

## EGEE-III Project Publishable Summary

The Enabling Grids for E-scienceE (EGEE, [www.eu-egee.org](http://www.eu-egee.org)) project closed in April 2010, having nurtured innovative, world-class research across Europe and around the globe, some of which would not have been possible otherwise. Its computing infrastructure, software tools and services and its supporting staff carried on the work begun in 2001 by the pioneering European DataGrid project. This work has been developed continually through three successive project phases: EGEE-I, -II, and -III.

The project's main achievements are:

- The creation of the largest collaborative production grid infrastructure in the world for e-science, demonstrating that such a production infrastructure can be used by a wide range of research disciplines, and producing scientific results in these disciplines which would otherwise not have been possible.
- With EGEE, scientists were able to do more work, on a larger scale, and get results in a shorter time frame. EGEE has formed collaborations within Europe and world-wide, and allowed Europe to collaborate as a whole with other regions. These collaborations will last for years to come.

At its close, EGEE represented a world-wide infrastructure of approximately to 200,000 CPU cores, collaboratively hosted by more than 300 centres around the world. By the end of the project, around 13 million jobs were executed on the EGEE grid each month.

A new organisation (EGI.eu) has been created to continue the coordination and evolution of the European Grid Infrastructure (EGI) with the EGEE Grid forming the foundation. This transition is an important step in ensuring the European research community has access to a distributed computing infrastructure to maintain its leadership position in research and support its work in global collaborations for many years to come. EGI.eu is a long-term organisation, not dependent on short-term funding cycles. It is supported by National Grid Infrastructures, the building blocks of the pan-European Grid.

### **Technological overview: What is a grid?**

Grids are effectively a mechanism for bringing together computing and storage resources located in, owned and operated by different organisations for secure access. By connecting through internet networks a grid is a means for sharing computer power and data storage capacity. It permits the creation of virtual research communities making use of computers located all over the globe to become an interwoven computational resource for large-scale, compute- and data-intensive grand challenges.

### **Main results**

**Applications:** As EGEE's mission is to enable global science, having a diversity of scientific applications running on this infrastructure is one of its key achievements.

This last year has seen the start up of the Large Hadron Collider. The successful collection, distribution, and analysis of data from the four LHC experiments represents a major achievement and demonstration of the grid's success. While the bulk of usage comes from the high-energy physics, astronomy, astrophysics, computational chemistry, earth sciences, fusion, and computer science communities, the full user community runs applications from research domains as diverse as multimedia, finance, archaeology, and civil protection. Many communities are using the infrastructure to examine questions they could never have addressed before.

At its close EGEE-III supports diverse application communities through a series of distinct services. These range from providing assistance in porting new applications to the infrastructure, to a registry of all applications already using the infrastructure that allows new users to discover and benefit from applications that are already being used. Many communities have developed their own applications, services or tools around the EGEE infrastructure – some of which have been recognised and placed in the RESPECT catalogue to enable their reuse. Both these services will be continued within EGI.

Service activities: EGEE offered a broad range of user support services, such as infrastructure operations, including network support, “help-desk” support, documentation, and training.

EGEE's operations were expanded and optimised during EGEE-III to include new Regional Operations Centres, notably in Canada and Latin America, and interoperation with other infrastructures – both in Europe and around the world. In terms of networking support, EGEE-III continuously improved the EGEE Network Operation Centre and migrated the tools to a new institution which will host them for EGI. Additionally new network monitoring and troubleshooting tools have been developed and deployed on sites within the production infrastructure. EGEE's service teams have prepared gLite middleware to be compatible with new internet protocols (IPv6). All source code has been publicly released to help the setting up of the network task for EGI.

Though computing resources increased exponentially during EGEE-III (coinciding with the start up of the LHC) the number of software defects found in production has, however not increased in proportion (remaining flat), thanks in part to the certification and release process managed by the service activities, which ensures a stable flow of software updates for improving the reliability of the infrastructure.

**Middleware:** Middleware is a crucial component of any Grid infrastructure as it provides the ‘glue’ to link the hardware resources within the Grid. The gLite middleware binds the EGEE resources into a single infrastructure to provide seamless access for the project's user communities. The overall goal of the middleware engineering group is to provide and maintain selected middleware services of the gLite distribution to satisfy needs the users. As part of the preparation for the transition to EGI, much attention was devoted to the review of the software process and to restructuring, and to the transition of the middleware development to a formal body, the gLite Open Collaboration which will ensure its continued support.

