Mem. S.A.It. Vol. 80, 461

Memorie della

Past and future of ESA Earth Observation Grid

L. Fusco and R. Cossu

Directorate of Earth Observation Programmes, European Space Agency, via Galileo Galilei, Frascati, Italy, e-mail: [luigi.fusco;roberto.cossu]@esa.int

Abstract. Due to its intensive data processing and highly distributed organization, the multidisciplinary Earth Science (ES) applications community is uniquely positioned for the uptake and exploitation of Grid technologies. In this paper, we describe a number of initiatives that the European Space Agency is carrying focusing on a ES e-collaboration platform that makes use of Grid and SOA technologies. Starting from the experience gained so far with ESA Grid Processing on Demand, and the results of the EC funded DEGREE project, we will discuss the vision of a dedicated ES platform. The aim is enabling scientists to locate access, combine and integrate historical and fresh Earth-related data from space, airborne and in-situ sensors archived in large distributed repositories. The big challenge is allowing Earth Science communities to easily and quickly derive objective information and share knowledge based on all environmentally sensitive domains. GENESI-DR project is already moving the first steps in developing this platform.

Key words. Grid, Earth Observation, ESA

1. Introduction

Planet Earth is increasingly in danger due to the strong presence and impact of the human life. In response to this, the United Nations has supported environmental conventions attempt to define internationally agreed protocols (e.g., Kyoto, Biodiversity and Montreal) to limit and monitor status of our global environment. The World Summit on Sustainable Development, Johannesburg 2002 (WSSD), highlighted the urgent need for coordinated observations relating to the state of the Earth. It established the ad hoc intergovernmental Group on Earth Observations (GEO) Earth Observations website (2008), co-chaired by the European Commission, Japan, South Africa, and the United States, and tasked it with the development of an initial ten year Implementation Plan by February 2005, also known as Global Earth Observation System of Systems (GEOSS) ten year Implementation Plan GEOSS (2008). This plan establishes the intent, operating principles, and institutional frame, which was the base at the third Earth Observation Summit, in Brussels, for the establishment of the intergovernmental GEO. The role of the European Earth Science community in the development and implementation of this global infrastructure is of paramount importance. Preceding the WSSD-2002, the European Commission (EC) and the European Space Agency (ESA) initiated, jointly supported by all ESA and EU member countries, a parallel international initiative called Global Monitoring for Environment and Security (GMES) Kopernikus (2008), which is now considered to be the European con-

Send offprint requests to: L. Fusco

tribution to GEOSS. The implementation and the systematic monitoring of international Environmental conventions need data, tools and world-wide infrastructures to gather and share the data. A common dedicated infrastructure will permit the Earth Science communities to derive objective information and to share knowledge in all environmental 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). Furthermore, each specific Earth Science (ES) domains community has existing methods, approaches and working practices for gathering, storing and exchanging data and information. These are likely to impose a considerable constraint on the impact and increased effectiveness generated by a shared e-Infrastructure approach. The challenge in front of us today, is to offer a framework that allows scientists from different Earth Science disciplines to have access, to combine and to integrate all historical and present Earth-related data from space, airborne and in situ sensors available from all digital repositories dispersed all over Europe together. The ESA Science and Application Department of Earth Observation Programmes Directorate at ESRIN is involved in several activities that make use of emerging technologies to address these challenges. In this paper we give an overview of such activities focusing on Grid technology. We start giving an overview of the ES challenges, related international initiatives, and key requirements scientific requirements. We provide a discussion on the use of emerging Information and Communication Technologies (ICTs) for EO data access and utilization. We describe the EO Grid Processing on Demand (G-POD) environment G-POD (2008) used in ESA, with examples of applications that exploit Grid technology either for massive data processing or fast access to large volume of data. The paper moves then to the vision of, and the first steps towards a dedicated ES Grid Platform. To this end, two ESA participated projects are presented, namely Dissemination and Exploitation of Grid for Earth science (DEGREE) DEGREE (2008) and Ground European Network for Earth Science Interoperations - Digital Repositories (GENESI-DR) GENESI-DR (2008). While DEGREE gathered and analyzed the main requirements of ES as respect to a Grid based platform, GENESI-DR is applying these findings to develop a prototype of ES e-collaboration platform.

