RESEARCH ARTICLE

A roadmap for a dedicated Earth Science Grid platform

Roberto Cossu · Monique Petitdidier · Julian Linford · Vincent Badoux · Luigi Fusco ·
Benoit Gotab · Ladislav Hluchy · Giuditta Lecca · Fabrizio Murgia · Camiel Plevier ·
Philippe Renard · Horst Schwichtenberg · Wim Som de Cerff · Viet Tran ·
Gerald Vetois

Abstract Due to its intensive data processing and highly distributed organization, the multidisciplinary Earth Science applications community is uniquely positioned for the uptake and exploitation of Grid technologies. Currently Enabling Grids for E-sciencE, and other large Grid infrastructures are already deployed and capable of operational services. So far however, the adoption and exploitation of Grid technology throughout the Earth Science community has been slower than expected. The Dissemination and Exploitation of GRids in Earth sciencE project, proposed by the European Commission to assist and accelerate this process in a number of different ways, had between its main goals the creation of a roadmap towards Earth Science Grid platform. This paper presents the resulting roadmap.

Introduction

One of the most fundamental challenges facing humanity at the beginning of the 21st century is to respond effectively to the global changes that are putting increasing pressure on the environment and on human society.

For a long time the observations taken from various sensors have been an effective way to discover and study the behaviour of the Earth System with all its components—such

R. Cossu (✉) · L. Fusco
Directorate of Earth Observation Programmes, European Space
Agency ESA-ESRIN,
Frascati 00044, Italy
e-mail: roberto.cossu@esa.int

M. Petitdidier
Laboratoire Atmosphères, Milieux,
Observations Spatiales,
Saint-Quentin Versailles University,
10-12 Avenue de l'Europe,
Velizy 78140, France

J. Linford
SERCO SpA,
Frascati, Italy

V. Badoux · P. Renard
Université de Neuchâtel,
Neuchâtel, Switzerland

B. Gotab
Institut de Physique du Globe de Paris,
4 Place Jussieu,
Paris, Cedex 05 75252, France

L. Hluchy · V. Tran
Ustav informatiky, Slovenska akademia vied,
Slovenska, Slovak

G. Lecca · F. Murgia
Center for Advanced Studies, Research and Development in
Sardinia (CRS4),
Pula 09010, Italy

C. Plevier
Dutch Space BV,
Leiden, Netherlands

H. Schwichtenberg
Fraunhofer Institute for Algorithms and
Scientific Computing SCAI,
Sankt Augustin, Germany

W. S. de Cerff
Koninklijk Nederlands Meteorologisch Instituut,
De Bilt, Netherlands

G. Vetois
Compagnie Générale de Géophysique,
Paris, France

as the atmosphere, the ocean, the solid Earth and so forth–and the impact of human being. The volume and the quality of the observations are increasing due to global and permanent observation networks and satellites. As a result, vast amount of data sets and databases related to different disciplines are nowadays archived in different locations and owned by different organizations.

Until the second half of the twentieth century, the different components of the Earth System were studied almost independently ones of each other. In the second half of the twentieth century, a better knowledge of the Earth components emerged—including their interactions and, their forcing and feedback mechanisms. A new concept was proposed based on a global approach of the Earth System. Nowadays, our total knowledge about this complex system is contained in models and measurements; how we put them together has to be managed cleverly. Coupled equations making up complex models have been developed to describe the Earth System and their resolutions have required supercomputers.

With regard to the data (observations and model outputs), there has been the explosion in the size of databases, with data from a plethora of different sources. The fact that petabytes of already acquired data are distributed in different locations and even in different formats makes it difficult for scientists to discover and access heterogeneous data needed for deriving knowledge about the Earth System. To face this data deluge the Earth Science (ES) community has started to use new technologies such as the Web and Grid. The Web services were rapidly adopted in particular to facilitate the access to data, their processing and visualisation. Based on this technology, the Earth Science community has developed portals, which present information from diverse sources in a unified way, considering several standard and specific-ES web services. In some cases the portal provides access to resources for dynamic processing of ES specific datasets and for high performance processing. However, even if they were to be made available, an efficient infrastructure to handle and treat very large data sets and combine them with computational resources for exploitation and modelling is still missing. The technical challenge is to put together databases and computing resources to answer the ES challenges.

Initiatives like the Global Monitoring for Environment and Security (GMES),[1] the Infrastructure for Spatial Information in Europe (INSPIRE)[2] and the Global Earth

Observation System of Systems (GEOSS)[3] base their approach on Web-service infrastructures and relevant data-access solutions to Earth Science data, e.g. specifications for data sharing by Open Geospatial Consortium (OGC)[4] or the Open-source Project for a Network Data Access Protocol (OPeNDAP).[5] 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 Grid services are based on stateful and transient services, while Web services are in general stateless. Therefore many existing initiatives and projects are addressing the harmonization of Grid and Web architectures for ES applications, especially concerning geo-information services. The Open Grid Services Architecture (OGSA),[6] a service oriented architecture describing the capabilities and specifying the services needed to enable a Grid, represents an evolution towards a Grid architecture based on Web-service technologies.

Grid computing is in general considered as a solution for scaling up computing power and storage capacity by sharing resources among institutions. Due to the huge need in intensive computing it has been expected that the ES community should adopt and exploit Grid as rapidly as for the Web services. As opposite the adoption of Grid in the ES has been slower than expected. This delay is mainly due to insufficient information and difficulties to implement an usual ES environment. The Dissemination and Exploitation of Grids in Earth Science (DEGREE) Project[7] was proposed to the European Commission (EC) in order to assist and accelerate this process in a number of different ways. Over a period of 2 years, 2006–2008, the consortium of ES academic and private research institutions, space agencies, and computer science institutes, conducted a survey of the tools, software, and protocols commonly used by the ES community and those provided by Grid projects. DEGREE addressed three main topics: data management, job management, and portals to integrate Earth science and Grid tools. The survey results, which are published on the DEGREE web site, were discussed with Grid project developers and applications proposed to test critical functionalities (Som de Cerff et al. 2009). Following those surveys and identification of the gaps, one of the main outcomes of DEGREE is a roadmap (Cossu et al. 2008) aimed to set out the key steps needed for the ES

---

[1] http://www.gmes.info/. GMES is an initiative for the implementation of information services dealing with environment and security. It is based on observation data received from Earth Observation satellites and ground based information.

