

The Advance Technology Fusion for EO Application

L.Fusco, R. Cossu, J. van Bemmelen, V.Guidetti, E.Li Santi, and A.Rossi*

Directorate of Earth Observation Programmes, European Space Agency, via Galileo Galilei 2, Frascati, Italy, luigi.fusco@esa.int

ABSTRACT

Due to its intensive data processing requirements and highly distributed organisation, the Earth Science applications community is uniquely positioned for the uptake and exploitation of Grid technologies. In this paper we describe a number of initiatives which are based on an Earth Science e-collaboration platform that makes use of Grid and Service Oriented Architecture technologies. The aim is the enabling and the simplification of life of scientists in discovery, access, combining and integration of all on-line available ever-generated Earth-related data from wherever they are and by what ever sensor they were generated.

Keywords: Earth Science, Grid technology, Service Oriented Architecture, data access

1. INTRODUCTION

The ESA Earth Observation Programmes Directorate at ESRIN has focused on the development of a dedicated Earth Science (ES) Grid infrastructure, named Earth Observation Grid Processing on-Demand (G-POD) (Fusco et al. 2006). This environment provides transparent, fast, and easy access to the large ESA Earth Observation (EO) archive together with ESA available Grid computing and dynamic storage resources. ESA has recently offered scientists the possibility to perform bulk processing and/or validation of their *own* algorithms exploiting this environment. Obtained results, related to a wide range of EO applications, have demonstrated how the ES community can benefit from Grid and emerging SOA technologies. Based on this experience, a number of initiatives have started to develop an ES e-collaboration platform that makes use of Grid and SOA technologies.

Among the recent projects we mention:

- The EU-IST DEGREE Project (<http://www.eu-degree.eu/>), which aims to assist and accelerate the uptake and exploitation of Grid and emerging SOA technologies in ES. DEGREE has among its objectives the identification of new research activities in this field that will be documented in a White Paper.
- The EU D4Science project (<http://www.d4science.eu/>). The main objective of this project is to deploy and use in real-world scenario's the e-Infrastructures built by the European Grid for E-science (EGEE), currently the largest international grid infrastructure, with a current capacity of over 20.000 CPUs, and DILIGENT (<http://www.diligentproject.org/>) to advance the science and decision making of scientific communities related to Environmental Monitoring and Fishery Resource Management.
- The EU GENESI-DR project (<http://www.genesi-dr.eu/>) which will provide the base for a world-wide e-Infrastructure for Earth Science integrating new scientific and technological derived paradigms in operational infrastructures in responds to the latest Earth Science requirements considering emerging technologies to improve performance in delivery, recent progress in geo-spatial data sharing across distributed servers and emerging Grid-based e-infrastructures. It has the challenge of establishing open Earth Science Digital Repositories access for science users, that is, it will establish seamless access to and share all online available data, information, products and knowledge originating from space, airborne and in-situ sensors. GENESI-DR shall operate, validate and optimise the integrated access and build on existing digital data repositories from main Earth Science players in Europe (including ESA, the German, Italian and French Space Agencies, Infoterra, KSAT, NILU and CNR) to

demonstrate how Europe can best respond to the emerging global needs relating to the state of the Earth, a demand that is unsatisfied so far.

2. DEGREE – Dissemination and Exploitation of Grid for the Earth Science community

The Dissemination and Exploitation of GRids in Earth science (DEGREE) project (<http://www.eu-degree.eu>) was a Specific Support Action, funded within the Grid call of EC FP6. The project aimed to promote the Grid culture within the different areas of ES and to widen the use of Grid infrastructure as platform for e-collaboration in the science and industrial sectors and for selected thematic areas which may immediately benefit from it. The DEGREE project was also tackling certain aspects presently considered as barriers to the widespread uptake of the technology, such as perceived complexity of the middleware and insufficient support for certain required functionality. Collective Grid expertise gathered across various ES application domains was exchanged and shared in order to improve and standardise on application specific services. The use of worldwide Grid infrastructures for cooperation in the extended ES international community has been also promoted. In particular the following objectives were achieved:

- Disseminate, promote uptake of Grid in a wider ES community and integrate newcomers.
- Reduce the gap between ES users and Grid technology.
- Explain and convince ES users of Grid benefits and capability to tackle new and complex problems.

DEGREE had among its objectives the generation of a roadmap setting out the key steps needed for the ES 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 as data archiving and preservation, data access, data harmonization. A first ES Grid Roadmap document was generated as basis for discussion among a wider ES community, in order to derive a final roadmap paper. The roadmap will facilitate and speed up the adoption of Grid and generally the e-infrastructure for ES, where adoption of the emerging technologies has so far been slower than expected, despite the large potential benefits.

