

GENESI-DEC: a federative e-infrastructure for Earth Science data discovery, access, and on-demand processing

Roberto Cossu¹, Fabrizio Pacini², Fabrice Brito³, Luigi Fusco¹, Eliana Li Santi¹, and
Andrea Parrini²

¹ European Space Agency, Earth Observation Science, Applications and Future
Technologies Dept, via Galilei, 00044 Frascati, Italy

² Elsag Datamat spa, Space Business Unit, User Segment Operations - Advanced
Technologies Group, via Laurentina 760, 00143 Roma Italy

³ Terradue Srl, Via Lunati 10, 00044 Frascati, Italy

roberto.cossu@esa.int, fabrizio.pacini@elsagdatamat.com, fabrice.brito@terradue.com,
luigi.fusco@esa.int, eliana.li.santi@esa.int, andrea.parrini@elsagdatamat.com

Abstract. 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, including Earth Sciences, Natural Resources Management, Environmental Monitoring system and human society dimensions. This paper introduces the European Commission co-funded project GENESI-DEC. 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.

Keywords: earth science, data access, data discovery, processing on-demand, grid, data federation, interoperability, digital earth.

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 data repositories, and encompassing all its systems and forms, including Earth sciences, natural resources management, environmental monitoring system and human society dimensions. The deployment of such representation would significantly improve the way we collect, analyse and communicate information on our environment, which is one of the main challenges of the Shared Environmental Information System. This paper introduces the European Commission co-funded project GENESI-DEC (Ground European

Network for Earth Science Interoperations – Digital Earth Community) [1]. 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. The objectives of GENESI-DEC are:

- Enlarge infrastructure: to enlarge the existing GENESI-DR [2] 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.

In order to reach the objectives listed above, the project proposes a repetitive approach built around a *virtuous cycle of innovation*. In this approach GENESI-DEC (system, services, capabilities, etc.) will be validated against Digital Earth Community needs for accessing and sharing Digital Earth data and resources, and enhanced twice in the project life-cycle and as needed, based on the results of the research activities, taking into account developments and constraints in the areas of standardisation, technologies and policies.

Section 2 gives an overview of the GENESI-DR infrastructure and its evolution towards GENESI-DEC. Section 3 provides details on the discovery mechanism promoted and adopted by GENESI-DR and GENESI-DEC. Section 4 analyses and the major research topics that will be addressed in the project. Section 5 describes the selected Digital Earth Communities to be used for requirements gathering and system validation. Section 6 concludes this paper by presenting the expected impacts.

2 GENESI-DEC e-infrastructure

GENESI-DEC build upon the experience gained and the infrastructure deployed in the predecessor project GENESI-DR. The GENESI-DR project has established an Earth Science infrastructure federating heterogeneous digital data repositories. Through the GENESI-DR Web Portal, users can search and access resources physically located at the DRs.

Figure 1 provides an high level overview of the GENESI-DR system. The Central Discovery Service, hosted at the so-called Central Site, allows discovery of data existing in heterogeneous Digital Repositories. It can be accessed by users via the internet through a user interface, the GENESI-DR Web Portal, or by external

applications via the exposed machine-to-machine interface. The GENESI-DR Web Portal provides data discovery capabilities and access to processing services that make use of high performance Grid-based computing resources to allow for scaling up computing power and storage capacity.



Figure 1 Overview of GENESI-DR

The machine to machine interface is based on OpenSearch, a collection of technologies allowing websites and search engines to publish search results in a standard and accessible format (see Section 3 for more details).

The goal of the GENESI-DR project has been to build an Earth Science infrastructure aggregating heterogeneous repositories: GENESI-DEC will focus on how to use the expertise gained and the infrastructure set up with GENESI-DR for the benefit of scientists who need to access and process Digital Earth data. Specific user communities (referred to as Digital Earth Communities) have been and will be selected and specific use cases involving such communities have to be implemented by GENESI-DEC.