[2] http://inspire.jrc.ec.europa.eu/ INSPIRE is a European Commission initiative to build a European spatial data infrastructure beyond national boundaries.

[3] http://www.earthobservations.org/ GEOSS seeks to connect the producers of environmental data and decision-support tools with the end users of these products, with the aim of enhancing the relevance of Earth observations to global issues.

[4] http://www.opengeospatial.org/

[5] http://opendap.org/

[6] http://www.globus.org/ogsa/

[7] http://www.eu-degree.eu

community to move towards achieving its Grid objectives. These are believed to play a fundamental role to achieve some of the big challenges that ES is nowadays facing. The major objectives of the roadmap are set out as a journey progressing through a list of objectives to be achieved in order to arrive at the final overall objective, i.e. the establishment of a well-identified ES Grid platform. Such a roadmap involves different scientific and technical actors in ES and Information Technology, policy makers, and so on. For this reason, it has been decided by the authors to herein present the roadmap at the highest possible abstraction level, trying to address an as wide as possible audience. The DEGREE roadmap has been considered as one key input for the implementation of the Ground European Network for Earth Science Interoperations—Digital Repositories (GENESI-DR)[8] project, which is currently aimed at providing a single access point for scientists to discover, access, and process a huge amount of heterogeneous ES data.

The aim of this paper is to present this roadmap to a wide audience. For details, either technical or application related we refer the reader to Cossu et al. (2008). The paper is organized as follows: "Grid and ES community experience" introduces the reader to the Grid, with special focus on projects relevant to ES. "Grid ES platform" analyses ES requirements and challenges relevant to the scope of the paper, and discusses how Grid can help ES scientists in addressing their challenges. "Earth Science Grid roadmap" details the roadmap proposed by DEGREE project to establish an ES Grid platform. Conclusions are drawn in "Conclusions". The paper is mainly written keeping in mind the European context, but many findings and considerations can be easily extended to a worldwide context.

Grid and ES community experience

Introduction to Grid

Grid characteristics

In this subsection, we present an overview of the main Grid characteristics without discussing detailed implementations and solutions. For such details we refer the readers to (Foster and Kesselman 1999; Wang et al. 2009; Castro et al. 2008; various authors 2006).

The term "Grid computing" originated in the 1990s as a metaphor for making computer power as easy to access as an electric power grid (Foster and Kesselman, 1999). Today the term "Grid" is used for a wide variety of different solutions. A widely used distinction is between Computational and Data Grids according to their focus on the provision of compute cycles or data related solutions; into Science Grid or Business Grids according to the category of users and applications; or into Health Grids, Earth Observation Grids or Finance Grids, according to the user community involved.

As a common denominator, a Grid can be characterized as a system of software and hardware that allows for coordinated and secure sharing of resources that are not centrally controlled. These resources can among others be in the form of computational capabilities, data repositories, software services or applications. Coordinating this resource sharing typically involves the formation of one or several so-called Virtual Organisations (VO). A VO determines the conditions under which resources are and given access to. For example, scientists of the same discipline might share their computational resources, algorithms and data files in a Virtual Organisation.

Grids are often based on long term collaboration between partners (resource providers and user communities) with a trust agreement to guarantee security and confidentiality. A Grid is in general more flexible, as the resources could be aggregated after certification or withdrawn partly or entirely in a short time. As a matter of fact the resources are under the control of the domain to which they belong.

Middlewares

The software used to achieve the previously described features, , is called a Grid-Middleware. The Grid-Middleware is responsible for the operation of the Grid, and allows uniform access to the often highly heterogeneous resources. As a Grid often spans multiple administrative domains, hardware and software installations vary greatly. The Grid-Middleware usually hides this heterogeneity to the user. Until today, several, often architecturally different, Grid-middlewares have been developed or are still being actively developed.

Commonly used middleware in Europe are for example gLite,[9] UNiform Interface to COmputing REsources[10] (UNICORE), Globus Toolkit,[11] Advanced Resource Connector (ARC), and Grid Resources for Industrial Applications (GRIA[12]). The most widely used is the Globus toolkit that has the largest range of higher level services and permits the users to easily build their own services, in particular interface with Web services. It is used currently

---

[8] http://www.genesi-dr.eu, Ground European Network for Earth Science Interoperations—Digital Repositories

[9] http://glite.web.cern.ch/glite/
[10] http://www.unicore.eu/
[11] http://www.globus.org/
[12] http://www.gria.org/

by thousands of sites in Business and Academia but mostly not inter-connected. gLite is the middleware of the largest Grid deployment today, Enabling Grids for E-SciencE (EGEE[13]), that is also part of the Worldwide LHC Computing Grid (WLCG) for the analysis of the petabytes of data that will be produced by the European Organization for Nuclear Research's (CERN) Large Hadron Collider (LHC) 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, around 300 sites are part of the EGEE project. They provide more than 80,000 CPUs, more than 20 petabytes of storage, and they are capable of running up to 100,000 concurrent jobs (Wang et al., 2009). The EGEE grid solution was deployed in Europe and also in the Mediterranean area and Latin America. The Grid middleware GRIA, is designed specifically to support business users of the Grid by supporting the core requirements for business Grid applications (Castro et al. 2008; Various Authors 2006). UNICORE is a Grid computing technology developed in two projects funded by the German ministry for education and research (BMBF). UNICORE is used in daily production at several supercomputer centres worldwide, for example in the Distributed European Infrastructure for Supercomputing Applications (DEISA). Beyond this production usage, UNICORE serves as a solid basis in many European and international research projects. ARC is developed and used in the NorduGrid activity since the year 2002, of which the actual Grid currently consists of approximately 34000 CPUs.

Interoperability

Recognizing this variety of technological solutions, efforts have been carried out to homogenize, or allow for interoperability of Grids. In 2006, the Global Grid Forum and Enterprise Grid Alliance, two initiatives for standardisation efforts in the area of Grid, merged to form the Open Grid Forum (OGF[14]). To enable interoperability between the above mentioned middlewares, the OGF published among others OGSA, a service oriented architecture describing the capabilities and specifying the services needed to enable a Grid. It is based on and respects Web-Service Standards such as the Web Service Description Language (WSDL), WS-Addressing, SOAP and others.

