

Scientific dataset management system for the research institute based on Onedata

Tuesday, 1 October 2024 18:00 (1 hour)

As the volume and complexity of scientific data continue to grow, the efficient management of data across its entire lifecycle has become paramount. In this context, we have decided to create a system for CEITEC Research Institute, which would allow emerging data sets to be registered and managed, using the existing Onedata system as the data layer.

At its core, Onedata oversees the entire data lifecycle, commencing with the acquisition of data from various connected instruments (cryo-EM, NMR, light microscopy) at the moment of data generation. The automated processes employed by the system enable the organisation of acquired data into coherent datasets, enriched with metadata harvested directly from the instruments themselves and the execution of workflows designed to generate data-aware metadata annotations where feasible, in accordance with defined metadata schemas established in specific fields. This facilitates the creation of FAIR datasets which are ready for publication in thematic data repositories, as and when required.

The ability to integrate heterogeneous storage capacity with heterogeneous high-performance computing (HPC) platforms, such as Jupyter notebooks and Kubernetes container clouds, is a significant advantage. By facilitating the connection between storage capacity and direct access to compute resources, Onedata enables access to compute resources for data analysis, thereby accelerating scientific discovery.

Finally, the ability to share live data via Onedata enables data sharing within and beyond the research group. Once the analysis has been completed, the system is prepared to allow scientists to easily complete and publish the final dataset to the thematic data repositories.

The objective of this poster is to illustrate the development of tools that will facilitate and streamline data sharing among scientific communities at the national and international levels. These tools are intended to support the principles of FAIR and Open Science.

Topic

Data innovations: Data Management/Integration/Exchange

Primary authors: Mr SVOBODA, Tomáš (Masaryk University); Mr ROSINEC, Adrian (Masaryk University); Dr RACEK, Tomas (Masaryk University); Mr HANDL, Josef (Masaryk University); KRENEK, Ales (CES-NET); Mrs SVOBODOVA, Radka (Masaryk University)

Presenters: Mr SVOBODA, Tomáš (Masaryk University); Mr ROSINEC, Adrian (Masaryk University)

Session Classification: Demonstrations & Posters