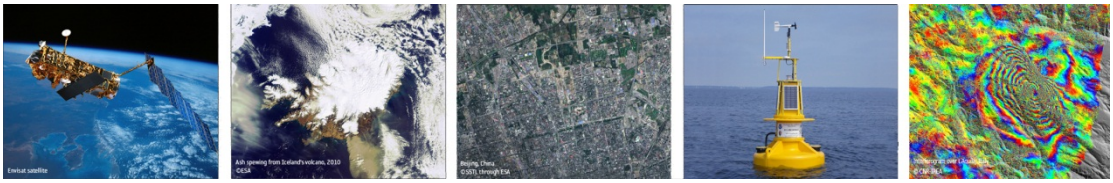
Publishable Summary



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 29 months. The contact person is Roberto Cossu (roberto.cossu@esa.int).

Introduction: **Digital Earth** is a visionary concept for the virtual representation of the Earth that is spatially referenced, interconnected with the world's digital data repositories, and encompassing all its systems and forms, includine Earth Sciences, Natural Resources Management, Environmental Monitoring system and human society dimensions. GENESI-DEC will establish open data and services access, allowing European and worldwide Digital Earth Communities to seamlessly access, produce and share data, information, products and knowledge. This will create a multi-dimensional, multi-temporal, and multi-layer information facility of huge value in **addressing global challenges such as biodiversity, climate change, pollution and economic development**. GENESI-DEC evolves and enlarges the platform developed by the predecessor GENESI-DR project by federating to and interoperating with existing infrastructures.

GENESI-DEC involves key partners of ESFRI projects and collaborates with key actors of Digital Earth and Earth Science initiatives, including the International Society of Digital Earth and GEO-GEOSS. Thus efficient use of already existing and planned developments is guaranteed.

Objectives: The objectives of GENESI-DEC are:

- *Enlarge the Infrastructure*: To enlarge the existing GENESI-DR infrastructure 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.

Digital Earth Communities: GENESI-DEC pays a great attention to all the aspects related to User Communities and in particular to Digital Earth Communities. To this end a dedicated activity (WP3) is in charge of:

- liaising with the wide Digital Earth community and relevant infrastructures projects,
- analysing and eliciting the needs and functionalities required by the Digital Earth Communities, also by defining high level use cases;
- documenting the community requirements and passing them to the technical activities of the project which undertake the infrastructure design and development work;
- identify and develop specific applications matching the community needs;
- once the upgraded system is deployed, validate GENESI-DEC to assure that the community needs are correctly addressed by the infrastructure.

WP3 significantly contributes to the main goals of the project by both guiding the enlargement of the GENESI-DEC infrastructure via interoperation with other e-infrastructures and building multidisciplinary services and customised Digital Earth facilities.

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. For each of the Digital Earth Community, e-infrastructures and/or digital repositories of interest have been identified and specific Use Cases defined. These have been analysed and used to derive user requirements, which have been documented and transferred to the technical activities.

Five applications have been implemented in the first project phase, which enrich GENESI-DEC with joint visualization capabilities of surface and seafloor data as well as aircraft and satellite data, and also with orthorectification and land cover mapping capabilities. The successful validation of the newly developed GENESI-DEC functionalities against the user requirements has been carried out.

In the second project phase new communities have been identified and analysed and collaborations established with new FP7 projects. Thirteen new use cases have been devised (and corresponding requirements specified): these use cases, while serving the needs of the identified communities, fully address the challenging Project objectives, from discovery and access of heterogeneous data, to discovery and usage of processing services and applications to multidisciplinary and inter-user collaboration.

New applications have been designed and implemented, as needed from the new use cases. They add up to the five already developed in the first project phase and provide GENESI-DEC with new capabilities, including ship detection services, disaster management support services that allow e.g. earthquakes damages assessment and analysis; services for the Calibration and Validation of data.

Moreover, two of the applications delivered in the first project phase (the Visualization Tool and Near real Time Orthorectification Service) have been further improved.

In the third project phase WP3 has carried out the validation activities on the final GENESI-DEC infrastructure, which were are aimed at verifying the compliance with the elicited user requirements.

The GENESI-DEC Platform: GENESI-DEC evolves from the predecessor GENESI-DR project, which has given a significant and recognized contribution in designing and implementing a multidisciplinary 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 has inherited a federative infrastructure hosting more than 166 dataset series**.** Dataset series include satellite data, in situ data, images acquired by airborne sensors, digital elevation models and model outputs. At the end of the first year of activity, the number of dataset series has increased to 440. **During the second and third reporting periods**, the number of series has increased from 440 to more than 1500. The work done in the first project year aimed at increasing and guaranteeing 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) and has also led to a significant increase of available products within the series.

Significant improvements in interfacing the CEOS WGISS Integrated Catalogue (CWIC) have been obtained, thus it is now possible to seamlessly discover data from US, Chinese and Brazilian providers. The metadata related to the Italian Ministry of Environment have been consolidated so to be fully compliant to the INSPIRE directive.