Security

Space remotely-sensed data and ground measurements play a very important role in ES for variability and for volumes. European archives contain already many PetaBytes and many more sources are already planned for the near future. As ES data are usually not free of charge and complex data policy governs their access, ES community pay a great attention to Grid security. Due to high-competition among teams, confidentiality on the application ported (e.g., code, data) is also a key issue for Grid security. Grid security is provided by a pervasive and seamlessly integrated security mechanism, essential for access control in a shared environment. As an example, in the EGEE infrastructure, each user is authenticated by means of a personal certificate delivered by one or more certificate authorities recognized by the project. This certificate indicates that the user belongs to a recognized partner of the project. Once the user has obtained a certificate, he/she needs to join a VO that gathers people with similar interests, common goal and requirements, and able to work collaboratively and to share resources (data, software, CPU, storage space, etc) regardless of their geographical location. It is mandatory to be part of a VO to be authorized to access the resources made available for this specific VO. Different VOs may share and access the same resources. All the members of a VO may have not the same rights relatively to access the resources or to perform some actions such as running specific applications, updating or deleting data. Then inside a VO different groups can be created with different rights and roles. The concept of virtual organization to access resources in Grid application areas is well suited to the current description of projects in the Earth Science.

Project classes

Since the emerging of the Grid concept, a large number of projects on Grid technology have been carried out in many countries. Those projects may be divided in three classes. In the first class, Grid is an object for research and for investigation in order to prepare the Next Generation of Grid (CoreGrid,[15] NESSI[16]). The second class is concerned by developing middleware and deploying it on testbeds (see details in the following of this chapter). The third class is focused on developing architectures and tools implemented on middleware, developed in the previous class, to fulfil requirements of specific applications. This last class deploys small or large test-beds for the assessment of Grid middleware and tools and as a consequence influences the

---

[13] http://www.eu-egee.org/
[14] http://www.ogf.org/

[15] http://www.coregrid.net/
[16] http://www.nessi-europe.com/Nessi/

technology development. The ES community is more concerned by this third class of projects

Experience of ES community in Grid

The ES community (research and industry) has already large experience in Grid and web services. The European ES community has explored the Grid technology since 2000 via European projects for research such as Data-Grid[17] and EGEE in different ES disciplines. Those infrastructure projects are being driven by the user communities and based finally on the gLite middleware. Other projects such as int.eu.grid[18] or Cyclops[19] aim at developing and deploying services linked to gLite for specific applications. The European Centre for Medium-Range Weather Forecasts (ECMWF) participated to the European project SIMDAT[20] devoted to industrial or operational applications, using Gria middleware. The European Space Agency (ESA) Grid Processing on Demand (GPOD)[21] (Fusco et al. 2007) offers world-wide scientists with an Earth Observation (EO) dedicated and user friendly environment. The Earth Science Grid (ESG)[22] is aimed at providing a seamless and powerful environment that enables the climate research.

The fact that many disciplines of the ES community have been well organized since decades, in regard to the metadata and data format, data exchange, algorithms and models, finds a natural implementation in the concept of VOs. Although most references presented in this paper come from the European ES community, other key Grid infrastructures, deployed in the USA and elsewhere are used by the Earth Science community, such as Open Science Grid (OSG),[23] and TeraGrid.[24]

Experience with DataGrid and EGEE

The first ES application ported in DataGrid involved scientific teams from the Koninklijk Nederlands Meteorologisch Instituut (KNMI), the Institut Pierre Simon Laplace (IPSL), the University of Tor Vergata, and the European Space Agency (ESA) and represented a classical scenario for handling space Earth Observation data and ground measurements. In fact, the Global Ozone Monitor-ing Experiment (GOME) sensor, mounted on board of the European Remote Sensing (ERS) satellite, is able to measure the total column of Ozone in the atmosphere. Different algorithms allow the scientists to estimate the vertical profiles of Ozone starting from the total column. As opposite, Ground-based Lidars measure directly and the vertical profile of Ozone. While satellites provide global coverage with limited time coverage, ground-based instruments provide measures only in a narrow geographical area but continuously. Hence scientists are interested in comparing/validating the Ozone profiles obtained processing GOME acquisitions with Lidar measures. The DataGrid ES application retrieved Ozone profiles from GOME data using a previously trained neural network. The validation was done by comparison with the ozone profiles measured with Lidars (Iapaolo et al. 2007).

The main results and advantages of using Grid could be summarized in:

- 7-years of GOME data were processed to extract Ozone profiles. Two versions of the neural network algorithm were used. Approximately 70,000 files were stored and handled on Grid;
- No need to download the data as they are directly accessible from the Grid computing resources for all the authorized partners;
- No need to convert data due to different computing platforms;
- All the retrieved data were validated with all the Lidar data available. To this end, only one metadata base with geospatial extension for collocation of satellite and Lidar data was created instead of one per institute.

In conclusion, the use of Grid made easier and fruitful the collaboration among the different teams involved in the project.

Other applications in various ES disciplines (Various Authors 2009) have been ported on EGEE to get more computing resources for parametric and statistical studies or when a specific event such as an earthquake occurs, or to fill up a database. EGEE has offered computing resources that permit in a reasonable time to process and store large sets of data, and to compute a statistically significant ensemble of simulations. It permits also to run jobs and to share data, algorithms and software among a VO or a group. Both of these capabilities concern important needs in ES. Another positive point is the collaboration of different ES disciplines and other domains that is very fruitful to find solutions of common problems.

In the framework of the EGEE project, the infrastructure has been deployed in hundreds of sites all over Europe. This Grid is interoperable with other Grid infrastructures deployed in China, Japan, Latin America, Mediterranean countries and USA. The experience gained with EGEE

---

[17] http://eu-datagrid.web.cern.ch
[18] http://www.i2g.eu/
[19] http://www.cyclops-project.eu/
[20] http://www.ecmwf.int/services/grid/simdat/
[21] http://eogrid.esrin.esa.int/
[22] http://www.earthsystemgrid.org/
[23] http://www.opensciencegrid.org/
[24] http://www.teragrid.org/

points out the flexibility and scalability of such a Grid in adding new resources and end-users partners. The ES community, typically working in collaboration on a worldwide base is very keen to be involved in joint such deployments.