2. Earth Science challenges and mission

The second half of the twentieth century saw full emergence of the global-scale concept in the atmosphere, the solid Earth, the hydrosphere, the cryosphere, the biosphere and the anthroposphere. All of these components are interlinked by a network of forcing and feedback mechanisms that affect the other components. Our perception of the Earth as a system finds its scientific expression in a modeling spectrum running from the conceptual approach to the highly computational one. These models encapsulate our understanding of the Earths processes, their dynamic behaviour, their relations and their feedbacks. Quantitative models, by their very nature, consist of equations whose solution requires input data. Evaluating the models also requires data. Hence our total knowledge about the system is contained in our models, our measurements and how we put them together have to be managed cleverly. Improved understanding of how to represent the dynamics of the Earth processes, together with the explosion in computing power, has allowed the construction of ever more powerful computer codes capable of solving the coupled equations making up an Earth System model with a spatial resolution of the order of ten kilometers. An equally influential development has been the explosion in the size of databases, with data from a plethora of different sources, and the interconnectivity of computer systems and databases. Putting these together has led to enormous advances in the methods of model-data fusion and data assimilation. The data are the basic input to any ES application, directly or indirectly as they validate the results. The data come from sensors on different platforms, satellite, plane, boat, balloon, buoy or mast, or located at ground on the land. The data are distributed in different locations according to their topic and volume. A data policy is associated to each set of data and the data policy may change as a function of time. It is explained by the fact that some data are very sensitive for country economy or security (GMES programme), or for discovery. Petabytes of already acquired data are presently underexploited (less than 10%), because to get the results in a reasonable time not enough computing resources, tools and algorithms are available. However, even if they were to be made available, an efficient infrastructure to handle and treat very large data sets is still missing. In order to facilitate the access to data, their processing and visualisation, the ES community has developed Portals i.e. Web interface that integrates mainly several standard and specific-ES Web services. In some cases the portal provides access to service-based Grid for high performance processing. In conclusion, more resources and services are required to face the ES challenges. In particular one major point is the data management. New technologies, such as Grid computing, could be a very promising solution.

3. International initiatives

Major challenges of ES are the implementation and the systematic monitoring of international Environmental conventions. Different international initiatives and programs are concerned. Producing and managing better information about the environment has become a top priority for nations around the world. In July 2003, the Earth Observation Summit brought together 33 nations plus the European Commission and many International Organizations to adopt a declaration that signified a political commitment toward the development of a comprehensive, coordinated and sustained Earth Observation System to collect and disseminate improved data, information, and models to stakeholders and decision makers Earth Observations website (2008). An ad-hoc group of senior political officials from all participating Countries and Organizations, named the Group on Earth Observations (GEO), was formed to undertake this global effort. GEO was charged to develop a ”Framework Document” plus a more comprehensive report to describe how the collective effort could be organized to continuously monitor the state of our environment, increase understanding of dynamic Earth processes, and enhance forecasts on our environmental conditions. Furthermore, it was to address potential societal benefits if timely, high quality, and long-term data and models were available to aid decision-makers at every level, from intergovernmental organizations to local government to individuals. The result is a ten-year implementation plan for a Global Earth Observation System of Systems (GEOSS) GEOSS (2008).

GEOSS is envisioned as a large national and international cooperative effort to bring together existing and new hardware and software, making it all compatible in order to supply data and information at no cost. Outcomes and benefits of a global informational system will include:

– disaster reduction;
– integrated water resource management;
– ocean and marine resource monitoring and management;
– weather and air quality monitoring, forecasting and advisories;
– biodiversity conservation;
– sustainable land use and management;
– public understanding of environmental factors affecting human health and well being;
– better development of energy resources;
– adaptation to climate variability and change.

Kopernikus, formerly known as GMES (Global Monitoring for Environment and Security), is a European initiative for the implementation of information services dealing with environment and security Kopernikus (2008). It is based on observation data received from Earth Observation satellites and ground based information. These data will be coordinated, analysed and prepared for end-users. Kopernikus comprises a set of services for European citizens helping to improve their quality of life regarding environment and security. It is the European participation in the worldwide monitoring and management of our planet Earth and the European contribution to the Group on Earth Observation (GEO). Another European initiative is the Shared Environmental Information System (SEIS) SEIS (2008). This is a collaborative initiative of the European Commission and the European Environment Agency (EEA) to establish together with the Member States an integrated and shared EU-wide environmental information system. This system would tie in better all existing data gathering and information flows related to EU environmental policies and legislation. It will be based on technologies such as the internet and satellite systems and thus make environmental information more readily available and easier to understand to policy makers and the public. According to the SEIS concept, environmentally-related data and information will be stored in electronic databases throughout the European Union. These databases would be interconnected virtually and be compatible with each other. The proposed SEIS is a decentralised but integrated web-enabled information system based on a network of public information providers sharing environmental data and information. It will be built upon existing e-infrastructure, systems and services in Member States and EU institutions. Different international initiatives focus on Crisis mapping services. In this case, the main challenge is to provide timely access to crisis data which is crucial for supporting time-sensitive applications. The International Charter Space and Major Disasters SMD (2008) aims at providing a unified system of space data acquisition and delivery to those affected by natural or man-made disasters through Authorized Users. Each member agency has committed resources to support the provisions of the Charter and thus is helping to mitigate the effects of disasters on human life and property. Also RESPOND RESPOND (2008), which began during 2004, as one of the second trances of ESA GMES Service Element (GSE) GSE (2008) projects, is an alliance of European and International organizations working with the humanitarian community to improve access to maps, satellite imagery and geographic information.