GENESI-DR will evolve towards GENESI-DEC through:

- Addition of new data and services: GENESI-DEC will federate new Digital Repositories and will provide new services as needed by the specific use cases.
- Federation of existing infrastructures: some of the selected use cases need data that reside in existing infrastructures, so GENESI-DEC will interoperate with such infrastructures either via standard interfaces (following and enhancing as needed the GENESI-DR approach) or via implementation of specific gateways. Initially the following infrastructures will be considered: EMSO; SeaDataNet/GeoSeas; CEODE A-SDI.
- Creation of community dedicated views: While GENESI-DR services are accessible through the same user interface, the GENESI-DR Web Portal, which is independent of the type of user, GENESI-DEC will be accessible through several portals representative of the different communities and will provide dedicated visualization tools (see Figure 2). GENESI-DR provides a limited set of services, community specific and with hard-coded workflow: GENESI-DEC will provide a larger set of services, including generic (community-independent) and will allow users to build

services customised on the base of their specific needs, by means of workflow customisation tools. These latter, indeed, allow creating customised workflows made up of existing core processing services, which could be generic or specific of different communities so responding to multidisciplinary needs.

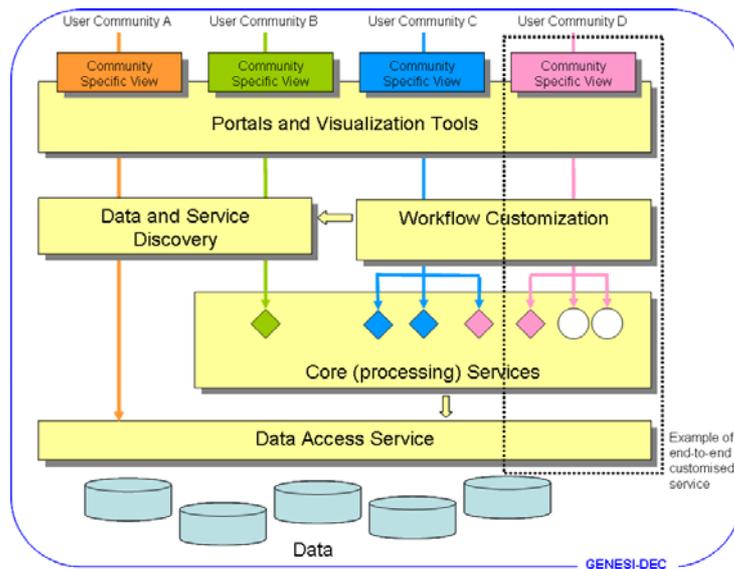


Figure 2 Community specific views and customized services

GENESI-DEC can be imagined as a ring where existing digital repositories holding Earth Science data as well as external e-infrastructures holding data and services can be integrated, to provide common and site-specific services. Research scientists, once connected to the GENESI-DEC ring, will eventually have access to all required data and services in a transparent and homogeneous way. The GENESI-DEC infrastructure will evolve progressively from a first deployment of existing services based on the most important improvements that have been identified in the final phase of GENESI-DR to its final configuration where several data and service providers and e-infrastructures will be federated. The main areas of improvements are: processing services, user workspace, search capabilities.

3 GENESI-DEC discovery service

As in GENESI-DR, data discovery services in GENESI-DEC will use OpenSearch interfaces. OpenSearch started in a community effort built around Amazon's A9.com and was originally intended to allow syndication of search results.

Coupled with geo-extensions, OpenSearch provides a collection of technologies that allow publishing search results in a format suitable for syndication and aggregation such as Atom, XML/RDF, KML, WKT, JSON among others. OpenSearch represents a way for websites and search engines to publish search results in a standard and accessible format. OpenSearch allows for hierarchical, drill-down methods of data discovery. Starting with the sole knowledge of the catalogue name, the GENESI-DEC users and applications can learn the series, dataset, and services associated with Earth Science data. OpenSearch Description Documents list search result responses for the given website/tool. It supports multiple responses, which may be in any format. GENESI-DEC metadata follows a DCLite4g model. Dublin Core Lightweight Profile for Geospatial is an extension vocabulary to Dublin Core proposed by the OSGeo Geodata Committee. Is a subset of the FGDC metadata element set, oriented towards practical repository management issues and provides classes that represent the Dataset Series and Dataset and their relations covering properties specific to a geographic application - file formats for GIS data, different coordinate systems.

The proposed OpenSearch geo extensions have been formalized within the Open Geospatial Consortium "OpenSearch Geospatial Extensions Implementation Standard" (09-084). This document codifies the OpenSearch GeoSpatial extensions in a form compatible with OGC specifications. Furthermore, GENESI-DEC partners are now responsible for the edition of the OpenGIS® Catalogue Service Implementation Specification, Version 3.0 - Part 4: OpenSearch Query Interface (10-032).