However the interface between ES environment and Grid middleware is not simple for the integration of specific applications or for the development of new models. For example in the handling of generic geospatial datasets, community specific Web services (e.g. Web Map Services) have been developed to visualise Geographic Information System (GIS) data and to access and process the data stored behind the screen. As it concerns GIS with the associated infrastructures, i.e. Spatial Data Infrastructure (SDI), that support information discovery, access, and use of this information in the decision-making process, the OGC has developed common standards and protocols to guarantee interoperability of data and services across a distributed network. As a case, in the framework of the Cyclops project, different applications, such as fire extension and flash flood mapping, need Grid for the data analysis and OGC standards for their visualization. A platform aggregating the different layers has built in the project (Mazzetti et al. 2009).

The recent enlargement of the ES VO in EGEE (in terms of both committed resources and users) has demonstrated the substantial progress achieved in awareness and utilization of the emerging EU Grid infrastructure and technology. However, it has also been pointed out that more and continuous effort is necessary to ensure long term success and fulfilment of the ES Grid community vision and objectives.

Experience with Gria

The SIMDAT project[25] delivered a Grid solution portfolio that is designed to meet the needs of industry. The portfolio offers Grid technologies and Grid-enabled Problem Solving Environments that can be used to build domain-specific Grid solutions. SIMDAT has developed new management infrastructure for Gria based on Service Level Agreements that allows service providers and customers to trade resources.

Among the industrial applications, there was one related to operational meteorology. The leader was the ECMWF with meteorological centres of Germany (Deutscher Wetterdienst), France (MeteoFrance), and United Kingdom (the UK Met Office) and another European Organization EUMETSAT. Data exchanges are the core task within the World Meteorological Organization (WMO). The meteoro-

logical centres are responsible for receiving and producing data and distributing them via the Global Telecommunication System (GTS); no computing resources are needed. A need of WMO is to upgrade its information system infrastructure with new technologies and offer new services. A prototype, called the Virtual Global Information System Centre (V-GISC) was developed and validated by other projects and meteorological centres (SIMDAT 2008). This work is now recognized as the pioneering project for the new emerging WMO information system infrastructure (WMO-WIS).

In this ES application Gria was not adopted even if some functionalities or protocols are used especially for security (SIMDAT 2006). The major Gria infrastructure drawback for this application is the fully connected nature of this grid architecture. For example VOs and the role of each member are administrated centrally then for a large number of users whose number and role could change dynamically, this centralization will be a burden. A decentralized architecture was adopted and a detailed description of the solution is in (Raoult et al. 2009).

ESA Earth Observation Grid Processing on Demand

Following its participation to DataGrid, the ESA Science and Application Department of Earth Observation Programmes Directorate has focused on the development of G-POD, a dedicated ES Grid infrastructure,[26] that provides scientists with the capability to plug, in this community environment, specific EO data handling and processing applications (Fusco et al. 2007). Coupled with high-performance and sizeable computing resources managed by Grid technologies, G-POD provides the necessary flexibility for building a virtual environment that gives applications quick access to data, computing resources, and results. Using a dedicated Web interface, each application has access to a powerful catalogue, i.e., the ESA Multi-mission User Interface System (MUIS), to select data from the G-POD storage elements. It furthermore communicates with the underlying Grid middleware, which coordinates all the necessary steps to retrieve, process, and display the selected products from the various available ESA and third-party missions.

As of early 2009, the ESA's centre for Earth observation is in Frascati (ESRIN) internal Grid infrastructure is composed of 5 clusters with a total amount of more than 600 cores and with storage capabilities of about 200 Terabytes. All the elements are all part of the same Grid LAN at the ESRIN, partially interfaced to other Grid elements in partners Earth Science organisations such as the

---

[25] http://www.scai.fraunhofer.de/about_simdat.html

[26] http://eogrid.esrin.esa.int/

Italian National Research Council (CNR), the University of Tor Vergata, Roma, and European Grid infrastructures (e.g. EGEE, and D4SCIENCE[27]).

The key feature of this Grid environment is the layered approach based on the Grid-engine, which interconnects the application layer with different Grid middlewares (at present interfaced with three different brand/releases of middleware, Globus, different versions of the EGEE middleware such as gLite, the one used for LHC). This characteristic enables the clear separation and development path between the EO applications and the middleware being used.

Thanks to this separation, ES scientists can focus on their applications without dealing with Grid technology issues and can take full advantage of the underlying Grid infrastructure.

In 2006, G-POD was formally opened to scientists through the "EO-Grid Processing-on-Demand Call for category-1 Proposals"[28] that explicitly aimed at «promoting the access and use of EO mission data available at ESA, offering on-line access to products with attached computing infrastructure and tools to assist the generation of "scientific added value products"».

It is worth noting that no Grid expertise was required by G-POD invited partners. As a result, EO scientists, rather than being discouraged by the otherwise perceived complexity of the Grid infrastructure, could focus on their EO applications appreciating the easiness in accessing data and computing resources provided by the underlying Grid technology.

Many other EO applications are fully operational and available through the ESA EO Grid portal.

Grid ES platform

ES challenges and priorities

Major challenges of ES, as in the implementation and the systematic monitoring of international environmental conventions, need data, tools and worldwide infrastructures to gather, share and process the data, and to run simulations or models. Today, information about the state of the Earth, relevant services, data availability, project results and applications are accessible only in a very scattered way through different operators, scientific institutes, service companies, data catalogues, etc. A common dedicated infrastructure would permit the ES communities to derive objective information and to share knowledge in all environmentally sensitive domains over a continuum of time (from historical measurement to real time assessment to short and long term predictions) and a variety of geographical scales (from global scale to very local facts). Such a global infrastructure needs to include access to historical data holdings and networks of sensors, broadband communications via ground and space, efficient, effective and distributed computing and storage resources to take care of and handle the scientific tools, methodologies, data, etc.

