EDITORIAL

Grid in earth sciences

**M. Petitdidier · R. Cossu · P. Mazzetti · P. Fox ·
H. Schwichtenberg · W. Som de Cerff**

The Earth Sciences (ES) community, with its mosaic of disciplines and players such as academia, industry, national surveys, international organizations, has specific requirements described next. In particular, any observation depends on four coordinates (three spatial dimensions and time) and then needs geospatial tools for its use. The data

M. Petitdidier (✉)
Laboratoire Atmosphères, Milieux, Observations Spatiales
(LATMOS), Institut Pierre Simon Laplace (IPSL),
10-12 Avenue de l'Europe,
78140 Velizy, France
e-mail: monique.petitdidier@latmos.ipsl.fr

R. Cossu
Directorate of Earth Observation Programmes,
European Space Agency ESA-ESRIN,
00044 Frascati, Italy
e-mail: Roberto.cossu@esa.int

P. Mazzetti
Istituto di Metodologie per l'Analasi Ambientale (IMAA),
Consiglio Nazionale delle ricerche, C.da S. Loja-Zona Industriale,
85050 Tito Scalo (PZ), Italy
e-mail: mazzetti@imaa.cnr.it

P. Fox
Rensselaer Polytechnic Institute,
110 8th Street,
Troy, NY 12180-3590, USA
e-mail: pfox@cs.rpi.edu

H. Schwichtenberg
Fraunhofer Institute for Algorithms
and Scientific Computing SCAI,
53754 Sankt Augustin, Germany
e-mail: horst.schwichtenberg@scai.fraunhofer.de

W. S. de Cerff
Koninklijk Nederlands Meteorologisch Instituut(KNMI),
PO Box 201, NL-3730 AE De Bilt, the Netherlands
e-mail: sdecerff@knmi.nl

policy is very complex and strict, as many real-time data may be strategic for the country and/or have an economic impact. In addition, the ES community provides short-term and medium-term predictions of weather and natural hazards in real-time and requires for those tasks immediate availability of resources by advance reservation or preemption. Model simulations of a host of phenomena relating to the Earth and its space environment need access to various large sets of data distributed geographically in different data centres. The various sources of data, among others, include satellite missions, observational networks, large instruments and simulations.

To face the data deluge and more and more complex simulations, the ES community has started to use new technologies such as the Web and Grid. Web services were rapidly adopted by the ES community in particular to access and download data. The ES community started using the Grid around 2000. The experience acquired around the world via several academic and R&D applications has demonstrated that Grid infrastructures could respond to the complexity and constraints imposed by ES applications. However, the interface between the ES software environment and Grid middleware is not simple for many applications. Consequently, Grid technology is not yet widely adopted in ES.

The term "Grid" emerged in the nineties, facilitated by the increase of network speeds, allowing the linking of time-efficient computing and storage resources (Foster and Kesselman 1999). The Grid responded to a pressing need for more computing resources to face and exploit the data deluge. The word "resources" has evolved and covers all that can be shared: computer, server, storage, database, services and so forth.

Basic concepts lying behind the "Grid" are: collaboration, user and provider communities, and security. Users as

well as resources must be authenticated by a certification authority and belong to a recognized *virtual organization (VO), i.e.* a user grid community authorized to access the resources, authenticated, and dedicated to this VO. The authentication credential, a personal certificate, certifies that the user is a known person belonging to a partner member of this user Grid community. Within a VO, sub-groups may be created to limit the access to some resources, especially the data or services. The concept of a virtual organization to access resources exists only in Grid application areas and is well suited to the current description of projects in the Earth Science.

Besides the Grid, the Web is a commonly used infrastructure for enabling ES applications. Many international initiatives-such as GEOSS (Global Earth Observation System of Systems), INSPIRE (Infrastructure for Spatial Information in Europe) and GMES (Global Monitoring for Environment and Security) base their architectural approach on the Web. Indeed most relevant solutions for providing access to Earth Sciences data, e.g. the OpenGeospatial Consortium (OGC) and the Open source Project for a Network Data Access Protocol (OPeNDAP) specifications for geospatial data sharing services-are based on Web services. Grid and Web service approaches have much in common as a result of their underlying Internet technology. However, they show some differences as well. For example, the Grids are based on asynchronous and stateful services while Web services are generally synchronous and stateless. Therefore, many existing initiatives and projects are addressing the harmonization of Grid and Web architectures for Earth Sciences applications especially concerning geo-information services.

In order to disseminate the Grid technology and continue building the Grid ES community, two "Grid" sessions were proposed and accepted for the first time in 2008 at the European Geosciences Union (EGU) General Assembly within the new section, Earth and Space Science Informatics (ESSI). Between the two sessions, 26 presentations, from Europe and the United States, provided an overview of the Grid successes and potentiality. The attendance, around 50 for both sessions, was relatively high in regard to the number of parallel sessions at the EGU meeting. Of the contributed presentations from the conference, 12 contributed papers were reviewed and selected to appear in this volume of Earth Science Informatics. These contributions represent an excellent cross-section of applications and developments that have taken place in several disciplines of Earth Sciences and the needs that have emerged and the developed solutions.