4. GENESI-DEC research activities

GENESI-DEC research activities will focus on two topics, i.e. use of semantic technologies for data discovery and workflow compositions, and security frameworks interoperability. Details are given in the following subsections.

4.1 Semantic technologies

GENESI-DEC will analyse and further enhance access and processing capabilities by taking into account ontology and geo-semantics.

Geo-semantics will be used in first instance to provide users with more effective and user-friendly search capabilities, by assisting them in the management and interpretation of (often huge amounts of) results. To this end already developed community specific ontologies will be used.

At a later stage, semantic technologies will be used to assist users during the definition of new processing services. In more details geo-semantic will be used to support users in the definition of customised task workflow.

4.2 Security Frameworks Interoperability

Security is paramount in Spatial Data Infrastructures (and more generally in research infrastructures) as broad engagement from resource/service providers and research communities is dependent on securing access to online data resources and services without disrupting end-users and applications developers accustomed (best practices/way of working). A number of initiatives have investigated cross-infrastructure federation and interoperability while addressing the need for user-friendly features such as single-sign on (SSO). The main aspects that GENESI-DEC will consider in this context are:

- architectural design and implementation of a federated security interoperability framework to meet GENESI-DEC stakeholders requirements;
- implementation of bridges to support the mapping of specific security models and technologies used by GENESI-DEC stakeholders and services to this framework;
- development of user-friendly paradigms and interfaces for sign-on and management of access rights.

The basic idea is to design and develop a brokering layer able, starting from the user identity credential in the origin framework, to issue short-lived security tokens suitable for the interaction with target resources/services and to support the delegation of these tokens in order to allow further interactions on behalf of the user.

Research in this task will leverage related work carried out in on-going initiatives/projects such as GENESI-DR, EGEE and GridShib. The solutions/approaches to security that will be analysed and integrated from the beginning of the project will be (a) Shibboleth with (b) X.509-based PKI and (c) OpenId/OAuth with (b) X.509-based PKI. This also reflects the fact that leading research and commercial infrastructures in Europe and world-wide providing access to storage and computational resources (e.g. EGEE, Amazon EC2/S3 etc.) support (b) for security whilst (a) and (c) are solutions well accepted and widely adopted by end-users especially for browser-based interaction patterns.

The framework itself may exploit quite different technologies to provide the mediation layer – among the options that will be considered are capability-oriented computing to dynamically provision policy translation and enforcement for cross-domain interactions.

5 Digital Earth Communities

GENESI-DEC aims to address Digital Earth Communities: a first set of these have already been selected, while additional ones will be identified and addressed during the project life.

While GENESI-DR was oriented towards the development and implementation of a digital repositories distributed infrastructure, GENESI-DEC will focus on a number of heterogeneous Digital Earth Communities, each with dedicated use cases:

- The seafloor and ocean observation community;

- The global atmosphere observation community using aircraft;
- The global change Earth Observation community;
- The territorial development and spatial planning community;
- The Black Sea catchment observation community.

In order to establish a real and effective link with these communities (so to properly gather and address their requirements) GENESI-DEC has already set up collaborations with related and relevant projects and/or organizations. Some of these projects have developed or are developing dedicated infrastructures (with data) to serve some of the needs of their communities. GENESI-DEC will federate to or interoperate with such dedicated DRs and infrastructures, bringing additional benefits to the Digital Earth Community.

GENESI-DEC establishes 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 a GEO-GEOSS work plan 2009-2011, the 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) to promote a common vision towards the GEOSS Common Infrastructure.

6 Conclusions

By enlarging the GENESI-DR platform, GENESI-DEC addresses and embraces a wide Digital Earth community. The project approach bridges the gaps among different disciplines of Digital Earth Community, providing a common, standardised, interoperable and multidisciplinary infrastructure, nonetheless able to address the specific needs of the single disciplines. By providing single and simple service interfaces to discovery, search, access and process on demand of many distributed environmental data holdings, GENESI-DEC reduces the *time-to-science* to a minimum and improves the effectiveness of many standard and complex operations for the Digital Earth Community.

Acknowledgments. GENESI-DEC is co-funded by the European Commission, FP7, grant agreement RI-261623.

References

1. GENESI-DEC, <http://www.genesi-dec.eu>
2. GENESI-DR, <http://www.genesi-dr.eu>