



Contribution ID: 22

Type: Poster

Refactoring and isolation data pipelines through the use of software containerization and continuous integration

At the IAS-8 institute of Forschungszentrum Jülich, the accurate and complete collection of measurement and environmental data is essential for subsequent analyses and modeling in many projects. Although the Bayeos server (<https://github.com/BayCEER/bayeos-server>) used at FZJ provides an open and standardized data platform for such data, the import and transformation of data from different sources is often difficult in terms of provision, traceability and subsequent adjustments. To address this problem, a flexible import and transformation pipeline for time series data was developed based on Python and a PostgreSQL-based integration database. There is a clean separation of import, transformation and aggregation processes, which also allows for easy customization. Each individual step of the defined pipeline runs as a container in a Docker environment. There is a template for a basic pipeline, which can be easily customized to define additional pre- and post-processing steps. This template has been successfully adapted for different existing data pipelines. Once this has been done, the containers are built automatically using the CI/CD pipeline of the DevOps platform Gitlab. In addition, Gitlab's own container registry ensures easy deployment and updating of the pipeline elements.

Slot length

Primary author: Mr BRUNS, Benjamin (Forschungszentrum Jülich GmbH)

Presenter: Mr BRUNS, Benjamin (Forschungszentrum Jülich GmbH)

Session Classification: Poster Session

Track Classification: Research Software (legacy): Computational Workflows