The major objectives identified in the roadmap are set out as a journey progressing through a list of points to be achieved in order to arrive at the establishment of a well identified ES Grid Platform. The main goals could be summarized such as:

- Support community building, support to education, and insertion of Grid training programmes,
- maximise approach to support large science community (as in GPOD) – single sign-on from all ES data providers,
- port to the Grid of new applications which may benefit immediately for having access to sensors, data, archives, etc.
- integrate in Grid environment a lot more specific data handling tools,
- facilitate access to the Grid for everyday users,
- solve outstanding blocking issues for ES applications deployment on Grid,
- increase development of dedicated ES Grid tools, based on re-use of newly available generic Grid frameworks, tools and services, based on emerging (WSRF) standards (e.g. OMII)
- consider ES Grid as Collaboration platform for science and service industry,
- plan the enhancement of ES Grid Platform with additional next-generation dedicated application level tools and services.

2.1 The roadmap

An ES Grid development roadmap was missing and its definition was of the main objectives of DEGREE. It is worth reminding that Grid is intended not as a goal but as a means for reaching the ES priorities. DEGREE partners believe that “Grid is a commodity ES users can benefit from to pursue their challenging objectives”.

At the time of the project, the situation was the following:

- First successful application deployment and demonstration of benefits that Grid infrastructure can bring to ES community.
- Middleware gaps identified, ES requirements gathered and analysed and communicated to the Grid developers.
- Need for increased awareness and participation of ES community to Grid activities in order to proceed (starting point of DEGREE project)

This was the starting point for the ES Grid roadmap, which is in line the expected achievements of the major ES communities programmes/initiatives, like GEOSS (<http://www.earthobservations.org/>), SEIS (<http://ec.europa.eu/environment/seis/index.htm>), and INSPIRE (<http://www.ec-gis.org/inspire>), is publicly available at (DEGREE Project, 2008). In the following we summarize the main points of the roadmap.

One of the major objectives to be addressed in the short term is *community building*. In parallel to building support to *education*, development of *Grid training programmes* 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.) (12-24months). Another major issue to be addressed in the short term is the definition and implementation of a standard approach to distributed data and metadata.

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. Other actions needed are: the integration, in Grid environment, of a lot more *specific data handling tools*.

The long-term objectives are fundamental for a real exploitation of grids in ES. 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 than they are used to).

3. D4SCIENCE – Distributed colLaboratories Infrastructure on Grid ENabled Technology for Science

The main objective of D4SCIENCE is to deploy and operate the e-Infrastructures built by the EGEE (<http://public.eu-egge.org/>) and DILIGENT (<http://www.diligentproject.org/>) projects to address the requirement of accessing distributed facilities for data management, processing and archiving as expressed among the others by the Earth Science User Community, represented in D4SCIENCE by ESA. Such an e-Infrastructures offers mechanisms that concurrently exploit networks, grids, and data in a seamless fashion and enable research scientific communities to operate regardless of the location of their research facilities.

In particular the project offers the means for creating Virtual Research Environments based on computation, storage and service resources shared at European level and then to customize these Environments by integrating domain-specific applications as requested by the international organizations leading the user communities involved in the project, i.e. FAO and ESA.

Beyond the benefits envisaged by the scientific communities with an adoption of the D4Science e-infrastructure the project is acting as a major catalyst for the cooperation and cross-fertilization among different user communities necessary for improving the way they address big challenges of Science and Society.

4. GENESI-DR – Providing transparent access to all Earth Science data

Ground European Network for Earth Science Interoperations - Digital Repositories (GENESI-DR) (<http://www.genesi-dr.eu/>), an ESA-led, European Commission (EC)-funded two-year project, kicked-off early 2008 and is taking the lead in providing reliable, easy, long-term access to Earth Science data via the Internet. Petabytes of data about our planet are available but distributed at different locations. Currently, information about the state of the Earth, relevant services, analysis results, applications and tools are accessible in a very scattered and uncoordinated way, often through individual initiatives from Earth

Observation mission operators, scientific institutes dealing with ground measurements, service companies, data catalogues, etc. Data access is a major logistic problem. The EC has funded GENESI-DR as a flagship project in Europe to help meet the challenge of facilitating life of scientists from different Earth Science disciplines located across Europe in discovery, access and use (combining, integrating, processing, ...) of historical and fresh Earth-related data from space, airborne and in-situ sensors archived in large distributed repositories.



Figure 1 – Dedicated ES infrastructure as envisaged by GENESI-DR project

GENESI-DR is a response to the need for science users to be provided with data and tools to access, combine, and integrate the Earth-related data for performing their analyses.

This need has led to the identification of the basic GENESI-DR infrastructure requirements:

- Capability, for Earth Science users, to discover data from different European Earth Science Digital Repositories through the same interface in a transparent and homogeneous way;
- Easiness and speed of access to large volumes of coherently maintained distributed data in an effective and timely way;
- Capability, for DR owners, to easily make available their data to a significantly increased audience with no need to duplicate them in a different storage system.

The first requirement is reflected in the GENESI-DR architecture on the Central Discovery Service, which allows users and applications to query information about data collections and products existing in heterogeneous catalogues, at federated DR sites.

This service can be accessed by users via web interface, the GENESI-DR Web Portal, or by external applications via open standardized interfaces exposed by the system.

More in detail, the Central Discovery Service identifies the DRs providing products complying with the user search criteria and returns the corresponding access points to the requester. This latter can refine its search towards the DRs, so that products complying with refined search criteria are identified and the corresponding metadata are returned to the client. These include the product access URL allowing product retrieval.