In the EGI-era the middleware deployed on the infrastructure will come from several projects. The gLite middleware developed within EGEE will evolve further as part of a new project: the European Middleware Initiative. EMI pulls together the middleware experts in Europe, most of whom come from the established community, so EMI preserves the expertise created by previous projects. Working together they will improve and standardise the major existing services from ARC, gLite, UNICORE and dCache to produce simplified and interoperable middleware.

**Training, dissemination and outreach:** Dissemination, outreach and communication are essential parts of EGEE, not only to communicate the achievements and successes of the project to the outside world, but also to attract new users, support existing users and reach out to the business community.

EGEE-III has maintained a high profile at over a dozen key events in the grid calendar, some of which have attracted thousands of delegates. Significant impact has also been achieved in generating press cuttings from press releases during EGEE major events, notably the annual conferences and User Forums, including coverage in the UK's Times, Telegraph and Guardian, as well as in many publications across Europe, the US and Asia.

A large number of articles have also been submitted to commercial publications such as HPCwire, project publications including BELIEF-II's Zero-In, and various magazines targeting policy makers such as research.eu and Seed magazine. The EGEE project liaised with the Journal of Grid Computing (JOGC) to produce a special issue of the JOGC with the theme 'EGEE applications and supporting grid technologies', to be published in September 2010. Over 120 articles have been submitted to the weekly online grid publication International Science Grid This Week (ISGTW). Many of these articles have subsequently been linked to by independent blogs and newsfeeds.

EGEE has reached out to businesses through the Business Task Force, which participates in the OGF-Europe Industry Expert Group, through the EGEE Business Associates (EBA) Programme and the wider Business Forum. These achievements have included presenting information on the website, running business days, organising business tracks at EGEE events, presentations to individual companies, issuing newsletters and e-announcements.

**Sustained benefits:** EGEE has been an investment for European and global research. And like any good investment, the payoffs will continue. The infrastructure itself will be cared for by National Grid Infrastructures, coordinated by the European Grid Initiative. The middleware engineered and tested will continue to be developed and improved through the European Middleware Initiative project and the gLite Open Collaboration. Expertise in dissemination and outreach will be preserved through a new phase of the GridTalk project, with which EGEE-III has worked closely. Collaborations formed between other European Infrastructures (in HPC and networking) will be retained in the European E-Infrastructure Forum – dedicated to harmonizing the offerings of infrastructures for large-scale public science projects. Many of the relationships developed with EGEE's Collaborating Infrastructure and Application projects in Europe will continue within EGI.

The work EGEE has performed with companies such as DigitalRibbon and projects such as RESERVOIR to integrate grids and cloud computing demonstrates how the grid concept can encompass the latest tendencies and techniques such as virtualization to increase the range of resources available to user communities.

### **Potential impact and use**

The public research community, the European Commission and governments, have invested heavily in data-supporting infrastructures in the past decade. E-Infrastructures are critical because science is fundamental to addressing the problems confronting our planet and our lives. Data is fundamental to science. The science we do now requires ever increasing data sets. Flexible, powerful computing systems are required to support this. Computing does not have the power to save our planet from global warming or energy shortages – it does, though, have an underlying role to play in making it happen.

For example, it has never been more important to have powerful and accurate climate information. Elaborate computer models are the primary tool used by climate scientists and bodies like the Intergovernmental Panel on Climate Change (IPCC) to report on the status and probable future of our earth. Models like those used by the IPCC need data from the atmosphere, land surface, ocean and sea ice, all originating from different communities, along with diverse accompanying metadata (data which describes data). The amount of data that climate scientists need to manage is enormous – on the petascale, yet a broad and global community needs to be able to access and analyze it. This is ideal for a grid solution – where information stored around the world can be woven together without moving databases. The Earth System Grid Federation (a project sponsored by the US Department of Energy, the National Science Foundation and the National Oceanic and Atmospheric Administration) will support the data for the next generation of climate models used by the IPCC in their next assessment.

### **Conclusion**

As the EGEE series of projects come to an end, the project reflects with pride that both the infrastructure and the knowledge it has created will be a lasting contribution for the European Research Area.

EGEE has achieved the largest collaborative production grid infrastructure in the world for e-science. Providing seamless access to a vast computing resource, 24/7, it has demonstrated that such a production infrastructure can be used by a wide range of research disciplines, producing scientific results which without Grid, would not have been possible to achieve. Through EGEE, scientists were able to do more science and on a larger scale, and get results in a shorter time frame. EGEE has formed collaborations within Europe and allowed Europe to collaborate as a whole with other regions worldwide. EGEE has acted as a good showcase of what is possible with production grids, and is pleased to see its legacy continue in a new e-infrastructure landscape under the aegis of the European Grid Initiative and its National Grid Infrastructures.