4. ES Key Scientific Requirements

Major challenges of ES are the implementation and the systematic monitoring of international Environmental conventions. To this end, data, tools and worldwide infrastructures to gather and share the data are needed. A common dedicated infrastructure would permit the ES communities to derive objective information and to share knowledge in all environmental 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. 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. For example, if we refer to remote sensing missions, only a limited community, with specific knowledge of what to search for, is today in a position to collect, compile and thus exploit the necessary EO information. Furthermore, the different ES community domains 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 infrastructure. However, these same differences impose considerable constraints which impact the increased effectiveness generated by the shared e-Infrastructure approach. ES requirements for accessing historical archives have greatly increased over the last ten years and the trend is likely to further accelerate in the future, mainly for the long term science and the long term environmental monitoring (with related services). As the time-span of EO data archives extends from a few years to decades, their value as a scientific time-series increases considerably especially for global change. In the next decade, the wealth of information currently locked inside the global data archives must be fully exploited and re-analysed on a global scale. Almost all aspect of ES can benefit from accessing long-time series of data Data play a central role in ES and different aspects need to be considered: data cataloguing, data discovery, digital right management for data, data access, data interoperability, metadata schemas, metadata exchange, and ontologies. In addition to data access, the requirement for intensive calculation due to large number of files (e.g., Monte Carlo approach, or re-processing of long series of data) also exists. The requirements for re-processing have been strongly increased over the last ten years. The science community expects rapid adaptation to new requirements, 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 equally covered in modern ground segments and operations concepts. Similar needs and concerns are raised by international forums and organizations as key recommendations like the 10-year GEOSS implementation plan and the 2006 Global Climate Observing System (GCOS) Systematic Observation Requirements for Satellite-Based products for Climate GCOS (2008), 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. Examples of application areas benefiting from ES long-term data archiving exploitation are wide ranging, include EC Policies with long-term perspectives; European and global environment monitoring and forecasting (of forests, land & soil, urban development, air quality, ecosystems and their management for protection of terrestrial, coastal and marine resources); management of energy resources (solar, etc.); development and humanitarian aid and 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, it is necessary to address the following priorities:

- a base for establishing a world-wide e-infrastructure for ES with European leadership;
- 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 for provisional model calibration); harmonisation at key ES data repositories limiting fragmentation of solutions;
- demonstration of solutions effectiveness and creation of the frame for approaching long term preservation of all type of ES data;
- validation of the effective capabilities required to access distributed repositories for new communities, including education, and assessment of benefits and impacts;
- integration of new scientific and technological derived paradigms in operational infrastructures in response to the latest ES requirements.

5. Emerging ICT technologies for EO data access and utilization

While conducting their research, Earth scientists are often hindered by the unnecessary complexity in locating and accessing the right EO data, products, and other information needed to turn data into results and knowledge, e.g. interpretation of the available data. We can state that the process of identifying and accessing data is typically the most expensive in time and effort. Of the different causes of this, those most frequently reencountered relate to:

- The physical discontinuity of data. Data are often dispersed over different data centres and local archives distributed all over Europe and abroad. Inherent to this, the different policies applied (e.g. access, costs), the variety of interoperability, confidentiality, and search protocols as well as the diversity of data storage formats. To access a multitude of data storage systems, users need to know how and where to find them and need a good technical/system background to interface with individual systems. Furthermore, often only the metadata catalogues can be accessed online, while the data themselves have to be retrieved offline.
- The diversity of (meta) data formats. New data formats are being introduced daily, not only due to the individual needs of a multitude of data centres, but also due to advances in science and instrumentation (satellites and sensors) creating entirely new types of data for research.
- The large volume of data. The total quantity of information produced, exchanged and requested is enormous and is expected to grow exponentially during the next decades, even faster than it did before. This is partly the result of the revolution in computational capacity and connectivity and advances in hardware and software, which combined together, are expanding the quality and quantity of research data and are providing scientists with a much greater capacity for data gathering, analysis and dissemination CSPR (2004). For example, the ESA Envisat satellite launched in early 2002, with ten sensors on-board, increases the total quantity of data available each year by some 500 Terabytes, while the ESA ERS satellites produced roughly five to ten times less data per year. Moreover, large volume data access is a continuous challenge for the Earth Science community. The validation of Earth remote sensing satellite instrument data and the development of algorithms for performing the necessary calibration and geophysical parameters extraction often require a large amount of processing resources. Often it also requires highly interactive access to large amounts of data to improve the statistical significance of the process. The same is true when users need to perform data mining or fusion for specific applications.
- The unavailability of historic data. Scientists do not only work with fresh data, they also use historic data, e.g. global change research over long time periods. Here, different problems can be distinguished. First, it is evident that often no metadata are defined, or no common metadata standards are being used, and auxiliary knowledge needed by scientists to understand and use the data is missing, e.g. associated support information in science and technical reports. Although the problem exists also for fresh data, it is exacerbated when using historic data. Metadata will be at the heart of every effort to preserve digital data in the next few decades. It will be used to create maintenance and migration programs and will provide information on collections for the purpose of orienting long-term preservation strategies and systems P. Gauthier et al (2003). Second, there are insufficient preservation policies in place for accessing historical data. After longer periods of time new technologies may have been introduced, hardware and software been upgraded, formats may have changed, and systems replaced. For example, it is almost impossible today to read files stored on 8-inch floppy disks which were popular just 25 years ago. Vast amounts of digital information from just 25 years ago are lost for all practical purposes H. Besser (1999).
- The many different actors involved. Science is becoming increasingly international and interdisciplinary, resulting in an increased total number of different actors involved (not only human). For example, ESA currently serves directly many thousand of users in the Earth Observation domain, many of whom need to exchange data, information and knowledge.