Furthermore, the different ES community domains are fully involved in dedicated thematic physical infrastructures as proposed by the European Strategy Forum on Research Infrastructure (ESFRI)[29] projects. They use specific existing methods, approaches and working practices for gathering, storing and exchanging data and information. A shared e-infrastructure approach is aimed at facilitating the linking up of disparate community localized infrastructures.

Data play a central role in ES and different aspects need to be considered: data cataloguing, data discovery, digital rights management for data, data access, data interoperability, metadata schemas, metadata exchange, and ontologies. In addition to data access, the requirement for intensive calculation is also important. For example, statistical approaches, the creation of databases or the frequent re-processing of long series of data need intensive computation.

The requirements for re-processing have been strongly increased over the last 10 years. The science community expects rapid adaptation to new needs, fast implementation of new algorithms in routine processing chains, and even fast re-processing of entire archives with these new algorithms. These requirements have to be covered in modern ground segments operational concepts.

Similar needs and concerns are raised by international forums and organizations as key recommendations like the 10-year GEOSS implementation plan (GEOSS 2005) and the 2006 GCOS[30] (Global Climate Observing System) "Systematic Observation Requirements for Satellite-Based products for Climate", where clear statements are made concerning the long term custody of present and future satellite data records and associated metadata, including the provision of open access to these records.

---

[27] http://www.d4science.eu
[28] http://eopi.esa.int/esa/esa

[29] http://cordis.europa.eu/esfri/ ESFRI is a strategic instrument aimed at supporting a coherent and strategy-led approach to policy-making on research infrastructures in Europe, and at facilitating multilateral initiatives leading to the better use and development of research infrastructures, at EU and international level.
[30] http://gcos.wmo.int/

Application areas, benefiting from ES long-term data archiving exploitation and intensive computation are wide ranging and include EC Policies with long-term perspectives. Some examples are the following:

- European and global environment monitoring and forecasting of forests, land and soil, urban development, air quality, ecosystems and their management for protection of terrestrial, coastal and marine resources;
- Management of energy resources (solar, etc.);
- Management of water resources needed for health, food security including sustainable agriculture and combating desertification; water resource management through better understanding of the water cycle;
- Civil protection and disasters monitoring;
- Global climate change; climate understanding for assessing, predicting, mitigating and adapting to climate changes, as well as the improvement of weather information, forecasting and warning.

Summarizing, in future European scale initiatives it is necessary to address the following ES priorities:

- a base for establishing a ES community specific world-wide e-infrastructure;
- guaranteed, reliable, easy, effective, and (standardized) operational access to a variety of data sources, and demonstration on how the same approach can be extended to provide seamless access to all ES data (e.g., integrating remote sensed data and in-situ measurements);
- harmonisation at key ES data repositories limiting fragmentation of the solutions[31];
- demonstration of the effectiveness of the different solutions and creation of the framework for long term preservation of all type of ES data[32];
- validation of the effective capabilities required to access distributed repositories for new communities, including education;
- integration of new scientific and technological derived paradigms in operational infrastructures in response to the latest ES requirements.

Benefits of a Grid based Earth Science platform

As discussed in the previous "ES challenges and priorities", ES community would benefit from a common dedicated generic infrastructure to derive objective information and to share knowledge in all environmental sensitive domains; ES has a need for gathering, storing and exchanging data and information and this is difficult to be collected or managed in one single institute or organization.

A Grid-based ES federated infrastructure aimed at interoperations of services will provide:

- the common dedicated infrastructure for federating and sharing all the available, difficult to reach, ES data sets at the different scales;
- the e-collaboration platform for advancing in multidisciplinary applications with need sharing of experiences, models, tools, and information;
- the required computation resources for large scale ambitious projects requiring simulations and/or for re-processing of entire archives with new algorithms.

Grid improvements for a successful ES Grid based platform

As stated in "ES challenges and priorities", the ES community plays the role of users interested in using and validating the development of emerging ICT technologies. The community has expressed in various projects strong and specific requirements to middleware developers. Example of middleware functionalities improvements that would make even more successful an ES Grid based platform are detailed in (Som de Cerff et al. 2007; Som de Cerff et al. 2008; Schwichtenberg et al. 2007; Schwichtenberg et al. 2008; Viet Tran et al. 2008; Linford et al. 2008). The key requirements for better services can be grouped as in the following:

Simplify access to heterogeneous data

- Simple Application Programming Interfaces (APIs) for data discovery;
- Interfaces for access to heterogeneous and federated data resources;
- Fast transfer of large files and/or large number of different small files;
- Transparent handling of various data formats and access protocols, widely used in the community, e.g. NetCDF[33] and HDF file formats for multidimensional arrays; OpenGIS and OPenDAP networks access protocols;

---

[31] This specific point is at the centre of the GENESI-DR project, which aims to validate emerging standards for federation of large ES repositories.

[32] ESA has initiated a dedicate programme for Earth Observation Long Term Data Preservation (http:// http://earth.esa.int/gscb/ltdp/

---

[33] http://www.unidata.ucar.edu/software/netcdf/

Metadata and data management

- User/role based access control to metadata and data;
- Handling of special types of databases beyond the relational model and SQL queries,; spatial databases are standard backend in OpenGIS Web-map services;

Reliability of processing (e.g., near real time job execution)

- Reliability (all jobs must finish before some deadlines) which translates in. efficient Quality of Service (QoS) and fault tolerance mechanism;
- Advanced reservation of resources for scheduled jobs and for risk alert;

Security

- Standardized Authentication and Authorization mechanisms for usage in portals as well as in direct Grid access;
- User identify management, access permissions control, account settings and customization of the individual user's environment;

Integration with ES Web portal

- Structured, customized organization of the Portal pages according to dedicated application-related themes, activities and functions;
- Easy customizations of the Portal information and content by the real-time integration of contributions from individual users;

e-collaboration

- Provision of specific tools for e-collaboration.

Some of those requirements have been partly fulfilled. Those requirements were addressed to Grid developers in the framework of DEGREE in order to be included in the next Grid Generation; recent projects such as GENESI-DR are implementing a number of such functionalities.

