Abstract: Contribution to NFDI4Earth

The National Research Data Infrastructure (NFDI) is intended to systematically develop, sustainably secure and make accessible the data holdings of science and research and to network them (inter)nationally. It will be established in a process driven by the scientific community as a networked structure of consortia acting on their own initiative.
DKRZ is co-applicant of NFDI4Earth (https://www.nfdi4earth.de/), whose funding will begin in October 2021 and will initially run for 5 years.

NFDI4Earth supports 14 small flexible pilot projects (1-year projects) to integrate a broad community and innovative developments into the NFDI4Earth. The pilots cover a broad range of topics (technical implementations, new methods, new standards and FAIRness, reproducibility, interoperability, and others).

A) *Observations closer to Model Data (OcMOD)*
Within this NFDI4Earth initiative, DKRZ together with the German Weather Service holds a one-year pilot project to bring "Observations closer to Model Data (OcMOD)".
Background: Researchers have to obtain model data and commonly used observational data from different sources: model data from e.g. the DKRZ and observational data from e.g. servers from public authorities. All data have to be prepared to be in the same formats and standards before it can be used for further analysis. A broad range of users (e.g. climate modelers, impact modelers, climate scientists, providers for climate services and education) are dealing with same the effort and troubles with the same sets of data.
The aim of the project is to bring observational data close to the model output, to provide easy access data to from public authorities and to increase the number of users of various disciplines, and to provide this data in standardized formats for easy usage. As a pilot for this project we will integrate the DWD reanalysis dataset COSMO-REA6 (Kaspar et al. 2020) to the infrastructure of ESGF (Earth System Grid Federation - https://esgf.llnl.gov/) and publish it at the WDCC. The challenge is not only integrating the observational data into ESGF but also to create a common workflow together with the public authority DWD.

B) *Enhancing Earth system model evaluation with data cube enabled machine learning*
Another very interesting pilot is focusing on enabling the use of cloud ready data and machine learning methods (here we focus on causal discovery) for routine Earth system model evaluation. The pilot also takes advantage of already available data from Earth system models and Earth observational data hosted at the DKRZ. The result of the pilot will be the release of an enhanced ESMValTool version that includes the data cube concept and innovative machine learning diagnostics. The pilot will contribute the infrastructure for a model evaluation approach with innovative ML methods included accessible through the NFDI4Earth one-stop shop. In addition, the enhancement of the existing platforms ESDL and ESMValTool can be used to further engage the NFDI4Earth community.

This application requests resources for pilot projects of NFDI4Earth. Two Pilots are presented here as examples, but the resources are used to carry out all pilots that require DKRZ resources.