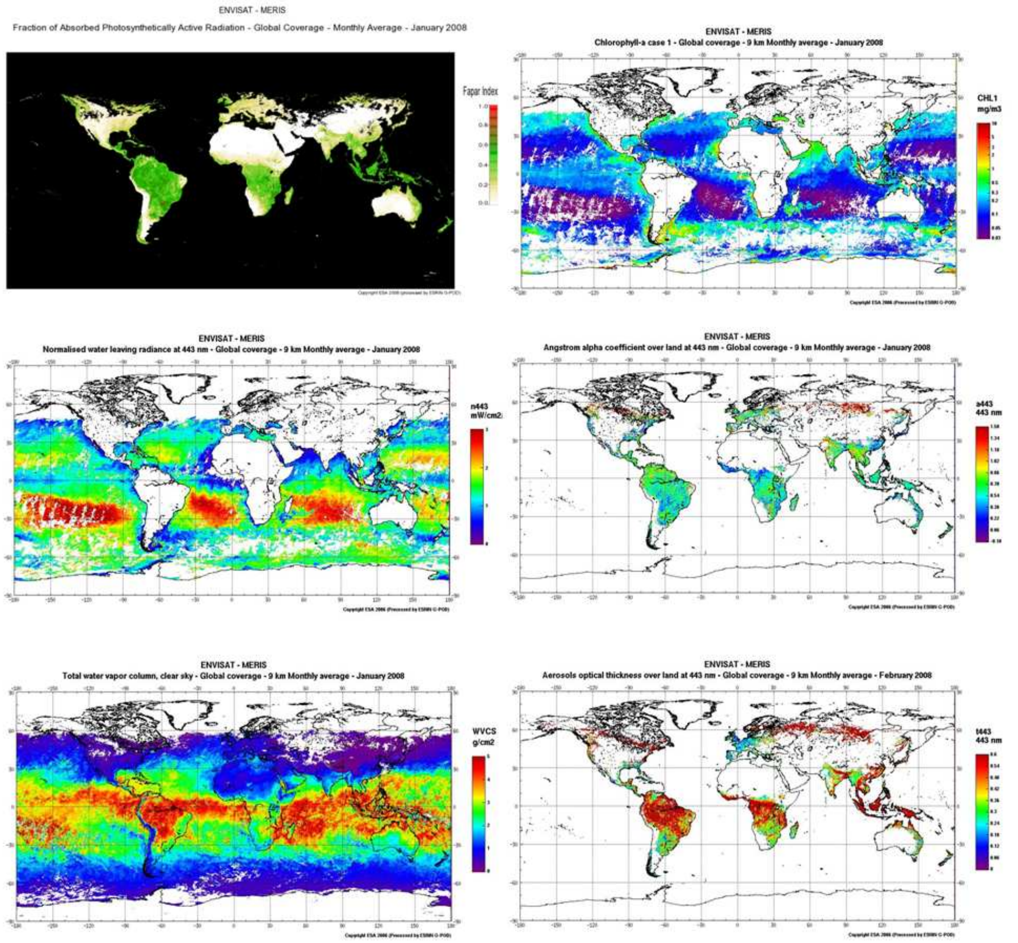

Fig. 1. Examples of MERIS Level-3 products generated in near real time in G-POD.