Earth Science Grid roadmap

The European Commission expects the creation of a permanent and sustainable e-infrastructure, the European Grid Initiative (EGI), with a transition period from 2010 to 2013. During this period any application domain will be organized in specialized support centres (SSC) to serve its targeted community in the manner which is most useful to that community. Then ES domain needs a roadmap, today missing, to provide a guideline for the future ES SSC. The ES roadmap is the main objective of this paper. It aims to set out the key steps needed for the ES community to deploy the Grid-based platform as described in the previous section.

It is worth reminding that Grid is herein intended not as a goal but as a means for reaching the ES priorities identified in the previous section. Grid is a commodity ES users can benefit from to pursue their challenging objectives. Nowadays the situation is the following:

- First successful deployment of applications and demonstration of benefits that Grid infrastructure can bring to ES community (Fusco et al. 2007; Renard et al. 2009).
- Middleware gaps identified, ES requirements gathered and analysed and communicated to the Grid developers (Som de Cerff et al. 2009).
- Need for increased awareness and participation of ES community to Grid activities in order to proceed (starting point of DEGREE project)

This is the starting point for the ES Grid roadmap. The major macro-objectives of the roadmap are instrumental to final overall objective, i.e. the establishment of a well-identified ES Grid platform. As such they are related with:

- Providing of an easy and fast access to large volumes of dispersed data;
- Providing computational resources.
- Building up an e-collaboration platform;

These macro-objectives should be reached with a stepwise approach, involving several, sometimes transversal "micro"-objectives. As an example, DEGREE project concluded that all the three above mentioned objectives require community building: to identify and collect ES data owned and hosted by different organizations; to increase the number of available computing resource; to "provide" the actors for a wide and fruitful e-collaboration. In a similar way, porting to Grid of applications from different ES disciplines and facilitating the access for everyday ES users to Grid are all mandatory to reach the overall objectives.

The roadmap has been derived based on an analysis for which DEGREE project has considered various aspects such as existing infrastructure, middleware, network and bandwidth, experiences gained within the EGEE project, etc.

The result is a list of points to be kept as succinct as possible in order to focus on key objectives, as it is more effective to emphasize a few key points to be visited frequently and consistently, rather than leaving too many points scarcely visited.

In this paper, and more specifically in this section, we provide a snapshot of these objectives and related roadmap rather than details that would results in a too technical dissertation. In addition, a detailed roadmap involves consideration related to different aspects such as metadata description, workflow management, web portals etc, and

Table 1 Legend of the symbols used in Tables 2 to 4

	Objectives of fundamental importance for ES, identified during the ES community consultation meetings
	ES specific objectives
	Key objectives identified during the technical analysis of the DEGREE technical work packages
	GEOSS key objective
	SEIS key objective
	INSPIRE key objectives

would probably result in a too complex dissertation (DEGREE involved partners with different expertises in these fields and this allowed their comprehensive analysis; the full detailed roadmap is available at http://www.eu-degree.eu/DEGREE/internal-section/wp6/DEGREE-D6.1.2_v2.8.pdf)

We have arranged the key points by category, specifying for each of them the expected achievements and the time period, from short-term to medium-term to long-term.

In addition, keeping in mind the high level description we want to provide, we have highlighted some key objectives by assigning an icon to indicate their specific nature and provenance. We have highlighted: objectives that have been identified as of fundamental importance by consulted ES scientists, i.e., by the user community; objectives that are quite specific of ES; more technical objectives that have been identified as critical during the technical analysis of the DEGREE project activities. This is believed to provide a tool for a more rapid reading of the roadmap by different readers, such as ES scientists and Grid developers.

For each time period, we have also summarized the corresponding expected achievements of three major ES communities' programmes/initiatives, i.e., GEOSS (GEOSS 2005), Shared Environmental Information System (SEIS),[34] and INSPIRE. While many other EC funded projects or National initiatives exist, GEOSS, SEIS, and INSPIRE represent the framework where these projects build (Charvat et al. 2009).

The main objectives are shown in Tables 1, 2, 3, and 4. The legend of the symbols used in such tables is illustrated on Table 1.

Short-term objectives

One of the major Union of Geodesy objectives to be addressed in the short term is *community building*. Several activities are on going, like DEGREE dissemination, EGEE, European Geosciences Union (EGU), and OGC-OGF joint discussion meetings.

Community building has a large governance component. For a Grid to achieve efficiency of scale, many organizations must participate (by installing Grid nodes, Grid access middleware, tools, datasets, etc.), yet without legislation and directives this will probably not happen. For this reason we urge the key actors at the political level (e.g. WMO, Committee of Earth Observation Satellite, GEO, space agencies, International and Geophysics, European Geoscience Union, etc) to engage in a lasting dialogue aimed at addressing the key issues raised by this roadmap.

In parallel to building support to *education*, development of *Grid training programme*s are needed.

A big effort should also be devoted in *supporting large science community* (as in G-POD). This includes single sign-on authentication to all ES data providers and implementing rules and methods for authorization to access restricted data (support for user roles, directory service, etc.). Another major issue to be addressed in the short term is the definition and implementation of a standard approach to distributed data and metadata.

Finally *identification of ES Grid community resources* is needed: Grid infrastructures; Grid VOs; Certificate Authorities recognized by the Grid; ES Grid centres of expertise; Dissemination and documentation resources.

---

[34] http://ec.europa.eu/environment/seis/index.htm. SEIS is a collaborative initiative to establish an integrated and shared environmental information system. This system would tie in better all existing data gathering and information flows related to EU environmental policies and legislation.

It is worth noting that in parallel to the top-down approach based on the involvement of "political level" actors, a bottom-up approach is also recommended. Demonstrator can create interest and momentum for further adoption of Grid. The adoption of Grid "at the bottom level" can stimulate the interest in "politic level" actors. So additional objectives ranging from short to medium term shall be hook-up with a number of initiatives and develop the needed Grid components; build demonstrators; and search for user feedback, define standards and make available tools compliant to this standard.