First, Grid technology requires a special infrastructure (i.e. a specific configuration of network, servers, storage, computing clusters) and uses specific software to manage and access it: the middleware (e.g., gLite, Globus Toolkit,

Gria). The existing middleware has been developed with different characteristics to fulfil requirements of the resource and service providers, the application end-users, the type of collaboration or partner management, etc. The software in most widespread use is the **Globus toolkit** that has the largest range of high-level services and permits users to easily build their own services, in particular to interface with Web services. It is currently used by thousands of sites in business and academia, however most of these sites are not inter-connected. An example is given in the architecture build for processing of digital elevation models (Lanig and Zipf). **gLite** is the middleware of the largest EU Grid deployment today, Enabling Grids for E-Science (EGEE, http://www.eu-egee.org/), that is designed for the analysis of the petabytes of data that will be produced by the European Organization for Nuclear Research's (CERN) Large Hadron Collider experiment in Geneva. Access to EGEE is not restricted to high energy physics and is currently used by other scientific communities mainly in public research, including Bioinformatics, Earth Sciences, and Astronomy. In March 2009, EGEE was deployed at more than 300 sites. It provides more than 80,000 CPUs, more than 20 petabytes of storage, and it is capable of running up to 100,000 concurrent jobs. Fernandez-Quirelaz et al. (this issue) and Clévédé et al. (this issue) present applications ported to gLite on climate and seismology, respectively; and Mazzetti et al (this issue) present an application related to the civil protection area for forest fires. The Grid middleware **GRIA** is designed specifically for business Grid applications by supporting their core requirements. The European project SIMDAT focuses on Grids for industrial product development, and developed various applications using and enhancing GRIA. The virtual organization in the Meteorological activity is one of the applications of the SIMDAT project (Raoult et al.).

The middleware has to be able to integrate different services and tools, already existing or not yet developed, to fulfil user requirements. There are outstanding questions of how to evaluate the effectiveness for newly developed and deployed Grid tools and middlewares. Som de Cerff et al. (this issue) propose test suites, elaborated from a variety of Earth Sciences applications. These suites provide testing for the functional and non-functional aspects of the Grid infrastructure in real applications. An atmospheric chemistry test suite is discussed in regard to data access tools and results shown.

As noted earlier, there are several efforts developing infrastructures in ES based on both Grid and Web services, as well as peer-to-peer architectures, e.g. the OPeNDAP Back-End-Server extensions in the paper by Garcia, et al. (this issue) indicate how service oriented applications can be deployed in a Grid environment. It seems essential that Grid systems and Web service systems of the future be fully

compatible or at least that interoperability of the two approaches is essential for Earth Science.

Data are a critical in ES because they are used for climatological or event studies and as input in models and simulations. First, the existing databases are geographically distributed in different data centres according to the sensors, the project and/or the topic. Typically, a strict data policy limits their use to authorized people or institutions. Different Grid tools have been developed to provide an interface with databases and offer different services. Several examples are presented in this issue; AMGA has been built as an interface with EGEE that provides the different services needed (Fernandez-Quirelas et al.). Other examples are provided by GRelC (Fiore et al.) and the Open Grid Service Architecture (OGSA; Lanig and Zipf, Raoult et al.).

Whether the data are observations or simulation outputs, they need geospatial technology for mapping features on the surface of the Earth, in particular geographical information services. Specific geographical web services have been developed and used by many scientific and societal applications. In many cases, high performance computing and storage capacity are required and the need to link Grid computing to those (usually) Web services has been one solution. The Open Geospatial Consortium (OGC) developed common standards and protocols to promote interoperability of data and services across a distributed network. Different applications in this issue, such as forest fire management (Mazzetti et al.), flood prediction (Kussul et al.), and digital elevation modelling (Lanig and Zipf) have defined architectures linking OGC Web Services (OWS) and the Grid.

The papers selected for this issue span a range of ES disciplines that have already used or tested Gris applications: atmospheric chemistry (Som de Cerff et al.), climate (Fernandez-Quirelas et al.), hydrology (Lecca et al.), seismology (Clévédé et al.), solar astronomy (Garcia et al.) as well as activities linked to operational meteorology (Raoult et al.), cartography (Lanig and Zipf), and risk management for fire (Mazzetti et al.) and flood (Kussul et al.). In each case, new functionalities or services have been developed and provided to fulfil the specific application requirements. Simple workflows have been developed to manage and monitor the execution of jobs and avoided

some failures (see Clévédé et al.; Fernandez-Quirelas et al.; Garcia et al.). Importantly, despite an increasing complexity, platforms including all the components for a given application or a set of applications have been developed (Murgia et al.). There are several examples given in this issue. One has been tested for the civil protection scenario such as for forest fires (Mazzetti et al.), the other, AQUAGRID, is devoted to sub-surface hydrology applications (Lecca et al.).

Many believed that by allowing all components of the information technology infrastructure - computational capabilities, databases, sensors, and people - to be shared flexibly as true collaborative tools, the Grid would allow new classes of applications to emerge. This vision has started to be a reality. However not all of the potentialities offered by the Grid have been exploited, especially e-collaboration that would permit not only sharing of computer and storage resources and services, but also the sharing of knowledge in different ways. In an effort to highlight progress and chart a future course, the European FP6 project DEGREE (Dissemination and Exploitation of Grids in Earth Science) developed and elaborated upon a roadmap with different phases along with key objectives coming from geosciences program requirements. Starting from the specific requirements of ES applications a roadmap towards an Earth Science Grid platform is proposed and discussed (an article by Cossu et al. will appear in a future version of this journal describing that effort).

The papers presented in this special issue provide an excellent overview of the developments in Grid technology in use for Earth Science and successful results. We hope to encourage further developments in Grid technology to meet the ES needs. We also want to encourage ES scientists to embrace this useful technology to assist in their research.

Reference

Foster I, Kesselman C (1999) Globus: A Toolkit-Based Grid Architecture. In: Foster I, Kesselman C (eds) The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, 1999, 259–278