

## Towards a distributed Earth Science Data Infrastructure

---

**Joost Van Bemmelen<sup>1</sup>, Luigi Fusco, Roberto Cossu, Eliana Li Santi**

ESA

*Via Galilei, Frascati, Italy*

*E-mail: name.surname@esa.int*

This paper elaborates on initiatives and experiences at the European Space Agency establishment in Frascati (Italy) in implementing a distributed Earth Science data infrastructure. Of particular interest are experiences in the ESA-lead projects GENESI-DR and GENESI-DEC (short for Ground European Network for Earth Science Interoperations –Digital Repositories / Digital Earth Communities), the first one having established a data infrastructure providing transparent access to a network of (mainly European) Digital Repositories, the second one aiming at extending it to a worldwide e-Infrastructure serving heterogeneous communities interested in Planet Earth data, services and applications.

Furthermore, potential contributions to the Global Earth Observation System of Systems that would enlarge the world-wide GEOSS Common Infrastructure's capabilities are described, in response to data producers and consumers expressed needs regarding re-use and dissemination of Earth Observation data across different regions in the world. GENESI-DEC has been recently integrated in the GEO Portal to demonstrate how it can contribute to the enhancement of the GEOSS Common Infrastructure with new data access and processing capabilities.

GENESI-DEC is proactive in several OpenGeoSpatial Consortium (OGC) working groups on the following topics: Catalogue Services for the Web, Web Processing Service, Ordering Services for Earth Observation Products Standard, OpenSearch GeoSpatial Standard, Publish/Subscribe Standard, Web Map Context Implementation.

*The International Symposium on Grids and Clouds and the Open Grid Forum  
Academia Sinica, Taipei, Taiwan  
March 19 - 25, 2011*

---

<sup>1</sup> Speaker

## 1. Introduction

Digital Earth is a visionary concept for the virtual representation of the Earth that is spatially referenced, interconnected with the world's digital knowledge archives, and encompassing all its systems and forms, including Earth Sciences, operational Natural Resources Management and Environmental Monitoring. In order to properly address the needs of several Earth Science communities and Policy makers, it is of fundamental importance to realize that vision by enabling and supporting the interconnection and integration of the relevant digital and human knowledge across relevant stakeholder communities. Digital Earth will underlie the current initiatives, which take full advantage, in the present digital age society, of the advancement in emerging technologies (space, instrumentation, ICT...) to ensure the monitoring and safeguarding of the Planet Earth, which is increasingly in danger due to the strong presence and impact of the human life. In fact, in response to alarming situation, the United Nations has supported environmental conventions attempt to define internationally agreed protocols (e.g. Kyoto, Biodiversity and Montreal) to limit and monitor the status of our endangered planet. 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) [1], 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. The GEOSS ten year Implementation Plan establishes the intent, operating principles, and institutional frame for developing a world-wide federated infrastructure for integrating available data and information to provide responses to the GEO institutional and scientific community.

The implementation of the Digital Earth (in the vision of GEOSS) needs data, tools and world-wide infrastructures to gather and share the data, services and processing resources for building Virtual Research Facilities, which will permit the Digital Earth Community to share and derive objective information and increase their knowledge in socio-economic and 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.

This 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 Earth, its state, socio-economic indicators, relevant services, data availability, project results and applications is accessible only in a very scattered way through different EO mission operators, scientific institutes, service companies, data catalogues, etc. For example, if we refer to Remote Sensing missions, only a limited community, who already knows what to search for, is today in a position to collect, compile and thus exploit the necessary EO information.

The European Space Agency is very active in implementing a distributed Earth Science data infrastructure. Of particular interest are experiences in the ESA-lead projects GENESI-DR [2] and GENESI-DEC [3] (short for Ground European Network for Earth Science

Interoperations –Digital Repositories / Digital Earth Communities), the first one having established a data infrastructure providing transparent access to a network of (mainly European) Digital Repositories, the second one aiming at extending it to a worldwide e-Infrastructure serving heterogeneous communities interested in Planet Earth data, services and applications.