In the third reporting period, all the operations needed to make the Web Portal interoperating/interface with the services and tools developed by WP6 and WP7 have been performed. A new processing environment, based on OGC WPS standard, has been deployed and an aggregator node has been deployed to store and publish the information related to the available WPS services, so that these are discoverable in the same way as dataset series are.

Research activities: The two major GENESI-DEC deliveries, as described while presenting the meaning and the products of the virtuous cycle of innovation, will include the outcomes of the two joint research activities (JRAs) identified by the consortium.

The first JRA (WP6- Geosemantics, Ontology and Workflows) is concerned with the use of semantics and ontologies in assisting data discovery and geospatial services composition. **During the second year** of the project, the semantic framework developed in the first year has been also enhanced by enriching the content and quality of the knowledge database and by developing new services that allows not only an easier data discovery, but also processing services discovery, data and services tagging, and workflow composition (e.g. identification and selection of data to be used for a particular processing service for a specified time period and over a specific geographic area).

The second JRA (WP7 – Security Frameworks Interoperability) is tackling the security in heterogeneous federated repositories. The outcomes of WP7 are expected to allow users to log in once to the system and gain access to the resources and services available in the federation without being prompted to log in again, despite the physical location and ownership of the different data and/or resources selected. **In the second year of activity,** WP7 has improved the components developed during the first year. The main developments include a Shibboleth/OpenID bridge, which is planned to be used in the ESA SSO operational environment, and an OpenLDAP attribute mapper, to provide the users with an easier access to secured data. Following the proof-of-concept implemented and tested by WP7 in GENESI-DEC, SeaDataNet is now also implementing an OpenID/Shibboleth extension to its authentication/authorisation service normally driven by CAS. This will allow a smoother access to data for users from other communities, among which GENESI-DEC users.

Furthermore a test repository using OpenID and the LDAP attribute mapper is also planned to be deployed. Security requirements in processing services have also been extensively analysed and addressed.

In the third reporting period a new set of thesaurus has been introduced to describe and annotate resources (dataset series, services and workflows). In details GENESI-DEC concepts, SBA and GEMET thesaurus have been used. GCMD has been integrated thanks to NASA department that has been very happy to be collaborating with GENESI-DEC,

inserting the project WP6 among the developer user of GCMD. The harvesting has been extended to also encompass the services catalogue harvesting procedure. The information collected during this phase, allows enriching the knowledgebase with information links between services and dataset series. It makes possible to reach all dataset series that can be used by a service starting from the service element and vice-versa.

The update functionality in the harvesting process has been integrated. The new function allows updating the knowledgebase upon catalogue changes without requiring the recreation from scratch of the knowledgebase. A new approach to the annotation of resources has been identified and implemented: automatic in the harvesting phase and manual in the discovery phase. The automatic tagging now manages a set of more than 5000 terms (without GCMD) during the harvesting phase. A new function has been created in order to perform, with a time selectable slot, a garbage collection operation. This functionality allows enforcing full alignment of the knowledgebase with the registered catalogues.

The discovery operations have been redesigned to increase the response time. Managing around 2000 resources entries with around 10 annotations each, means an average of 2000+(2000*30) triple entries in the semantic store that have implied an optimization of the related SPARQL queries.

Standardisation and dissemination activities: Since the beginning of the project, 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. Another major key collaboration is in the frame of environmental Research Infrastructures, and more specifically with **ESFRI** projects.

GENESI-DEC integration in the GEO Portal has demonstrated how it can contribute to the enhancement of the GEOSS Common Infrastructure with new data access and processing capabilities. A number of successful demonstrations of GENESI-DEC capabilities in easily discovering and accessing heterogeneous data have been made included the one at the GEO Plenary, in November 2011. The enhancement of the GEOSS Common Infrastructure will be addressed by the EC-FP7 GEOWOW project starting from the GENESI-DEC achievements.

Within the frame of the GEO-GEOSS work plan 2012-2015 GENESI-DEC is contributing in providing data access and supported the GEOWOW project to prepare a common vision towards the GEOSS Common Infrastructure.

GENESI-DEC has also contributed to the GEO Hazard Supersites. This is an initiative of the geohazard scientific communities aimed at providing access to space-borne and in-situ geophysical data of selected sites prone to earthquake, volcano or other hazards. Terradue, partner in GENESI-DEC, is providing Cloud Virtual archive for the biggest ESA SAR data repatriation. This huge amount of SAR data, over than ten thousand images, are becoming accessible to science communities dealing with interferometry, landslide and change detection. The virtual archive is coupled with complementary services including data discovery. This implements simple interface such as OpenSearch and results in Atom, RDF and KML format.

