



GENESI-DR – A single access point to Earth Science data

R. Cossu (1), P. Goncalves (2), and F. Pacini (3)

(1) European Space Agency, ESRIN, (2) Terradue, (3) Elsag Datamat

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. 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. 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. The use of high-speed networks (GÉANT) and the experimentation of new technologies, like BitTorrent, will also contribute to better services for the Earth Science communities.

GENESI-DR (Ground European Network for Earth Science Interoperations - Digital Repositories), 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. This project 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.

GENESI-DR builds a federated collection of heterogeneous digital Earth Science repositories to establish a dedicated infrastructure providing transparent access to all this and allowing Earth Science communities to easily and quickly derive objective information and share knowledge based on all environmentally sensitive domains. The federated digital repositories, seen as services and data providers, will share access to their resources (catalogue functions, data access, processing services etc.) and will adhere to a common set of standards / policies / interfaces. The end-users will be provided with a virtual collection of digital Earth Science data, irrespectively of their location in the various single federated repositories.

GENESI-DR objectives have lead to the identification of the basic GENESI-DR infrastructure requirements:

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

Data discovery is based on a Central Discovery Service, which allows users and applications to easily query information about data collections and products existing in heterogeneous catalogues, at federated DR sites. This service can be accessed by users via web interface, the GENESI-DR Web Portal, or by external applications via open standardized interfaces exposed by the system. The Central Discovery Service identifies the DRs providing products complying with the user search criteria and returns the corresponding access points to the requester. By taking into consideration different and efficient data transfer technologies such as HTTPS, GridFTP and BitTorrent, the infrastructure provides easiness and speed of access. Conversely, for data publishing GENESI-DR provides several mechanisms to assist DR owners in producing a metadata catalogues.

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, addressing major environmental and health issues in Europe, with a particular focus on air quality
 - SeaDataNet 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.

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 today's 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.