The paper will describe the GENESI-DR achievements in terms of federation of distributed environmental data repositories, easy and fast access to heterogeneous data (airborne, in situ, satellite), effective data and service discovery, high performance (Grid-based) processing. It will discuss how GENESI-DEC will build upon such achievements for responding to specific needs expressed by heterogeneous user communities such as Disaster management, Seafloor and Ocean Earth Observation, Global Change Earth Observation, Global Atmosphere Observation and Agriculture Monitoring, with the final goal of establishing a multidisciplinary collaborative platform. Therefore, issues like multidisciplinary collaboration, creation of virtual research communities, interoperability between heterogeneous and regional infrastructures, integration of new scientific and technological paradigms in operational infrastructures in response to the latest Digital Earth requirements, operations harmonization, security and control will be discussed.

Furthermore, potential contributions to GEOSS that would enlarge the world-wide GEOSS Common Infrastructure's capabilities will be described.

## 2. GENESI-DR

Ground European Network for Earth Science Interoperations - Digital Repositories (GENESI-DR) (<http://www.genesi-dr.eu/>) is a project, co-funded by the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement n° 212073 addressing work programme topic INFRA-2007-1.2.1: Scientific Digital Repositories and implemented by a consortium led by the European Space Agency, aimed at providing reliable, easy, long-term access to Earth Science data via the Internet. GENESI-DR was concluded end May 2010.

The GENESI-DR project has achieved important objectives in providing reliable, easy, long-term access to Earth Science data and associated on-demand processing resources via the Internet. It has designed and implemented a multi-disciplinary platform, providing discovery capabilities of scattered and heterogeneous data, providing easy and fast access to such data, on demand computing resources, making easier the dissemination of newly generated results.

At the end of the project, 16 different Digital Repositories and 3 external e-infrastructures (interfaced via specific gateways) hosting more than 160 dataset series are integrated in GENESI-DR, for a daily increasing number of more than three million products. Dataset series refer to heterogeneous resources and include satellite data, in situ data, images acquired by airborne sensors, digital elevation models and model outputs.

GENESI-DR had a dedicated work package on data access policy issues. Various workshops regarding this topic have been organised to facilitate discussions among data providers with as main outcome that policies cannot that easily be changed; it is often a process of years before changes in policies can be accepted and as well implemented. As such, GENESI-DR has decided to respect the data access policies of the individual data providers.

Today, GENESI-DR provides the users with a single interface with the technical tools for discovering (and accessing) data but it is the data owner that grants them, as appropriate, with the *credentials* for accessing the data. In such cases, the user shall first accept the data policy and licenses, before actually being granted the access to the data.

GENESI-DR achievements have been disseminated in the GEO and GEOSS as example of effective management of large volumes and diverse types of Earth Observation data. The achievements of GENESI-DR are tangible and the success of the project is beyond question. However, a critical analysis of the present achievements highlights the following:

- All collaborations demonstrated the value of the federated approach, a next step is the GENESI-DR infrastructure to be evolved so it can be used in operations. In particular, some activities now consider GENESI-DR as their primary access point to ES data.
- GENESI-DR should also be extended to encompass the wider and multidisciplinary Earth Science user communities (we address them as Digital Earth Communities), including research, related institutional, industry & operational actors; collaborations have demonstrated the potential value of extending the infrastructure even to encompass socio-economic studies.
- Communities still need customised services in their specific environment while sharing data for collaboration. Each community has in fact its own tools, needs, habits, and it is necessary to make the core and essential GENESI-DR services (like data discovery, access and on demand processing) customisable as much as possible on these driving factors.

The challenge in front of us today is to offer a framework that allows scientists and users from different Digital Earth disciplines to discover, access, and process Earth-related data acquired by heterogeneous sensors as well as multi-sources information available in Digital Repositories (DRs) dispersed worldwide. This increase in data and resources access must be combined with the possibility to easily customize services according to the needs of the specific Digital Earth discipline, so easily building customised virtual research facilities. This is the rationale behind the follow-up GENESI-DEC project, which is currently enlarging and upgrading the GENESI-DR infrastructure addressing this real need of several Virtual Research Communities within the Digital Earth context.

### 3. GENESI-DEC

Ground European Network for Earth Science Interoperations - Digital Earth Communities (GENESI-DEC) (<http://www.genesi-dec.eu/>) is a project, co-funded by the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement n° 261623 addressing work programme topic INFRA-2010-1.2.3: Virtual Research Community and implemented by a consortium led by the European Space Agency, aimed at providing reliable, easy, long-term access to Earth Science data via the Internet. The project kicked off on the first of May 2010 and its duration is two years.

