Earth Sciences data access and preservation with gLibrary

Veronica Guidetti (1) and Antonio Calanducci (2)

(1) ESA-ESRIN, Via Galileo Galilei,00044 Frascati, Rome, Italy (veronica.guidetti@esa.int), (2) INFN, Catania, Italy (antonio.calanducci@ct.infn.it)

ESA-ESRIN, the European Space Agency Centre for Earth Observation (EO), is the largest European EO data provider and operates as the reference European centre for EO payload data exploitation. EO data acquired from space have become powerful scientific tools to enable better understanding and management of the Earth and its resources. Large international initiatives such as GMES and GEO, supported by the European Commission, focus on coordinating international efforts to environmental monitoring, i.e. to provide political and technical solutions to global issues, such as climate change, global environment monitoring, management of natural resources and humanitarian response.

Since the time-span of EO data archives extends from a few years to decades, their value as scientific time-series increases considerably, especially for the topic of global change. It will be soon necessary to re-analyse on global scale the information currently locked inside large thematic archives. Future research in the field of Earth Sciences is of invaluable importance: to carry it on researchers worldwide must be enabled to find and access data of interest in a quick and easy way.

At present, several thousands of scientists, principal investigators and operators, access EO missions' metadata, data and derived information on a daily basis. Main objectives may be to study the global climate change, to check the status of the instrument on-board and the quality of EO data. There is a huge worldwide scientific community calling for the need to keep EO data accessible without time constrains, easily and quickly.

In collaboration with ESA-ESRIN, INFN, the National Institute for Nuclear Physics, is implementing a demonstrative use case where satellite remote sensing data, including in-situ data and other kind of digital assets, are made available to the scientific community via gLibrary (https://glibrary.ct.infn.it), the INFN digital library platform.

gLibrary can be used to store, organise, browse, retrieve, annotate and replicate any kind of digital asset on data grids or distributed storage environments.
It provides digital assets preservation capabilities, making use of distributed replication of assets, decoupling from the underlying storage technology, and adoption of standard interfaces and metadata descriptions. In its future development gLibrary will investigate and possibly provide integration with grid and HPC processing services, including the ESA G-POD facility (http://eogrid.esrin.esa.int). Currently, gLibrary features encompass fast data access, quick retrieval of digital assets, metadata handling and sharing (including text annotation), high availability and scalability (due to its distributed architecture), (meta)data replication and, last but not least, authentication and authorisation.

Much of the experimentation is on-going at EC and international level to provide coordinated and interoperable access to EO data and satellite imagery including any kind of related digital assets (metadata, documents, product guidelines, auxiliary data, mission/sensor specifications, environmental reports). The work with gLibrary comes as a best effort initiative and targets a full interoperability with ESA EO data dissemination, recovering and processing services and intends to demonstrate the benefit the scientific community can gain from this kind of integrated data access. It contributes to respond to the Earth Sciences data users' needs, moving forward the technology development to facilitate a very interactive EO information sharing, analysis and interoperability on the Web.