The International Council for Science for example, deals with data access issues on a global scale CSPR (2004). In Europe, different initiatives are supported by the European Commission (EC), e.g. as part of their specific action on Research Infrastructures, which aims to promote the development of a fabric of research infrastructures of highest quality and performance, and their optimum use on a European scale to ensure that researchers have access to data, tools, and models they need. The ESA Science and Application Department of Earth Observation Programmes Directorate at ESRIN is involved in different initiatives focusing, in particular, on the use of emerging technologies for data access, exploitation, user information services and long-term preservation. For example, Fusco & van Bemmelen (2004) provides an overview of the use of Grid, Web services, and Digital Library technology for long-term EO data preservation. Note that the same ICT technologies are used by other communities for accessing data in general. In the EO community all emerging technologies can play a major role: infrastructures based on high-speed networks could support data access and drastically speed-up the transfer of the enormous quantities of data; Grid middleware could support the management of distributed heterogeneous resources including storage, processing power and communication, offering the possibility to significantly improve data access and processing times; digital libraries could help users locating data via advanced data mining techniques and user profiling. A shared distributed infrastructure integrating data dissemination with generic processing facilities shall be considered a very valuable and cost-effective approach to support Earth Science data access and utilization. Among other specific technologies which have had an important role in the ES community, Web services in particular have played a key role for a long time. Web services technologies have emerged as a de facto standard for integrating disparate applications and systems using open standards. One example of a very specialized ES Web service is the Web mapping implementation specification proposed by the OpenGIS Consortium OpenGIS (2008). Thanks to Web services, Internet has become a platform for delivering not only data but also, most importantly, services. After a Web service is deployed on a Web server and made discoverable in an online registry of services, other applications can discover and invoke the deployed service to build larger, more comprehensive services, which in turn deliver an application and a solution to a user. Web based technologies provide also an efficient approach for distributing scientific data, extending the distribution of scientific data from a traditional centralized approach to a more distributed model. Some Web services address catalogue services to help users to locate datasets they need or at least narrow the number of datasets of interest from a large collection. The catalogue contains metadata records describing the datasets. Web services provide the fundamental mechanism for discovery and client-server interaction and have become a widely accepted, standardized infrastructure on which to build simple interactions. On the other hand, Grids were originally motivated by the need to manage groups of machines for scientific computation. For these reasons, Web Services and Grids are somehow complementary and their combination results in Grid services (e.g. Open Grid Services Architecture). Note that there, in the very recent past, there is a convergence approach in the definition of the Web Processing Service, by the Open GIS Consortium and by the Global GRID Forum OGF (2008).

6. Grid technology for data access and processing: ESA Grid-Processing on Demand

Following the participation to DATAGRID, the first large European Commission funded Grid project EDG (2004), the ESA Science and Application Department of Earth Observation Programmes Directorate at ESRIN has focused on the development of a dedicated Earth Science Grid infrastructure, under the name Earth Observation Grid Processing on-Demand, G-POD G-POD (2008), Fusco et al (2003). 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 transparent, fast, and easy access to data, computing resources, and results. Using a dedicated Web interface, each application has access to a catalogue like the ESA Multi-mission User Interface System (MUIS) Landgraf&Fusco (1998) Landgraf&Fusco (1998) and storage elements. It furthermore communicates with the underlying Grid middleware, which coordinates all the necessary steps to retrieve, process, and display the requested products selected from the large database of ESA. This makes G-POD an ideal environment for processing large amounts of data, developing services which require fast production and delivery of results, comparing approaches and fully validating algorithms. In particular the G-POD concept and technology solves the equations:

- Move processors close to the data in a flexible and controlled way, thus leave the data wherever they are archived, reduce dissemination costs and effort
- Resources can be shared (data, tools, computing resources), thus reducing investments and running costs and reducing data flows to the minimum, with clear reliability and performance improvements.

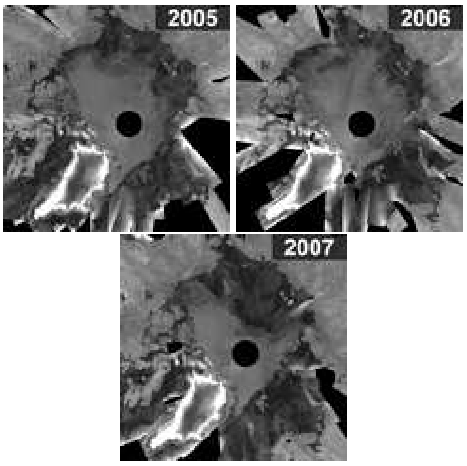

Fig. 2. Radar mosaics of the Arctic region related to the early September of 2005, 2006, and 2007. The mosaics have been generated in G-POD and show the shrinking of the Arctic ice coverage.

A number of selected ESA Earth Observation missions and related software tools have been integrated by ESA into G-POD to provide for facilitating data handling and analysis. G-POD is nowadays used for global scale product generation, research activities, rapid mapping and so on. ESA has recently offered scientists with the possibility to perform bulk processing and/or validation of their own algorithms exploiting this environment. Technically speaking, the ESRIN controlled infrastructure currently has a computing element of more than 200 PCs, with storage elements of more than 120 Terabytes, all part of the same Grid LAN in ESRIN and partially interfaced to other Grid elements in other ESA and partners facilities such as the European Space Research and Technology Centre (ESTEC), the European Space Astronomy Centre (ESAC), and EGEE EGEE (2008). The key feature of this Grid environment is the layered approach based on the GRID-ENGINE which interconnects the application layer with different Grid middleware (at present interfaced with three different brand/releases of middleware, Globus Globus (2008), LCG LCG (2008), and gLite gLite (2008). This characteristic enables the clear separation and development path between the Earth Observation applications and the middleware being used.

