Contribution ID: **108** Type: **not specified**

EOSC support to transparent access to Copernicus Sentinel image data for wider uptake in science

Monday, 2 November 2020 14:45 (15 minutes)

Earth Observation (EO) is a relatively new domain on the European Open Science Cloud. EO data access has long suffered from restrictive licenses and opaque and proprietary distribution systems, which has, by an large, hindered wide uptake in science, in particular beyond the traditional remote sensing and geospatial analysis disciplines. Massive new EO data streams which are distributed under a full, free and open license include those from the European Copernicus program's Sentinel sensors since 2014 and US Landsat since 2010. Currently multiple Petabytes of high resolution Sentinel-1 (SAR) and -2 (optical) sensor data are available for thematic research and monitoring applications in maritime and land science disciplines.
Still, even with open licenses, EO data access remains complex and combining such data with geospatial reference data for targeted analysis is hard for novice users. Extensive knowledge of sensor-specific data organization, map projections and formats is often required. Some data, for instance Sentinel-1, requires complex processing to create "analysis ready"data sets. Similarly, feature data sets suffer from a plethora of, outdated, data formats that only (still) exist due to long deprecated proprietary solutions. The old paradigm in EO data analysis was that 80% of a researcher's time was spent on data pre-processing and preparation, and 20% on analysis. This radically changed with the introduction of Google Earth Engine (GEE), Google's cloud infrastructure that hosts complete sensor data collections closely coupled to its massive parallel processing capacity. By abstracting data access and integrating ever more sophisticated analysis methods in its library of geospatial analysis routines, science users are able to compose their analysis in scripts that can be executed interactively or in batch. With that, the paradigm is more than inverted, i.e. 95% of research time is now spent on programming and testing the analytical logic that underlies scalable and reproducible science methods.

In this demonstration, we show how EOSC resources can be used to emulate some of GEE's functionalities. We have developed this in an Early Adaptor Project, which inherits developments that were originally implemented on Copernicus DIAS, which is European cloud infrastructure that is closely coupled to Sentinel data archives. One of the DIAS instances (CloudFerro) is federated in EOSC. Amongst others, we demonstrate hybrid solutions that combine pre-extracted time series from PostgreSQL/Postgis with direct access to image subsets. The time series can be analyzed and visualized in Jupyter Notebooks or client side python, including machine learning routines, through the use of RESTful services. Image extracts are used both in pre-configured visualization and for full resolution higher level image processing (e.g. segmentation, structural analysis). Our development further provides pointers on how optimized data formats, smart caching and prediction and support to "on the fly"processing may further advance DIAS utility in this domain, and achieve the overall goal to "take space data out of the space domain", and integrate deeper in applied science.

Presenter: LEMOINE, Guido (European Commission, Joint Research Centre)

Session Classification: Demos 3