Analyzing large-scale Earth Observation data repositories made simple with OpenEO Platform

Edzer Pebesma¹, Patrick Griffiths², Christian Briese³, Alexander Jacob⁴, Anze Skerleva⁵, Jeroen Dries⁶, Gilberto Camara⁷, and Matthias Mohr¹

¹University of Münster, Institute for Geoinformatics, Germany (edzer.pebesma@uni-muenster.de)
²ESA
³EODC
⁴EURAC
⁵Sinergise
⁶VITO
⁷GEO

The OpenEO API allows the analysis of large amounts of Earth Observation data using a high-level abstraction of data and processes. Rather than focusing on the management of virtual machines and millions of imagery files, it allows to create jobs that take a spatio-temporal section of an image collection (such as Sentinel L2A), and treat it as a data cube. Processes iterate or aggregate over pixels, spatial areas, spectral bands, or time series, while working at arbitrary spatial resolution. This pattern, pioneered by Google Earth Engine™ (GEE), lets the user focus on the science rather than on data management.

The openEO H2020 project (2017-2020) has developed the API as well as an ecosystem of software around it, including clients (JavaScript, Python, R, QGIS, browser-based), back-ends that translate API calls into existing image analysis or GIS software or services (for Sentinel Hub, WCPS, Open Data Cube, GRASS GIS, GeoTrellis/GeoPySpark, and GEE) as well as a hub that allows querying and searching openEO providers for their capabilities and datasets. The project demonstrated this software in a number of use cases, where identical processing instructions were sent to different implementations, allowing comparison of returned results.

A follow-up, ESA-funded project “openEO Platform” realizes the API and progresses the software ecosystem into operational services and applications that are accessible to everyone, that involve federated deployment (using the clouds managed by EODC, Terrascope, CreoDIAS and EuroDataCube), that will provide payment models (“pay per compute job”) conceived and implemented following the user community needs and that will use the EOSC (European Open Science Cloud) marketplace for dissemination and authentication. A wide range of large-scale cases studies will demonstrate the ability of the openEO Platform to scale to large data volumes. The case studies to be addressed include on-demand ARD generation for SAR and multi-spectral data, agricultural demonstrators like crop type and condition monitoring, forestry services like
near real time forest damage assessment as well as canopy cover mapping, environmental hazard
monitoring of floods and air pollution as well as security applications in terms of vessel detection
in the mediterranean sea.

While the landscape of cloud-based EO platforms and services has matured and diversified over
the past decade, we believe there are strong advantages for scientists and government agencies to
adopt the openEO approach. Beyond the absence of vendor/platform lock-in or EULA’s we
mention the abilities to (i) run arbitrary user code (e.g. written in R or Python) close to the data, (ii)
carry out scientific computations on an entirely open source software stack, (iii) integrate different
platforms (e.g., different cloud providers offering different datasets), and (iv) help create and
extend this software ecosystem. openEO uses the OpenAPI standard, aligns with modern OGC API
standards, and uses the STAC (SpatioTemporal Asset Catalog) to describe image collections and
image tiles.