Many EO applications are fully operational and available through the ESA EO Grid portal. Some of them exploit the Grid technology for bulk processing of huge amount of products allowing e.g. the user to obtain large-scale high-resolution mosaics of EO data, other to provide fast access to large volumes of data. It is important to note the wide diversity of EO application themes, such as: meteorology, chemistry of the atmosphere, oceanography, simulations, operational generation of Level 3 products, generation of different products relevant to Essential Climate Variables (ECVs) OpenGIS (2008) defined by GCOS, and production of maps for fast damage assessment. Methodologies for the analysis of multi-source data, time series, and data assimilation are being considered. Many universities, institutions, research centres, international organizations, and private companies are involved. Figure 1 shows some examples of MERIS Level-3 products generated in near real time in G-POD. In the following subsections, we details two G-POD applications, i.e., mosaicking of ASAR product for Arctic ice monitoring, and Fast Access to Radar Imagery for Rapid Mapping.

6.1. The monitoring of the Arctic ice

On September 2007, the following news was published on ESA web site: the area covered by sea ice in the Arctic has shrunk to its lowest level this week since satellite measurements began nearly 30 years ago, opening up the Northwest Passage a long-sought short cut between Europe and Asia that has been historically impassable ESA (2008a). The news had a world wide echo. The study showed several mosaics (see Figure 2.) Each mosaic was created from nearly 200 images acquired by the Advanced Synthetic Aperture Radar (ASAR) instrument aboard ESAs Envisat satellite. All the mosaics were generated in G-POD environment.

On August 2008, ESA web news reported that following last summer's record minimum ice cover in the Arctic, current observations from ESA's Envisat satellite suggest that the extent of polar sea-ice may again shrink to a level very close to that of last year ESA (2008b). Also in this case a series of mosaics of the Arctic Ocean created from images acquired from the Advanced Synthetic Aperture Radar (ASAR) instrument aboard Envisat were made public (see Figure 3).

6.2. G-POD for rapid mapping: FAIRE

Among the different applications we wish to detail the Fast Access to Imagery for Rapid Exploitation (FAIRE) Cossu, Bally, Colin, and Fusco (2008) Cossu, Bally, Colin, and Fusco (2007). FAIRE is a G-POD integrated service to support image analysts in crisis and damage mapping for disaster management, where near real time access to both recent and historical data plays a key and critical role.

EO is a recognized source of information for disaster management. EO based crisis mapping services are generally delivered via projects such as GSE RISK EOS and GSE RESPOND alongside with the International Charter Space and Major Disaster, which enable to provide timely access to crisis data from a variety of EO missions. The all-weather capability of Synthetic Aperture Radar (SAR) data provides useful input to crisis and damage mapping. This is particularly relevant for flood monitoring. For these reasons, G-POD FAIRE was deployed to provide rush crisis mapping products combining SAR based observations with other EO crisis data. The underlying Grid technology accelerates the access and exploitation of ERS-2 SAR and Envisat ASAR in order to enable accurate, rapid and large coverage observations of flooded features. Through a dedicated user-friendly web interface, authorized users can browse for the required products specifying the geographical area of interest as well as the acquisition time, and, if required, limiting the search for a given mode or pass. All data acquired after June 2005 are available in the system. New acquisitions become accessible approximately after 30 minutes from the Level 1 generation in ESA Ground Segment. The service allows the user to calibrate and coregister a multitemporal series of (A)SAR data. Data can be corrected from topographic effects (using SRTM DEM dataset) and projected to specified projection systems, e.g., UTM.

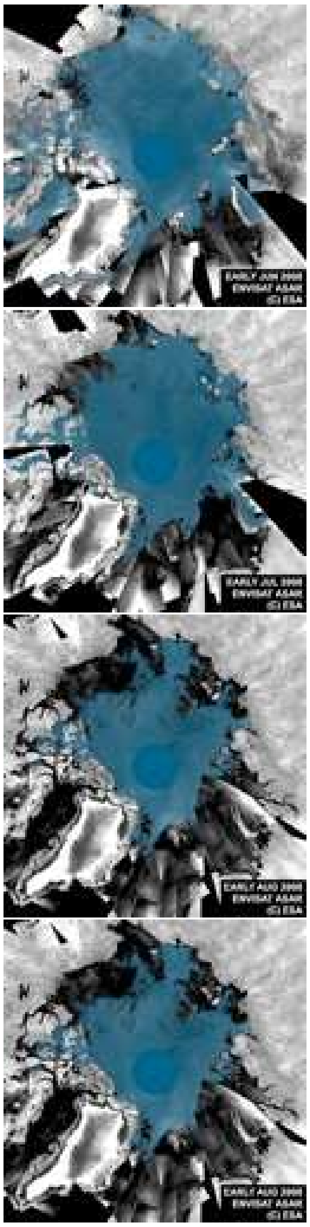

Fig. 3. Radar mosaics of the Arctic region related to (from top to bottom) the early June , July , August , and September 2008. The mosaics have been generated in G-POD.

