EDITORIAL

Development of virtual organizations, applications and services for earth science on grid e-Infrastructures

Can Özturan · Vasso Kotroni · Emmanouil Atanassov

Grids emerged as a platform to share computational resources in the early 2000s. The grid concept was generalized to include the sharing of not just computational resources but also other resources such as data, devices, and applications in the so called virtual organizations (VOs). Grids will help us to move into the so-called e-Science era. E-Science refers to computational knowledge generation using various combinations of massive computational resources, data, and devices in highly distributed network environments. E-Science is the next natural step beyond the web-based platforms of the last decade that offers distribution of scientific data and running of applications through web interfaces on client–server architectures. Traditional web-based platforms have several shortcomings. Firstly, data and computational resources are basically decoupled. It is the duty of each scientist to manually download and manage all the data himself. Automation of this process may require the writing of various scripts which may be quite difficult for a user.

C. Özturan (✉)
Department of Computer Engineering, Boğaziçi University,
Istanbul, Turkey
e-mail: ozturaca@boun.edu.tr

V. Kotroni
National Observatory of Athens,
Athens, Greece
e-mail: kotroni@meteo.noa.gr

E. Atanassov
Department of Grid Tech. and Applications, Institute for Parallel
Processing, Bulgarian Academy of Sciences,
Sofia, Bulgaria
e-mail: emanouil@parallel.bas.bg

Secondly, retrieving massive data over the web can be too slow and hence not practical. Grids can solve these problems by offering turn-key solution to such problems: Scientists can just acquire an account on a grid and become a member of a virtual organization associated with a field and instantly have access to all the resources such as applications, machines, devices, and data that they need. Hence, grids can help us to realize e-Science.

Southeastern Europe Grid e-Infrastructure for Regional e-Science (SEE-GRID-SCI) (SEE-GRID eInfrastructure for regional eScience, http://www.see-grid-sci.eu) User Forum took place in Istanbul, Turkey during 9–10 December 2009 and brought together grid-oriented Earth Science community from the region. The User Forum provided a platform for presentation of Earth Science projects in the area of application deployment and service development for grid infrastructures. The user forum focused on three areas: seismology, meteorology, and environment. Thirty one papers were presented in the user forum covering a wide spectrum of applications and application services. The ten papers appearing in this issue were selected because they contributed state-of-the-art grid software and approaches that have the potential for wide adoption in the areas of seismology, meteorology, and environment. The papers presented in the special issue can be categorized as follows:

- Three seismology papers : (i) Seismic data server application service, (ii) Numerical modeling of mantle convection, and (iii) Gridification of fault plane solution.
- Two meteorology papers: (i) Grid implementation of weather research and forecasting model and (ii) Multi-model multi-analysis ensemble weather forecasting on the Grid

- Five environmental papers: (i) Grid computing for atmospheric composition studies in Bulgaria, (ii) Quasi-Monte Carlo integration on the grid for sensitivity studies (iii) Processing of satellite images in space-time domain, (iv) Experiments on ESIP—environment oriented satellite data processing platform, and (v) Practical considerations on the GreenView application development and execution over the SEEGRID.

The tasks carried out by the SEE-GRID-SCI community to develop a regional e-Science platform can be summarized as follows:

1. Building of virtual organizations for seismology, meteorology, and environment in which scientists and engineers could collaborate by contributing applications and sharing of regional data,
2. Gridification of various seismology, meteorology, and environment applications mainly by the use of grid workflows,
3. Building of general purpose application services as well as VO specific application services that could be used by all VOs,
4. Training of interested researchers on the use of resources provided by the VOs.

Figure 1 shows the model used by the SEE-GRID-SCI community to develop an e-Science platform. At the lowest level, a grid infrastructure called SEEGRID was built as part of the greater EGEE project (Enabling Grids for e-Science (EGEE) Project, http://www.eu-egee.org/) with the involvement of several Southeastern European countries. A number of general application services were developed that could commonly be used by all the VOs. An example is the Parallel Grid Run-time and Application Development Environment (P-GRADE). This environment has been used by the Fault Plane Solution and the Mantle Convection seismology applications that appear in this special issue. Another general application service that has been used by most applications is a workflow engine. Some general application services can still be too low level for some scientists who want to immediately focus on his work rather than spend time in learning new tools. Because of this, another layer was introduced in SEE-GRID-SCI that provides VO specific application services. For example, scientists in the seismology field often need to access seismological data. Therefore, object oriented libraries called iterators were developed so that a high level and location-independent means of data access mechanism was
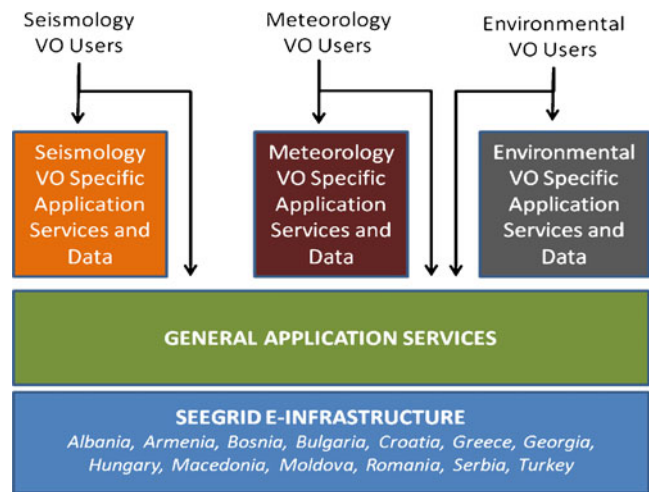
Fig. 1 Architecture of Southeastern Europe's grid based e-Science platform

provided to seismologists. Similar approaches were taken in the meteorology and environment VOs. Implementation VO specific application services have accelerated gridification of applications considerably by relieving the users from the need to learn additional tools outside their fields.

To summarize, whereas early grids originally emerged as infrastructures to share computational and storage resources, what they are evolving into are far more valuable than the mere cost of these resources. Computational resources usually have a lifetime of three years whereas data resources, and software have a lifetime of several decades. Recently, we are seeing the emergence of cloud technologies that may at first sight be seen as a competitor to grid technologies. However, this assessment is misleading. Clouds currently provide on demand raw machines and storage services to anyone on a pay-per-use basis. Because of the backing of industry, it is likely that clouds will provide these raw resources more economically than the grid infrastructures that are usually provided and maintained by nations. However, in order to realize e-Science, we need applications, applications services, field data, field specific communities, and field specific application services so that a scientist can readily use these field specific resources rather than spend months assembling these resources before he can get down to his field specific real work. Virtual organizations on grid infrastructures are exactly providing these high level field specific resources and therefore grids will continue to attract scientists for many years to come.