GENESI-DEC evolves and enlarges the platform developed by the predecessor GENESI-DR project by federating to and interoperating with existing infrastructures. It will provide the

multidisciplinary Digital Earth Community with mechanisms to share data and resources and with tools enabling collaboration and facilitating the building of user-configured virtual research facilities. GENESI-DEC involves key partners of ESFRI [4] projects and collaborates with key actors of Digital Earth and Earth Science initiatives, including GEO-GEOSS.

The objectives of GENESI-DEC are:

- **Enlarge Infrastructure:** To enlarge the existing GENESI-DR infrastructures in terms of data, resources availability and geographical extent.
- **Guarantee Service:** To provide guaranteed, reliable, easy, effective access to a variety of data, facilities, tools and services to an ever-increasing number of Digital Earth users from all disciplines.
- **Harmonise Federation:** To harmonise operations at selected key Digital Earth infrastructures limiting fragmentation of solutions.
- **Enable User Collaboration:** To enable multidisciplinary collaboration among Digital Earth users as well as the creation of user-configured virtual research facilities/test-beds.
- **Respond to Innovation:** To integrate new scientific and technological derived paradigms in operational infrastructures in response to the latest Digital Earth requirements.
- **Promote Virtualisation:** To stimulate, educate and support the creation of virtual Digital Earth research communities.

### 3.1 Digital Earth Communities

GENESI-DEC pays a great attention to all the aspects related to User Communities and in particular to Digital Earth Communities.

GENESI-DEC thoroughly analyses the Digital Earth Communities and their needs and consequently identify strategic disciplines to be considered in GENESI-DEC. At the beginning of the project, GENESI-DEC has started considering a predefined set of specific Digital Earth Communities within specific Earth Science domains/disciplines. In fact, thanks to the fact that GENESI-DEC is strongly built on the success of GENESI-DR, several liaison with external heterogeneous Digital Communities were already established at the project kick-off. These liaisons are mainly made via representative projects in the European and international scenario, relevant in the context of the Earth Sciences or research data infrastructures. Such liaisons are put in place in two different ways:

- by including in the consortium partners that are actively involved in the aforementioned projects (e.g., INGV for the EMSO project [5], DLR for the IAGOS project [6], MARIS for the SeaDataNet [7] project)
- by agreeing on and signing Memoranda of Understanding (e.g. the NASA GIOVANNI project [8], the University of Tokyo DIAS project [9])

Other communities and projects will be identified during the project lifecycle.

For each of the Digital Community, e-infrastructures and/or digital repositories of interest have been identified and specific Use Cases will be defined. A first set of Communities, infrastructures and uses cases is already identified.

During the project lifecycle, the set of Use Cases will be reviewed, enlarged, modified and enhanced, to take into account new identified needs of the already addressed communities

and/or to address other communities of interest. Other external infrastructures to be federated to or made interoperable with GENESI-DEC will be identified as well.

Specifying user requirements to support interoperation with new infrastructures: The defined use cases have been analysed and used to derive system requirements, which have been documented and transferred to the technical activities.

### 3.2 The GENESI-DEC Architecture

GENESI-DEC evolves from the predecessor GENESI-DR project, which has given a significant and recognized contribution in designing and implementing a multi-disciplinary platform. The platform developed by GENESI-DR provides discovery capabilities of scattered and heterogeneous data, easy and fast access to such data, on demand computing resources, and makes easier the dissemination of newly generated results. The GENESI-DR Architecture is realized by existing and newly developed services, interacting through SOAP and REST interfaces. GENESI-DEC inherited a federative infrastructure made up of 20 different Digital Repositories hosting more than 400 dataset series. Dataset series refer to heterogeneous and include satellite data, in situ data, images acquired by airborne sensors, digital elevation models and model outputs.

