Research data management for DETECT: Storage and data processing in the HPC environment of JSC

Within the Collaborative Research Centre 1502, DETECT, large amounts of research data from various sources are being produced and shared between the CRC partners and to the outside world. These sources comprise model input and output data, observational data from satellites and large networks as well as economical and statistical information affecting land use and land cover developments. Reanalysis and ensemble simulation results from the regional climate model TerrSysMP will reach a data volume in the order of Petabytes, which calls for HPC oriented storage and workflow strategies to enable effective data analysis and sharing.

Jülich Supercomputing Center (JSC) at the Research Center Jülich hosts one of Europe's largest supercomputing systems, JUWELS, providing high-performance computing power as well as high-performance and high-capacity storage resources from its versatile storage infrastructure JUST. The Service Project Z03 of DETECT set up a data project at JSC, which can hold the data amounts to be produced within the CRC and which provides the necessary infrastructure for HPC related workflows and long-term archiving. Special care, already in the planning phase, needs to be taken of data formats, metadata, and general structure of the repository in order to allow for FAIR data handling and effective data processing. netCDF is favored as data format for geoscientific data, as it already comes with established metadata standards and elaborated tools for modification, analysis, and visualization. In addition, the use of data cubes for faster extraction of sub-datasets will be explored. Alternatively to classical HPC access with ssh, JSC now also offers datalad, a distributed data management system. Datalad opens up an easy way to access JUST data from your remote Computer, allowing to handle management of large datasets and to select user-defined files and datasets for up- and download.

Primary authors: Dr STEIN, Olaf (Forschungszentrum Jülich);   NIKFAL, Amirhossein (Forschungszentrum Jülich)

Presenter: Dr STEIN, Olaf (Forschungszentrum Jülich)

Session Classification: Posters