EGU25-4155, updated on 02 Jul 2025
https://doi.org/10.5194/egusphere-egu25-4155
EGU General Assembly 2025

How open software, data and platforms are transforming Earth observation data science

Wolfgang Wagner[1,2], Matthias Schramm[1], Martin Schobben[1], Christoph Reimer[2], and Christian Briese[2]

[1]TU Wien, Department of Geodesy and Geoinformation, Vienna, Austria (wolfgang.wagner@geo.tuwien.ac.at)
[2]EODC Earth Observation Data Centre, Vienna, Austrua

One of the most time-consuming and cumbersome tasks in Earth observation data science is finding, accessing and pre-processing geoscientific data generated by satellites, ground-based networks, and Earth system models. While the much increased availability of free and open Earth observation datasets has made this task easier in principle, scientific standards have evolved according to data availability, now emphasizing research that integrates multiple data sources, analyses longer time series, and covers larger study areas. As a result of this “rebound effect”, scientists and students may find themselves spending even more of their time on data handling and management than in the past. Fortunately, cloud platform services such as Google Earth Engine can save significant time and effort. However, until recently, there were no standardized methods for users to interact with these platforms, meaning that code written for one service could not easily be transferred to another (Schramm et al., 2021). This created a dilemma for many geoscientists: should they use proprietary cloud platforms to save time and resources at the risk of lock-in effects, or rely on publicly-funded collaborative scientific infrastructures, which require more effort for data handling? In this contribution, we argue that this dilemma is about to become obsolete thanks to rapid advancements in open source tools that allow building open, reproducible, and scalable workflows. These tools facilitate access to and integration of data from various platforms and data spaces, paving the way for the “Web of FAIR data and services” as envisioned by the European Open Science Cloud (Burgelman, 2021). We will illustrate this through distributed workflows that connect Austrian infrastructures with European platforms like the Copernicus Data Space Ecosystem and the DestinE Data Lake (Wagner et al., 2023). These workflows can be built using Pangeo-supported software libraries such as Dask, Jupyter, Xarray, or Zarr (Reimer et al., 2023). Beyond advancing scientific research, these workflows are also valuable assets for university education and training. For instance, at TU Wien, Jupyter notebooks are increasingly used in exercises involving Earth observation and climate data, and as templates for student projects and theses. Building on these educational resources, we are working on an Earth Observation Data Science Cookbook to be published on the Project Pythia website, a hub for education and training in the geoscientific Python community.

References

Burgelman (2021) Politics and Open Science: How the European Open Science Cloud Became Reality (the Untold Story). Data Intelligence 3, 5–19. https://doi.org/10.1162/dint_a_00069

Reimer et al. (2023) Multi-cloud processing with Dask: Demonstrating the capabilities of DestinE Data Lake (DEDL), Conference on Big Data from Space (BiDS'23), Vienna, Austria. https://doi.org/0.2760/46796

Schramm et al. (2021) The openEO API–Harmonising the Use of Earth Observation Cloud Services Using Virtual Data Cube Functionalities. Remote Sensing 13, 1125. https://doi.org/10.3390/rs13061125

Wagner et al. (2023) Federating scientific infrastructure and services for cross-domain applications of Earth observation and climate data, Conference on Big Data from Space (BiDS'23), Vienna, Austria. https://doi.org/10.34726/5309