In the first year of activity, GENESI-DEC has worked to increase the number of series available through the infrastructure, to consolidate the set of metadata based on the feedback of User Communities (e.g., satellite atmospheric data) and to increase and guarantee the refreshing and updating of the catalogue with the acquisition of new data (for all the series of satellite data acquired by ESA and available online). Furthermore, the project team has worked to solve some open issues identified at the end of GENESI-DR and this has led to a new release of the GENESI SW. In particular, the Catalogue Access Service SW has been improved in terms of performances (e.g., time to response), features (e.g. querying capabilities), management (e.g., more user friendly interface to register new data, based REST interface); the GENESI-DR portal has been upgraded with new features that makes easier the discovery of the data and the development of highly customizable services.

While GENESI-DR's main focus was on "genesi-fication" of Digital Repositories (i.e., deploying new DRs or modifying existing ones to make them accessible through OpenSearch), GENESI-DEC considers as well pre-existing e-infrastructures, this way enlarging the federation of resources. Catalogue Access Services exposing OpenSearch interfaces were deployed and specific gateways designed and implemented to this end. In particular, thanks to the simplicity and flexibility of OpenSearch, which allows for discovery and access to data that are heterogeneous in nature (e.g., satellite data, in situ data, airborne data, models, maps), in format (e.g., hdf, netcdf, txt, xls, tiff, proprietary formats, etc), and in access protocol (e.g., http, https, ftp, gsiftp, OGC standards like wms, wfs, etc), OpenSearch is considered one of the key elements of GENESI-DEC today.

User requirements have been translated in system requirements and are currently under analysis. Related solution will be contained in the next GENESI-DEC SW release.

Specific research activities deal with the use of semantics and ontologies to improve discovery capabilities and interoperability between applications and services, and with the security in heterogeneous federated repositories. At present, a dedicated service allows the user to discover *geolocation of terms available in DBpedia* so support the users in identifying the geographical areas where to carry out the analysis. For instance, a user can browse for “wine regions”, or “desert” or “nuclear power stations in Japan”. An additional ontology service allows the users to browse the available series by categories or applications.

### 3.3 Standardisation and dissemination activities

GENESI-DEC has established key collaborations in the frame of Global Environmental initiatives, such as the Global Earth Observation System of System (GEOSS), which will provide decision-support tools to a wide variety of users, being a global and flexible network of content providers allowing decision makers to access an extraordinary range of information at their desk. This ‘system of systems’ will proactively link together existing and planned observing systems around the world and support the development of new systems where gaps currently exist.

Within the frame of the GEO-GEOSS work plan 2009-2011 and its DA-09-02a GEO task for effective management of large volumes and diverse types, an Alliance has been proposed among selected data centres/projects (in particular DIAS, GIOVANNI and GENESI-DR/GENESI-DEC) to promote a common vision towards the GEOSS Common Infrastructure.

GENESI-DEC has been recently integrated in the GEO Portal to demonstrate how it can contribute to the enhancement of the GEOSS Common Infrastructure with new data access and processing capabilities. This enhancement is the objective of the GEOWOW project shortly described in the next Section.

Furthermore, GENESI-DEC is proactive in several OpenGeoSpatial Consortium (OGC)[10] working groups on the following topics: Catalogue Services for the Web, Web Processing Service, Ordering Services for Earth Observation Products Standard, OpenSearch GeoSpatial Standard, Publish/Subscribe Standard, Web Map Context Implementation.

## 4. GEOWOW

The integration of GENESI-DR/DEC in the GEO-Portal is contributing to spread awareness in the GEOSS community of the validity of the GENESI-DR/DEC approach in general and of the advantages of the OpenSearch technologies in particular.

The development of GEOSS has multiple strands as identified in its 10-year implementation plan: to support the Societal Benefits Areas (SBAs), to harmonise and sustain Earth Observation and the use of modelling, to foster and make easier data sharing, to build capacities.

A key component of GEOSS is the GEOSS Common Infrastructure (GCI), which includes the above mentioned GEO-Portal and should allow users to discover, access and use the data and services available through GEOSS.

The GEOWOW project, short for GEOSS Interoperability for Weather Ocean and Water, has the overall aim of injecting the research and development necessary to mark a step change in the functionality and performance of the GCI. To this purpose the project will build on recent European research efforts in which some of its partners (ESA, Terradue, EC-JRC, CNR) have been key players such as GENESI-DR/DEC and EuroGEOSS and will follow the way paved by the integration of GENESI-DR in the Geo-Portal.