Fig. 4. Multitemporal composition obtained automatically in FAIRE and used to support image analyst in generating flood maps. The image is related the flood that affected Bihar State, India, in August 2008. Blue colour represents the areas most likely flooded. The image analyst has to further interpret such image determining which areas are really flooded and avoiding false alarms (due for example to other kind of changes).

In addition, by specifying a crisis data, pre-crisis data are temporally averaged to produce a backscatter reference image that is compared to the post-crisis product to generate a quicklook where most-likely flooded areas are highlighted (It is worth noting that the system does not provide any Value Added Product but also intermediate products that help image analysts in their job). An example of such a kind of product is shown in Figure 4. The system automatically retrieves data stored on different storage elements (e.g. distributed archive), identifies the jobs needed for accomplishing the task required by the user, and distributes them on different computing nodes of the Grid. Processing usually requires one to two hours. G-POD FAIRE system is now used operationally in different activities of RESPOND and Charter where ESA is involved. Figure 4 shows an example of product generated by FAIRE and related to the flood that affected the Bihar State, India, in August 2008. Further interpretation of this product was carried out in the context of the International Charter, call 218, and led to flood map shown in Figure 5.

7. Toward an ES e-collaboration platform

Based on the successful experience of G-POD, a number of initiatives have started in ESA to develop an ES e-collaboration platform that makes use of Grid and SOA technologies. In the following subsections, we focus our attention on two ESA participated projects, i.e., DEGREE and GENESI-DR. The first one was a Specific Support Action project aimed to widen the use of Grid infrastructure by showing how Grid services can be integrated within key selected ES applications, the latter will practically deploy an ES dedicated infrastructure providing reliable, easy, long-term access to Earth Science data via the Internet.

7.1. DEGREE

The Dissemination and Exploitation of GRids in Earth sciencE (DEGREE) project DEGREE (2008) 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.

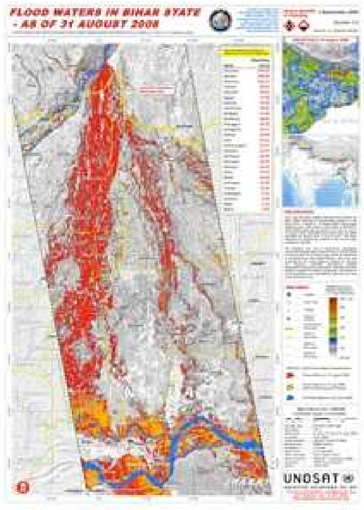

Fig. 5. Flood waters in Bihar State, 31/08/2008. Map derived by analysis of FAIRE output. Credit: Copyright ESA 2008 - Image processing, map created 01/09/2008 by UNOSAT. Further details at `http://www.disasterscharter.org/disasters/CALLID_218_e.html`

- 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.

Fig. 6. Dedicated ES infrastructure as envisaged by GENESI-DR project

7.2. The roadmap

An ES Grid development roadmap was missing and its definition was of the main objectives of DEGREE. It aims to set 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 harmonisation, reprocessing of data. 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. The major objectives of the roadmap were 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. Various aspects were considered 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. The key points have been arranged by category, specifying for each of them the expected achievements and the time period, from short-term to medium-term to long-term. The roadmap, which is in line the expected achievements of the major ES communities programmes/initiatives, like GEOSS Earth Observations website (2008) , SEIS SEIS (2008), and INSPIRE INSPIRE (2008), is publicly available at DEGREE-wp6 (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. Several activities are on going, like DEGREE dissemination, EGEE, 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, CEOS, GEO, space agency, IUGG, EGU) 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 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. 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. 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. Data retrieve, reprocessing and systematic data processing shall be standard services in Grid; the direct deployment to the Grid storage of the most used very large environmental databases and file 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. 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. 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).

7.3. GENESI -DR

The amount of information being generated about our planet is increasing at an exponential rate, but it must be easily accessible in order to apply it to the global needs relating to the state of the Earth. Ground European Network for Earth Science Interoperations - Digital Repositories (GENESI-DR) GENESI-DR (2008), an ESA-led, European Commission (EC)-funded two-year project, 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. So data access is a major logistical problem. The EC has funded GENESI-DR as a flagship project in Europe to help meet this challenge. GENESI-DR will allow scientists from different Earth Science disciplines located across Europe to locate, access, combine and integrate historical and fresh Earth-related data from space, airborne and in-situ sensors archived in large distributed repositories.