It is worth noting that at the time of writing this manuscript, some actions related to these short-time objectives have already been taken and are on going. Special sessions related to the exploitation of Grid in ES have been organized in Conferences and workshops, (e.g., at the 2009 General Assembly of the European Geoscience Union), a special issue of the Earth Science Informatics Journal on Grid computing in Geosciences was published (various authors, 2009) . The EC funded project GENESI-DR is currently demonstrating the federation of large number of heterogeneous ES repositories in a single access point for discovery, access, and processing.

More in general, a great attention has to be paid to ensure the interoperability among ES platforms and other infrastructures. In the very recent time, e-IRG has established a Task Force for Data Management with the objective to identify issues for metadata management and interoperability. Earth Science is well represented in this activity.

Medium-term objectives

The medium-term period is mainly oriented at solving outstanding blocking issues for ES applications deployment on Grid; porting of applications from different ES disciplines throughout a wide community; and facilitating access for everyday ES users to Grid. Main medium term objectives are summarized in Table 3.

Other actions needed are:

- the integration, in Grid environment, of a lot more specific data handling tools
- the development of standard services in Grid for data retrieving, reprocessing and systematic data processing;
- the direct deployment to the Grid storage of the most used very large environmental databases and file

Table 2 Short-term objectives identified in the DEGREE roadmap—Light gray background for the expected achievements. Dark gray background for specific objectives, as identified by user community, technical experts, and ES main initiatives (see Table 1 for a legend of the symbols used)

Short-term objectives	
	from 1 to 3 years
	Expected Achievements: ○ *expand the ES-Grid community in terms of participating organizations, teams and applications* ○ *significantly reduce the gaps towards a basic Grid infrastructure suitable for the needs of ES community* ○ *porting of applications from different ES disciplines throughout a wide community, initially for dissemination and moving gradually to more applications that produce useful results*
	…continued support for cross-fertilization between communities and the continuation of ES Grid promotion and dissemination activities…
	…establishment of Grid as a key component to be fully integrated within the emerging ES Research as well as operational large-scale infrastructures …
	…easy APIs for accessing data… integrated single-sign-on for ES community…
	… facilitate, within 2 years, the development and availability of shared data, metadata, and products commonly required across diverse societal benefit areas …
	… sharing of information and full decentralised access (as of 2009)…
	... metadata available for spatial data; discovery and view services operationally ,the EC establishes and runs a geo-portal at Community level

Table 3 Medium-term objectives identified in the DEGREE roadmap—Light gray background for the expected achievements. Dark gray background for specific objectives, as identified by user community, technical experts, and ES main initiatives (see Table 1 for a legend of the symbols used)

	Medium−term objectives
🕐	*from 2 to 5 years*
	Expected Achievements: o *Solve outstanding blocking issues for ES applications deployment on Grid, including data access and quality of service* o *Porting of applications from different ES disciplines throughout a wide community, publication of new scientific results which could not have been achieved without ES Grid infrastructures.* o *Facilitate access for everyday ES users to Grid.*
👥	*… the increase of applications ported to the Grid to let them benefit immediately for having access to sensors, data, archives, etc…. the facilitation of access to the Grid for everyday users …*
🌍	*…development of a Grid platform aimed at supporting the implementation of INSPIRE and GMES objectives for sharing of environment-related data and information within Europe and the provision of services to policymakers and citizens. …*
⚙	*… track changes tools for data (at VO level)…authentication and authorization to data (from data management point of view)…middleware should support dynamic access policies to data …to assure a Quality of Service, e.g., monitoring of queues… near-real time job execution… reliabilities of nodes…easy installation of Grid…*
GEO	*… facilitate, within 6 years, data management approaches that encompass a broad perspective of the observation data life cycle, from input through processing, archiving, and dissemination, including reprocessing, analysis and visualization of large volumes and diverse types of data …*
SEIS	*… architecture benefiting from the network of services developed under INSPIRE (as of 2010)*
🍍	*… download services operational, coordinates transformation services operational, newly collected and extensively restructured spatial data sets available, schema transformation and "invoke spatial data service" services operational, advanced metadata available for spatial data…*

repositories (e.g., space and weather archives, climate reanalysis, seismic profiles, catalogues and waveforms);

- the integration of all the features needed for *e-collaboration* among ES;
- the increased development of dedicated ES Grid tools base on standards and re-use;
- the consolidation *of results* from many ongoing projects and initiatives to concentrate the efforts in the creation of a common ES platform.

Long-term objectives

The long-term objectives are fundamental for a real exploitation of Grids in ES. Nonetheless, they will probably not be achieved unless medium term objectives are achieved. Main long-term objectives are summarized on Table 4.

The effort required for achieving these objectives is more about end user engagement, than it is about the deployment of any particular technology. Thus, particular emphasis should be given to: ease of adoption, ease of use, hiding complexity from the end-user; understanding and supporting the "usual way of business" for ES end-users (do not force users to learn "arcane" computer knowledge or work in a fundamentally different way then they are used to).

Conclusions

This paper was generated on the basis of the results obtained by the different DEGREE project activities, of discussions among DEGREE project partners, and of the

Table 4 Long-term objectives identified in the DEGREE roadmap—Light background for the expected achievements. Dark background for specific objectives, as identified by user community, technical experts, and ES main initiatives (see Table 1 for a legend of the symbols used)

Long–term objectives	
⏱	*From 4 to 8 years*
	Expected Achievements: o *Establish a standardised ES Grid platform fully integrating important new emerging ES standards* o *This ES Grid platform shall be able to provide ES users with easy and seamless access to distributed and diverse data and to support them in enlarging collaboration, creativity and productivity*
👥	*… an ES Grid platform …*
🌍	*…. make ES data readily available via Grid services, so that ES applications running on the Grid processing resources can be used by experts to perform their own analyses, integrating data from across different application domains … uploading of new results and datasets to increase the pool of available resources accessible via the Grid …*
✳	*… ES specific technical requirement on Grid middleware are fully implemented…*
GEO	*… establish, within 10 years, a system of systems to provide timely data and products for local, national, regional, and international policy makers … capacity building strategy that will have significantly strengthened the capability of all countries, and particularly of developing countries, to:* o *Use Earth observation data and products (e.g., process, integrate, model) following accepted standards;* o *Contribute to, access, and retrieve data from global data systems and networks;* o *Analyze and interpret data to enable development of decision-support tools and to advance understanding in the nine societal benefit areas;* o *Integrate Earth observation data and products with other data and products, for a more complete view and understanding of problems and derived solutions;* o *Improve infrastructure development in areas of poor observational coverage; and,* o *Develop recommended priorities for new or augmented efforts in capacity building.*
SEIS	*…coherent information resulting from more coherent monitoring systems and interoperability rules adopted under INSPIRE (as of 2015)*
🍊	*… newly collected and extensively restructured spatial data sets available*

feedbacks received by consultations with the wide ES community. Almost all collected feedbacks agreed on the importance of generating an ES Grid roadmap and on the main objectives it contains. The experience and feedback presented in this paper come from the European ES community.