A pre-requirements analysis carried out across multiple SBAs indicates that the main requirements for the GCI in respect to its desired functionality can be grouped as *needs for*:

- Improved services for discovery and access to data
- Processing services
- Improved visualisation services
- Setting-up more complex services e.g., by concatenation and/or workflow services
- Meeting specific data, derived priority parameters and application requirements identified by the communities that are behind each single SBA and/or across SBAs.

The project implementation will enrich and extend the GEO Common Infrastructure with services and/or functionality to offer science users improved access to and usability of data and services. It will also critically evaluate the progress made and contribute to the evolution of the GEOSS architecture with dedicated tasks, in partnership with the GEOSS Committees, users, and governance structures. A particular focus of the project is to support inter-disciplinary interoperability, conscious that many of the key scientific questions of our time require an integrated multi-disciplinary effort to be addressed adequately.

In summary, the main GEOSS interoperable infrastructure related project objectives are:

- To propose and validate a distributed architectural model federating Earth observation and other Earth Science data holdings, including specific communities' infrastructures, and put this model forward as the European contribution to the evolution of the GEOSS architecture.
- To develop innovative methods for harmonized access and use of heterogeneous data, services, and models to foster sharing of knowledge among multiple disciplines, and more integrated assessment necessary to advance global sustainability research.
- To contribute to the GEOSS Common Infrastructure (GCI) interoperability, standardisation and operability via developments and evolution.
- To develop and support services for data dissemination, access, use (and processing) for the selected SBAs, contributing to the development of assessment tools and monitoring methods for sustainable development.
- To establish, harmonise and promote data sharing and usage procedures consistent with the GEOSS Data Sharing Implementation Guidelines, and contribute to the development of the GEOSS Data CORE.

In parallel, the project will address specific thematic objectives to support the users of the Weather, Water and Ecosystem SBAs, specifically:

- Weather SBA – To provide harmonized and fast data access for meteorological hazards/ extreme events, eg floods, including pre-processing services aimed at making data of immediate use.



- Water SBA – To deploy an e-infrastructure giving access to in-situ and satellite data as needed by hydrological application and Run-off process;
- Ecosystem SBA – To support and enhance the access to in-situ and satellite ocean observations, to information on threats to ocean ecosystems, and to key ocean forecasts and projections - for research and assessment.

## 5. Conclusions

The implementation of the Digital Earth (in the vision of GEOSS) needs data, tools and world-wide infrastructures to gather and share the data, services and processing resources, which will permit the Scientists, Policy Makers and citizens to share and derive objective information and increase their knowledge in socio-economic and 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. ESA-ESRIN is very active in developing a distributed Earth Science infrastructure able to responds effectively to these needs. GENESI-DR and GENESI-DEC has significantly contributed to this task and have paid the way to the GEOWOW project in its challenging objective of enhancing the GEOSS Common Infrastructure in response to the requirements of several Social Benefit Areas.

## 6. Acknowledgement

This work has been partially supported by GENESI-DEC - Ground European Network for Earth Science Interoperations - Digital Earth Communities - (<http://www.genesi-dec.eu/>), co-funded by the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement n° 261623 and implemented by a consortium led by the European Space Agency.

## References

- [1] GEO-GEOSS: <http://www.earthobservations.org/>
- [2] GENESI-DR project, <http://www.genesi-dr.eu>
- [3] GENESI-DEC project, <http://www.genesi-dec.eu>
- [4] ESFRI: [http://ec.europa.eu/research/infrastructures/index\\_en.cfm?pg=esfri](http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri)
- [5] EMSO project: <http://www.emso-eu.org/management/index.php>
- [6] IAGOS project: [http://mozaic.aero.obs-mip.fr/web/features/information/iagos/buf/mainColumnParagraphs/0/document/IAGOS\\_Flyer.pdf](http://mozaic.aero.obs-mip.fr/web/features/information/iagos/buf/mainColumnParagraphs/0/document/IAGOS_Flyer.pdf)
- [7] SEADATANET project: <http://www.seadatanet.org/>
- [8] NASA GIOVANNI: <http://disc.sci.gsfc.nasa.gov/giovanni>
- [9] DIAS: <http://www.editoria.u-tokyo.ac.jp/dias/english/index.html>
- [10] OpenGeoSpatial Consortium: <http://www.opengeospatial.org/>