EXPLOITING EARTH OBSERVATION DATA AT EUROPEAN SCALE

An EGI Open Data & Compute Platform for Sentinel Data

Authors: EGI Strategy & Policy and UCST Team
Version: 5.0
Date: 24th June 2016
Contact: policy@egi.eu

EGI Vision

Researchers and innovators from all disciplines have easy, integrated and open access to the advanced digital capabilities, resources and expertise needed to collaborate and to carry out data/compute intensive science and innovation

EGI Mission

Create and deliver open solutions for science and research infrastructures by federating digital capabilities, resources and expertise between communities and across national boundaries

Copernicus program and the Sentinels: current challenges

The new generation of Earth Observation (EO) satellites from the Sentinel missions, developed by ESA for the Copernicus program, is generating large amounts of data. At the moment, this data is not easily integrated into processing chains outside the Copernicus ground segment. Very often, public and private institutions aiming to deliver end-user services based on EO data do not possess the computing power, the storage capacity or the software technology to cope with these new data flows. Handling an increasing volume of EO data is one of the main challenges for the community and hybrid clouds, coupled with big data management solutions and applications, are seen as a potentially efficient solution.

The design of such solutions needs the mix of expertise that can only be found in collaborations involving:

- *Data consumers*, to help defining and designing real added-value services to be integrated with the existing EO Exploitation Platforms
- *ICT experts and platforms operators*, with advanced knowledge on EO systems, to develop and provide hosting platforms for these services and to offer general solutions
- *E-Infrastructures*, to supply the computing and storage resources needed for data access and exploitation and provide the tools to manage the datasets in a distributed environment

The role of EGI

The main goal of EGI is to empower data- and compute-intensive research and innovation. By federating the capabilities of more than 350 publicly-funded resource providers, EGI has delivered unprecedented data analysis capabilities to more than 50,000 researchers from many disciplines and is now committed to bring innovation to the private sector.

Objective 1: Enable search and retrieval capabilities of Copernicus data and other reference EO datasets using e-Infrastructure services.

EGI offers the horizontal services and processes required to manage generic distributed infrastructures. Working on top of the EGI infrastructure, public and private entities willing to develop EO exploitation platform could focus on the end service without worrying about the complexities of the Copernicus data layer and underlying distributed environment. They could also benefit from adopting re-usable EGI components (e.g. single sign-on, accounting, etc.).
Furthermore, EGI will ensure the adoption of standard interfaces to make data and software reusable across services and domains, and portable across cloud infrastructures.
EGI services are being enhanced with capabilities to deal with Copernicus issues (vertical innovation). An example is the smart caching mechanisms extension of the Open Data Platform (ODP) [1].

Objective 2: Provide new e-Infrastructure data management capabilities to handle volume and variety

in big data processing and analysis for all research disciplines, crossing boundaries of infrastructure providers

EGI has an objective to enhance existing e-Infrastructure data management technologies with a focus on the interoperable EGI Open Data Platform [1].

Though the Open Data Platform, EGI can provide the capability of accessing and integrating huge volumes of different datasets in a single analysis task, thus facilitating data use and re-use by coupling data with analytic tools, software and the computing services required to create new knowledge.

Objective 3: Increase exploitation of EO data by public and commercial organizations through cloud services, e-Infrastructures

EGI plans to advertise its solutions for an easy development of EO data exploitation platforms to the public and private sectors. The integration of two ESA thematic exploitation platforms (TEPs), geohazard and hydrology, will demonstrate the benefits of working with the EGI infrastructure and will be advertised to other TEPs. The work on the geohazard exploitation platform (GEP) will allow reinforcing the partnership with the EPOS Research Infrastructure (RI) and EGI.

EGI services : an open data & computing infrastructure for Copernicus

Impact	*Services*
Seamless search and retrieval capabilities of Copernicus data and other reference EO datasets	Technology to federate distributed Copernicus data repositories geographically distributed and managed by different organisations: • service catalogue • accounting & monitoring • single sign-on: AAI Infrastructure
Data management capabilities to handle volume and variety in big data processing and analysis for all research disciplines	• Data management: registration and setup of EO data products into the EO metadata catalogue and related pre-staging through the smart cache in the EGI Open Data Platform
Standard interfaces to cross the borders	• OCCI standard interface • Data and software re-usable in various services and domains
A powerful execution environment co-located with Copernicus datasets	On-demand and massive processing: • standard OCCI interface to access a public cloud to run to run workflows, apps or bulk processing, • list of computing resources close to the pre-staged data • 325 resource providers, 650K CPU cores,~300 PB of disk and 200PB of tapes
Creation of custom (meta)data catalogues and virtual file-systems	Domain-specific metadata catalogues and smart caching mechanism to temporally replicate subsets of EO data hosted by various mirrors on the resource providers local storage

References

[1] *https://documents.egi.eu/document/2547*

[2] onedata.org

[3] www.openaire.eu