



Big Data challenges and solutions in building the Global Earth Observation System of Systems (GEOSS)

Paolo Mazzetti, Stefano Nativi, Mattia Santoro, and Enrico Boldrini
CNR, IIA, Monterotondo (RM), Italy (paolo.mazzetti@cnr.it)

The Group on Earth Observation (GEO) is a voluntary partnership of governments and international organizations launched in response to calls for action by the 2002 World Summit on Sustainable Development and by the G8 (Group of Eight) leading industrialized countries. These high-level meetings recognized 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. To this aim is constructing the Global Earth Observation System of Systems (GEOSS) on the basis of a 10-Year Implementation Plan for the period 2005 to 2015 when it will become operational.

As a large-scale integrated system handling large datasets as those provided by Earth Observation, GEOSS needs to face several challenges related to big data handling and big data infrastructures management. Referring to the traditional multiple Vs characteristics of Big Data (volume, variety, velocity, veracity and visualization) it is evident how most of them can be found in data handled by GEOSS. In particular, concerning Volume, Earth Observation already generates a large amount of data which can be estimated in the range of Petabytes (10¹⁵ bytes), with Exabytes (10¹⁸) already targeted. Moreover, the challenge is related not only to the data size, but also to the large amount of datasets (not necessarily having a big size) that systems need to manage. Variety is the other main challenge since datasets coming from different sensors, processed for different use-cases are published with highly heterogeneous metadata and data models, through different service interfaces. Innovative multidisciplinary applications need to access and use those datasets in a harmonized way. Moreover Earth Observation data are growing in size and variety at an exceptionally fast rate and new technologies and applications, including crowdsourcing, will even increase data volume and variety in the next future.

The current implementation of GEOSS already addresses several big data challenges. In particular, the brokered architecture adopted in the GEOSS Common Infrastructure with the deployment of the GEO DAB (Discovery and Access Broker) allows to connect more than 20 big EO infrastructures while keeping them autonomous as required by their own mandate and governance. They make more than 60 million of unique resources discoverable and accessible through the GEO Portal. Through the GEO DAB, users are able to seamlessly discover resources provided by different infrastructures, and access them in a harmonized way, collecting datasets from different sources on a Common Environment (same coordinate reference system, spatial subset, format, etc.).

Through the GEONETCast system, GEOSS is also providing a solution related to the Velocity challenge, for delivering EO resources to developing countries with low bandwidth connections.

Several researches addressing other Big data Vs challenges in GEOSS are on-going, including quality representation for Veracity (as in the FP7 GeoViQua project), brokering big data analytics platforms for Velocity, and support of other EO resources for Variety (such as modelling resources in the Model Web).