On this purpose, to meet the second requirement, flexibility and performance are taken into consideration by making use of different and efficient data transfer technologies such as HTTPS, GridFTP and BitTorrent.

To cope with the third requirement, the Architecture of GENESI-DR provides the DR owners with a mechanism (Catalogue Generator) to produce a metadata catalogue by simply harvesting metadata from their storage systems.

The Central Discovery Service communicates with the different DR catalogues through a web service gateway.

The GENESI-DR e-Infrastructure will be validated against specific applications in the land, atmosphere and marine domain, including:

- Near real time orthorectification for agricultural crops monitoring and Urban area mapping in support of emergency response, both led by the Joint Research Centre;

- GlobModel (<http://www.globmodel.info>), for real time data assimilation, addressing major environmental and health issues in Europe, with a particular focus on air quality, led by RUR, the University of Reading;
- SeaDataNet (<http://www.seadatanet.org/>), led by ENEA, with the aim to aid environmental assessments and to forecast the physical state of the oceans in near real time.

Their needs in terms of data types have been analysed when planning the initial deployment of GENESI-DR and, along with technical and policy issues, have driven the selection of the first set of DRs to be integrated in the infrastructure.

ESA, ENEA, NILU and KSAT belong to this first set and the completion of their integration is planned before the end of the year.

The project's goal for the next future is to deal with new requirements gathered from application and DR owners, as well as to increase the number of integrated DRs (adding Infoterra, CNES, DLR, ASI, JRC) and the number of validation applications as well as to establish links to other EU projects to validate the system.

Architectural designing effort is already being spent in order to face the expressed need for GENESI-DR users to be provided with processing services to be used on data of interest.

In fact, GENESI-DR will soon provide services to enable expert users exploiting computational and network resources in order to produce the final desired product. This means either passing input data to a processing service (an OGC Web Processing Service, for instance) available at some site or running a user application / algorithm on Grid resources on specified data sets.

GENESI-DR is also analysing common approaches to preserve the historical archives and the ability to access the derived user information as both software and hardware transformations occur. Ensuring access to Earth Science data for future generations is of utmost importance because it allows for the continuity of knowledge generation improvement. For instance, scientists accessing in 50 years from now the climate change data of the then past 50 years will be able to better understand and detect trends in global warming and apply this knowledge to ongoing natural phenomena. GENESI-DR works towards harmonising operations and applying approved standards, policies and interfaces at key Earth Science data repositories. To help with this undertaking, GENESI-DR is establishing links with the relevant organisations and programmes such as space agencies, institutional environmental programmes, international Earth Science programmes and standardisation bodies.

From an industry point of view, the project's infrastructure, which is based on state-of-the-art technologies such as Grid, will pave the way for them introducing new services in the Earth Science domain from a scientific and commercial viewpoint.

5. CONCLUSIONS

Data play a central role in different aspects of Earth Science. Petabytes of data about our planet are available but distributed at different locations. For these reasons, while conducting their research, Earth Science scientists are often hindered by difficulties locating and accessing the right data, products, and other information needed to turn data into knowledge. Grid and emerging ICT technologies are possible solutions for building up a data infrastructure, where data management, e-collaboration, services, and community building are possible.

ESA and EC are paying a great attention to data infrastructures. In this paper we have described EC funded projects DEGREE, D4SCIENCE and GENESI-DR. In all the cases, the vision is a data infrastructure able to provide reliable, easy, long-term access to Earth Science data via the Internet, so allowing ES scientists and users to easily and quickly derive objective information and share knowledge based on all environmentally sensitive domains. New ICT technology such as Grid seems very promising to face the ES challenges. These technologies should be a commodity for ES community that should pose itself as Grid users more than Grid developers. This means that the adoption of Grid and ICT technology in data infrastructure should be as transparent as possible to ES scientists and users.

6. REFERENCES

L.Fusco, R.Cossu, C.Retscher, "Open Grid services for Envisat and Earth observation applications" in High Performance Computing in Remote Sensing, Ed: Antonio Plaza, Taylor and Francis Group, Chapter 13.

DEGREE Project – Earth Science white paper on Grids. Retrieved October 25th, 2008 from the World Wide Web: https://www.eu-degree.eu/DEGREE/internal-section/wp6/DEGREE-D6.1.2_v2.8.pdf/view

DEGREE Project – Project Homepage: Accessed October 25th, 2008 from the World Wide Web: <https://www.eu-degree.eu/>

EGEE: Enabling Grids for E-science in Europe. project Homepage accessed October 25th, 2008 from the World Wide Web: <http://public.eu-egee.org>

D4Science: DIstributed colLaboratories Infrastructure on Grid ENabled Technology for Science, project Homepage accessed October 25th, 2008: <http://www.d4science.eu>

DILIGENT: A DIgital Library Infrastructure on Grid ENabled Technology. Project Homepage accessed October 25th, 2008: <http://www.diligentproject.org>

GENESI-DR: Ground European Network for Earth Science Interoperations - Digital Repositories, project Homepage accessed October 25th, 2008 from the World Wide Web: <http://www.genesi-dr.eu/>