board is discussing its transition to the EGI Council and the setup of the EGI organization with its funding. In May and June 2009, interested NGIs are expressing their willingness to contribute financially and with labor to the setup and operation of EGI, with first payments on membership fees expected in October 2009. In addition, a proposal for funding to the European Commission is being prepared, which hopefully provides the necessary seed money for the setup of EGI.

Assuming that everything on this roadmap occurs according to the plans, the EGI Infrastructure should be a "large-scale, production grid infrastructure built on national grids that interoperate seamlessly at many levels, offering reliable and predictable services to a wide range of applications". With this in mind, the future European grid infrastructure is offering a series of challenges, which need to be addressed by corresponding action within the IT community:

- Scalability: The expected size and the number of resources provided in the EGI infrastructure require a federated and hierarchical approach, not only to coordinate the interaction between the NGIs, but also to provide the corresponding functionality. For example, repositories on system configuration should be available on and across a local, regional, national, and international basis, such that two persons from different countries are able to collaborate across institutional and national boundaries.

- Manageability: Corresponding to the scale of the system and the intrinsic complexity of grid environments, management features are also a clear need in production infrastructures. This includes simple tools to manage resources, services, and VOs, but also dedicated management processes for various aspects of the grid environment, including sophisticated functionality for applying and utilizing Service Level Agreements (SLAs), and corresponding management processes. A possible example is the request of a user community for resources within a specific VO, where the grid should provide means to obtain these access capabilities fast and as easy as possible.

- Reliability: The complexity of the infrastructure is also a potential source of failures during the operation of the grid. Yet, stability is a necessary characteristic of production environments, which may not be the focus in experimental grids, proofs-of-concept implementations, or today's grid testbeds. For this reason, means of ensuring fault tolerance as well as rollback and recovery mechanisms must be included in the functionality of the environment. Corresponding verification and testing mechanisms must be provided to ensure that the middleware is of sufficient quality.

- Interoperability: Besides the size of the infrastructure, it is also of utmost important to respect the autonomy of the involved institutions, in particular also of the NGIs. This involves not only the possibility of NGIs to provide different middleware stacks or specific features of a middleware, but also the fact that different legal systems require different means of operation on some aspects. The goal must therefore be to specify the interfaces of the middleware, to certify is adherence to specific standards, and to describe the semantics of interoperability in a way that it can be mapped onto different national requirements.

- Integration: Aside from the combination of different grids in different institutions and countries, it is also time to increase the efforts on the integration of grids with other e-Infrastructures, e.g. supercomputers, clouds, or data repositories. In many service provisioning centers, the integration of these different technologies under a common roof is an everyday task, which, however, suffers from that fact that similar tasks are often addressed in different ways. For example, access mechanisms on HPC machines usually require the traditional username and password token, while grids more and more rely on certificates of different kinds. Thinking about the integration of these different information technologies may provide potential synergies for the operation of each of them.

- Higher level services: Assuming that the EGI grid infrastructure provides the basic layer of the European grid, it is clearly also possible to investigate topics related closer to the applications. With such higher level services and new functionality, e.g. graphical user interfaces for working on the grid, functionality for interactive access to grids and software technology for scientific visualizations, new application domains may become interested in using the grid, thereby continuing the growth of the user basis and the potential customers. In this sense, cloud systems are prominently featured in the discussion today may be seen as a higher level service abstracting from the low-level complex processes of grids. This higher level of abstraction thus allows easier access of the grid infrastructure and at the same time introduces new functionality (e.g. the payment mechanisms of grids.)
- Accounting & billing: Finally, the provisioning of grid services is something that user communities often expect as free or inexpensive. Yet, the costs for the services may be substantial and it is important to investigate these costs, the account for the usage of resources and to provide corresponding means of billing. In fact, within EGI it has been formulated that service charges are a possible source of funding the EGI operations. However, due to the lack of corresponding functionality, service charges are not applied initially. Ultimately, the goal is that individual user communities pay for those services, which they specifically need for their work. More advanced business models, e.g. trading of resources must be considered in this framework.

The functionality requirements indicated above provide a large field of investigation for computer scientists, but also a large number of opportunities for industrial businesses interested in providing solutions or services for the grid community. With the setup of the EGI, we are currently experiencing exciting times that are expected to transform the way we are using the e-Infrastructures today. Yet, there is much more which we didn’t even think about so far.