Contribution ID: **34**   Type: **Extended Presentation (25' + 5' for questions)**

Automating scientific dataset management and processing using Onedata

Wednesday 12 October 2022 17:30 (30 minutes)

Making experiment results FAIR is a known challenge. To cope with this task, we develop a system that provides an easy way to make data produced by specialized devices (such as cryo-em microscopes) available to the scientific community. We focused on making the system as easy as possible for data producers and also for users who use the datasets in their scientific computations. Our solution can be used to manage the storage of experimental data between several tiers of data storage. Beginning physical data storage of the experimental facilities where the data originates, national or scientific domain data storage services, and fast storage in computing facilities provided on both national and European levels.

The software is built on top of the global data access solution for science Onedata. Our software supports the whole process, from acquiring produced data from the device, setting up automatically all necessary Onedata parameters (access policy, metadata, ⋯), publishing the datasets, and archiving them in permanent storage. It implements varying policies of handling the data, e.g., expiration at the acquisition facility, archiving in multiple copies, and data publication after an embargo period. It can also export datasets to supported repositories or metadata to metadata catalogs. The life cycle of the data is defined in a YAML file which is attached to the dataset.

Primary author: SVOBODA, Tomáš (CESNET)

Presenter: SVOBODA, Tomáš (CESNET)

Session Classification: IBERGRID Contributions

Track Classification: Enabling and fostering Open Science adoption in EOSC