A dedicated infrastructure providing transparent access to all this will support Earth Science communities by allowing them to easily and quickly derive objective information and share knowledge based on all environmentally sensitive domains (Figure 6). The use of high-speed networks (GANT GEANT (2008)) and the experimentation of new technologies, like BitTorrent, will also contribute to better services for the Earth Science communities. In order to reach its objectives, the GENESI-DR e-Infrastructure will be validated against user needs for accessing and sharing Earth Science data. Initially, four specific applications in the land, atmosphere and marine domains have been selected, including:

- Near real time orthorectification for agricultural crops monitoring
- Urban area mapping in support of emergency response
- Data assimilation in GlobModel project GlobModel (2008), addressing major environmental and health issues in Europe, with a particular focus on air quality
- SeaDataNet SeaDataNet (2008) to aid environmental assessments and to forecast

the physical state of the oceans in near real time.

Other applications will complement this during the second half of the project towards the beginning of 2009. GENESI-DR also aims to develop 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 todays climate change data in 50 years will be able to better understand and detect trends in global warming and apply this knowledge to ongoing natural phenomena. GENESI-DR will work towards harmonising operations and applying approved standards, policies and interfaces at key Earth Science data repositories. To help with this undertaking, GENESI-DR will establish 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 projects 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. It is believed that GENESI-DR will push towards an easily accessible 'virtual repository', where an extremely large set of valuable scientific information will be available to small and medium enterprises and large companies, providing benefits to different actors, from data providers to service providers to end users.

8. 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 past, present and future initiatives like G-POD and EC funded projects DEGREE 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.

References

Group on Earth Observations website: `http://www.earthobservations.org/`

The Global Earth Observation System of Systems (GEOSS)10-Year Implementation Plan: `http://www.earthobservations.org/documents/10-Year%20Implementation%20Plan.pdf`

Kopernikus, former Global Monitoring for Environmental and Security website: `http://www.gmes.info/`

ESA Grid Processing on-Demand (G-POD) Web portal: `http://eogrid.esrin.esa.int`.

Dissemination and Exploitation of Grid foe Earth sciencE (DEGREE) website: `http://www.eu-degree.eu`

Ground European Network for Earth Science Interoperations - Digital Repositories (GENESI-DR) website: `http://www.genesi-dr.eu/`

Shared Environmental Information System (SEIS) website: `http://ec.europa.eu/environment/seis/index.htm`

International Charter Space and Major Disasters website: http://www.disasterscharter.org

RESPOND website: http://www.respondint.org/Respond

GMES Service Elements (GSE) website: http://www.esa.int/esaLP/SEMSVM0DU8E_LPgmes_0.html

Global Climate Observing System (GCOS) website: http://www.wmo.int/pages/prog/gcos/

International Council for Science (2004): Report of the CSPR Assessment Panel on scientific Data and Information, December 2004, ISBN 0-930357-60-4.

P. Gauthier, et al (2003): Ensuring the sustainability of online Cultural and Heritage Content: From an Economic Model to an Adapted Strategy, M2W FIAM - 2003.

H. Besser (1999): Digital Longevity, Chapter in Maxine Sitts (ed.) Handbook for Digital Projects: A Management Tool for Preservation and Access, Andover MA: Northeast Document Conservation Center, 2000, pages 155 166.

L. Fusco, and J. van Bemmelen, 2004, Data Science Journal, 3,222.

OpenGIS website: http://www.opengis.org.

Open Grid Forum 22 website: http://www.ogf.org/OGF22/

European Data Grid Project website: http://www.eu-datagrid.org.

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., 2003

G. Landgraf and L.Fusco, ESA bulletin 93, February 1998,

G. Landgraf and L. Fusco, on line http://www.esa.int/esapub/bulletin/bullet93/LANDGRAF.pdf

EGEE website: http://public.eu-egee.org

Globus website: http://www.globus.org

LCG website: http://lcg.web.cern.ch/LCG

glite website: http://glite.web.cern.ch/glite

ESA web news: Satellites witness lowest Arctic ice coverage in history, http://www.esa.int/esaEO/SEMYTC13J6F_index_0.html

ESA web news: Arctic ice on the verge of another all-time low, http://www.esa.int/esaEO/SEMCKX0SAKF_index_0.html

R. Cossu, Ph. Bally, O. Colin, L. Fusco, ESA Grid Processing on Demand for fast access to Earth Observation data and rapid mapping of flood events, European Geosciences Union General Assembly 2008, Vienna, Austria, 13 18 April 2008

R. Cossu, Ph. Bally, O. Colin, L. Fusco, Rapid mapping of flood events through the use of FAIRE- final, 6th Workshop on Remote Sensing for Disaster Management Applications

INSPIRE website : http://www.ec-gis.org/inspire

DEGREE Earth Science white paper on Grids: https://www.eu-degree.eu/DEGREE/internal-section/wp6/DEGREE-D6.1.2_v2.8.pdf/view

GANT website : http://www.geant.net/

GlobModel project website: http://www.globmodel.info

SeaDataNet website: http://www.seadatanet.org/