EMODNet (European Marine Observation Data Network) is an important infrastructure currently in development. SeaDataNet (community targeted in GENESI-DEC) has qualified as one of the core components, for archived marine in situ data. Especially its CDI (Common Data Index) services is important, which creates opportunities for metadata discovery and data access of this distributed infrastructure where around 70 data centers are currently

connected to. MARIS has developed OpenSearch interfaces on top of the SeaDataNet CDI catalogue which contains over 1 million marine in-situ datasets, as well as a machine to machine interface to allow ordering and download of datasets by external portals and applications with trusted users. In combination with the OpenID/Shibboleth extension of the AAA services, this improves discovery and access to marine in-situ data from external portals like GENESI-DEC.

GENESI-DEC has also established successful collaborations with several **Environmental ESFRI projects**, in addition to the ones represented in the consortium. INGV has developed OpenSearch interfaces to the MOIST catalogues, which contained data that will be included in the EMSO RI. DLR, partner in IAGOS, is now using OpenSearch to cataloguing its flight data. The ENVRI project which has among its goals to achieve interoperability among Environmental ESFRI infrastructures, considers GENESI-DEC as the basis for the technical development to be done for achieving this challenging objective. Preliminary results already include discovery and access of samples of data from EMSO, EuroArgo, ICOS, EPOS, Eiscat-3D.

Several other projects are making use of GENESI-DEC or OpenSearch for publishing their data or discovering data from other providers. For example, the ngEO project (Next Generation User Services for Earth Observation) of the ESA has adopted OpenSearch. The ngEO system is a strategic project for ESA's user services, designed to replace the current access system to ESA products (EOLi), used by the whole industry to apply for and acquire the products of practically all the agency's missions.

GENESI-DEC is proactive in several **Open GeoSpatial 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. GeoNetwork, a catalog application to manage spatially referenced resources, now supports OpenSearch Catalog interfaces.

Highlights and Conclusions: During the second and third periods of activity, GENESI-DEC has significantly enhanced the operational platform resulting from the first innovation cycle. This platform is aimed at offering Earth scientists reliable, easy, long-term access to Earth Science data via the Internet. In particular, GENESI-DEC has increased the number of products discoverable and accessible through the platform, has consolidated the set of metadata for the already registered series, has improved the reliability and freshness of data available. Research Activities have delivered a large set of components that have been fully integrated in the platform. These allow a semantically enriched search of data and processing services, the automatic identification of processing services available for the discovered data, data tagging and workflow chaining as well as a single-sign-on experience that simplifies data access for authorised users.

Many of these features are the response to the needs of the different Digital Earth Communities identified within the project that have allowed the definition of several challenging use cases that have been used in the third reporting period to validate the platform.

Several collaborations with projects and external bodies have been established with the aim of educating user communities to the use of Research Infrastructures, gathering user requirements. Particular attention has been given to GEO-GEOSS and to Environmental ESFRI projects.

The success story: GENESI-DEC has given a significant and recognized contribution in designing and implementing a multidisciplinary platform. This platform 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 range of data and services made available by GENESI-DEC is extensive. Millions of satellite data, in situ temperature profiles, data acquired by airborne sensors, sea surface and seafloor data, volcanic data, colour orthophotos and digital elevation models, are available through this platform.

GENESI-DEC is a federation of resources with catalogues that expose OpenSearch interfaces. Search clients, like the GENESI-DEC portal, can use OpenSearch description documents to learn about the public interface of the search engines in the federation. These description documents contain parameterized URL templates that indicate how the search client should make search requests, so allowing different communities to specify the parameters more relevant for their searches. OpenSearch description documents can be extended with foreign markup provided that all foreign elements and attributes are associated with an explicit XML namespace distinct from that of the core OpenSearch format.

GENESI-DEC has adopted OpenSearch and has successfully promoted this approach, by showing its flexibility, its easiness to be used, and by demonstrating how it can respond the needs of several different scientific communities. One of the biggest success story of the GENESI is in fact the adoption of OpenSearch by several projects and initiatives so putting the basis for an operational worldwide multidisciplinary environmental federation.

Frontier environmental research increasingly depends on a wide range of data and advanced capabilities to process and analyse them. GENESI-DEC has significantly contributed in starting the processing for a worldwide multidisciplinary environmental federation providing scientists with an easy discovery and access to heterogeneous data. The re-use and evolution of GENESI-DEC in ENVRI will speed up the construction of several European environmental infrastructures and will allow scientists to use the data and software from each facility to enable multi-disciplinary science.

As already mentioned, GENESI-DEC is already enabling, together with other technologies and approached currently considered by the GEOWOW project, GEOSS stakeholders to discover and access an increasing number of data. The decision of building GEOSS came from the recognition that international collaboration is essential for exploiting the growing potential of Earth observations to support decision making in an increasingly complex and environmentally stressed world. GEOSS will yield a broad range of societal benefits including reducing loss of life and property from natural and human-induced disasters, better understanding of our planet, better management of natural resources.