Starting from the analysis of the main ES requirements, challenges and objectives and considering also the visions that other ES initiatives are promoting, an ES Grid Roadmap has been outlined, setting out the main objectives, which the ES Community considers its goals.

In this paper, rather than providing detailed specification, our aim has been to describe the approach, highlight key characteristics and provide a general outline for the envisaged platform. Rather than detailing "how to do it" we have considered "what to do", mainly from the point of view of ES partners.

The vision for the envisaged Grid-based ES platform is in line with other EC projects, GEOSS strategic implementation plan and INSPIRE directives.

During the development it is important to always keep in mind factors as: the wide diversity of ES application themes and cultures; the large number of existing methodologies for the analysis of multi-source data, time series, and data assimilation are being considered; the ES need for trans-national collaborations involving universities, institutions,

research centres, international organizations, operational entities, public bodies and private companies.

References

Castro Leon, He J, Chang M, Parviz Peiravi (2008) The business value of virtual service oriented grids. Intel Press, ISBN-10 1934053104, ISBN-13: 978-1934053102

Charvat et al. (2009) INSPIRE, GMES and GEOSS activities, methods and tools towards a single information space in Europe for the environment, Ed. Charvat, Alberts, Horakova, ISBN 978-9934-8105-0-3

Cossu R, Linford J, Babik M, Petitdidier M, Schwichtenberg H, Som de Cerff W (2008) Earth Science white paper on Grids, DEGREE EU deliverable 6.1.2, http://www.eu-degree.eu/DEGREE/internal-section/wp6/DEGREE-D6.1.2_v2.8.pdf

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

Fusco L, Cossu R, Retscher, R (2007) Open grid services for Envisat and Earth observation applications. In: Plaza A (ed) High Performance Computing in Remote Sensing (chapter 13), Taylor and Francis Group, ISBN 978-1584886624

GEOSS (2005) The global Earth observation system of systems (GEOSS) 10-Year Implementation Plan. http://Earthobservations.org/docs/10-Year%20Implementation%20Plan.pdf

Iapaolo M, Godin-Beekmann S, Del Frate F, Casadio S, Petitdidier M, McDermid IS, Leblanc T, Swart D, Hansen G, Stebel K (2007) GOME ozone profiles retrieved by neural network techniques: a global validation with lidar measurements. J Quant Spectrosc Radiat Transfer 107:105–119

Linford J et al. (2008) Grid, E-collaboration and Soa portals in Earth science: state-of-the art survey. https://www.eu-degree.eu/DEGREE/internal-section/wp4/DEGREE-WP4D4.1%20ES%20Portals%20Survey%20v1_0.pdf

Mazzetti P, Nativi S, Angelini V, Verlato M, Fiorucci P (2009) A grid platform for the european civil protection e-infrastructure: the forest fires use scenario. Earth Sci Inf 2(1):53–62

Raoult B, Aubert G, Gutierrez M, Arciniegas-Lopez C, Correa R (2009) Virtual Organization in the SIMDAT Meteorological Activity: a decentralised access control mechanism for distributed data. Earth Sci Inf 2(1-2):63–74

Renard P, Badoux V, Petitdidier M, Cossu R (2009) Grid computing for Earth Sciences. Eos 90(14):117–119

Schwichtenberg H et al. (2007) Survey of existing data technologies in Earth Science and data usage policies. Available at DEGREE Web site. https://www.eu-degree.eu/DEGREE/internal-section/wp2/DEGREE-D2.1_v1.0.pdf

Schwichtenberg H et al. (2008) Report on grid based data management tools for ES applications https://www.eu-degree.eu/DEGREE/internal-section/wp2/DEGREED2.3-final.pdf

SIMDAT (2006) D12.2.1 Report on technologies assessed for grid infrastructure. Available at www.simdat.org

SIMDAT (2008) Virtual meteorolofical centre (VMC) documentation, user, administrator and reference manual. Version 0.9, available at www.simdat.org

Som de Cerff W, Petitdidier M, Boonne C, Huchly L, Lonjaret M, Schwichtenberg H, van de Vegte J, van der Wel F, Vilotte JP (2007) Report on ES family of applications and their requirements, DEGREE EU deliverable 1.1, http://www.eu-degree.eu/DEGREE/internal-section/wp1/DEGREE-WP1-D1-1.3.pdf

Som de Cerff WJ, Petitdidier M, Gemünd A, Gotab B, Horstink L (2008)Progress report on ES family of applications and their Grid requirements, DEGREE EU deliverable 1.2, http://www.eu-degree.eu/DEGREE/internal-section/wp1/DEGREE-D1.2u.pdf

Som de Cerff W, Petitdidier M, Gemünd A, Horstink L, Schwichtenberg H (2009) Earth science test suites to evaluate grid tools and middleware-examples for grid data access tools. Earth Sci Inf 2(1-2):117–131

Various Authors (2006) The emergence of grid and service-oriented IT: an industry vision for business success, different Authors, Tabor Communications, ISBN-10: 1427600252 , ISBN-13: 978-1427600257

Various authors (2009) Special issue on grid computing in geosciences. Earth Sci Inf 2(162):139

Viet Tran et al. (2008) Updated state of the art and gap analysis on application management and control technologies. Available at DEGREE Web site https://www.eu-degree.eu/DEGREE/internal-section/wp3/DEGREE-D3.3-final.pdf

Wang L, Jie W, Chen J (2009) Grid computing: infrastructure, service, and applications, CRC Press, ISBN 9781420067668