Contribution ID: **11**   Type: **Presentation (15' + 5' for questions)**

Open Data for DESY, HIFIS, NFDI and EOSC

Tuesday, 29 October 2024 16:10 (20 minutes)

DESY, one of Europe's leading synchrotron facilities, is active in various scientific fields, including High Energy and Astro particle Physics, Dark matter research, Physics with Photons, and Structural Biology. These fields generate large amounts of data, which are managed according to specific policies that respect restrictions on ownership, licenses, and embargo periods. Currently there is a move to make this data publicly available, as requested by funding agencies and scientific journals. To facilitate this, DESY IT is developing solutions to publish Open Data sets, making them easily discoverable, accessible, and reusable for the wider scientific community, particularly those not supported by large e-infrastructures.

In line with Open and FAIR data principles, DESY's Open Data solution will provide a metadata catalog to enhance discoverability. Access will be granted through federated user accounts via eduGAIN, HelmholtzID, NFDI, and later EOSC-AAI, enabling community members to access data using their institutional accounts. Interoperability will be ensured by using commonly accepted data formats such as HDF5, specifically NeXuS, openPMD and ORSO. Providing technical and scientific metadata will make the Open Data sets reusable for further analysis and research. The blueprint for DESY's Open Data solution will be shared through HIFIS and the wider community upon successful evaluation.

The initial prototype will consist of three interconnected solutions: the metadata catalog SciCat, the storage system dCache, and the VISA (Virtual Infrastructure for Scientific Analysis) portal. Scientific data, along with its metadata, will be stored in a specific directory on dCache and ingested into SciCat which provides direct access and download options. To ensure harmonization of scientific metadata among similar experiments, experiment-specific metadata schemata will be created for metadata validation before ingestion. A subset of technical and scientific metadata will be integrated into the VISA portal, allowing scientists to access datasets within it. The VISA portal allows to create computing environments with pre-installed analysis tools and mounted datasets, providing easy access to data and tools.

During the presentation, the system's architecture, its components, and their interactions will be discussed, focusing on the harmonization of metadata schemata and the roadmap for developing tooling and processes for ingestion and validation of metadata.

Primary authors: MILLAR, Paul (DESY); WETZEL, Tim (DESY IT (Research and Innovation in Scientific Computing)); Dr JANDT, Uwe (DESY)

Co-authors: Dr REPPIN, Johannes (DESY); Dr KOBUS, Julia (CAU Kiel); Dr PITHAN, Linus (DESY); FUHRMANN, Patrick (DESY)

Presenter: WETZEL, Tim (DESY IT (Research and Innovation in Scientific Computing))

Session Classification: IBERGRID

Track Classification: Enabling and fostering Open Science adoption