

Interoperable Workflow Efficiency: Exploring the Integration of OpenEO, CWL, and EOEPKA for Seamless Data Processing and Modeling

Tuesday, 1 October 2024 15:55 (15 minutes)

Contemporary HPC and cloud-based data processing is based on complex workflows requiring close access to large amounts of data. OpenEO process graphs allow users to access data collections and create complex processing chains. Currently OpenEO can be accessed via one of the clients in JavaScript, R, or Python. Direct access to data is provided via Spatio Temporal Asset Catalogues (STAC). As part of our ongoing research under the InterTwin project, the focus is on extending the capabilities of OpenEO to support the management and execution of OGC Application Packages.

An Application Package allows users to create automated, scalable, reusable and portable workflows. It does so by creating, for example, a Docker Image containing all of the application code and dependencies. The workflow is described with Common Workflow Language (CWL). The CWL document references all of the inputs, outputs, steps, and environmental configurations to automate the execution of an application.

The execution is handled by an Application Deployment and Execution Service (ADES) coming from the EOEPKA project. It is a Kubernetes based processing service capable of executing Application Packages via OGC WPS1.0, 2.0 OWS service, and the OGC API-processes. In many ways, the core goals and objectives between EOEPKA and the InterTwin project align well. The focus is on allowing workflows to be seamlessly executed without the need of substantial code rewrites or adaptations to a specific platform.

OpenEO is on its way to become an OGC community standard. It currently supports a large set of well-defined cloud optimized processes that allow users to preprocess and process data directly in the cloud. The goal is to integrate the Application Deployment and Execution Service (ADES) from the Earth Observation Exploitation Platform Common Architecture (EOEPKA) project to create a fusion between OpenEO process graphs and Application Packages.

Application Package support in OpenEO is a means of providing users the ability to bring their applications directly to the platform. Instead of having to reimplement the code in a process graph, it is possible to wrap any existing application. The fusion of process graphs and CWL based application workflows extends OpenEO for users that would like to perform testing of their models, ensemble models etc. while utilizing the same process graph and direct access to data.

Many complex workflows require some kind of data preprocessing. This preprocessing can be done using OpenEO process graphs and then be directly sent for execution to an Application Package to run the actual process. OpenEO complements STAC by providing a standardized interface and processing framework for accessing and analyzing Earth observation data. Having all of the data and tools readily available on a single platform creates an accessible, interoperable, and a reproducible environment for users to create efficient workflows.

The ability to create standardized, reusable workflows using CWL and execute them on distributed computing resources via ADES can significantly reduce the time and effort required for data processing tasks. Researchers can focus on algorithm development and data analysis rather than worrying about infrastructure management or software compatibility issues.

Topic

Needs and solutions in scientific computing: Digital Twins

Primary authors: JACOB, Alexander (Eurac Research - ACCADEMIA EUROPEA DI BOLZANO); ZVOLENSKY, Juraj (Eurac); CLAUS, Michele (EURAC); CAMPALANI, Piero (Eurac Research); FERRARIO, iacopo (EURAC Reseach)

Presenters: JACOB, Alexander (Eurac Research - ACCADEMIA EUROPEA DI BOLZANO); ZVOLENSKY, Juraj (Eurac)

Session Classification: Cloud Compute federation and national initiatives