



A Hybrid Evaluation System Framework (Shell & Web) with Standardized Access to Climate Model Data and Verification Tools for a Clear Climate Science Infrastructure on Big Data High Performance Computers

Christopher Kadow (1), Sebastian Illing (1), Oliver Kunst (2), Uwe Ulbrich (1), and Ulrich Cubasch (1)

(1) Freie Universität Berlin, Institute of Meteorology, Berlin, Germany (christopher.kadow@met.fu-berlin.de), (2) Zuse Institute Berlin, Berlin, Germany

The project 'Integrated Data and Evaluation System for Decadal Scale Prediction' (INTEGRATION) as part of the German decadal prediction project MiKlip develops a central evaluation system. The fully operational hybrid features a HPC shell access and an user friendly web-interface. It employs one common system with a variety of verification tools and validation data from different projects in- and outside of MiKlip. The evaluation system is located at the German Climate Computing Centre (DKRZ) and has direct access to the bulk of its ESGF node including millions of climate model data sets, e.g. from CMIP5 and CORDEX. The database is organized by the international CMOR standard using the meta information of the self-describing model, reanalysis and observational data sets. Apache Solr is used for indexing the different data projects into one common search environment. This implemented meta data system with its advanced but easy to handle search tool supports users, developers and their tools to retrieve the required information. A generic application programming interface (API) allows scientific developers to connect their analysis tools with the evaluation system independently of the programming language used. Users of the evaluation techniques benefit from the common interface of the evaluation system without any need to understand the different scripting languages. Facilitating the provision and usage of tools and climate data increases automatically the number of scientists working with the data sets and identify discrepancies. Additionally, the history and configuration sub-system stores every analysis performed with the evaluation system in a MySQL database. Configurations and results of the tools can be shared among scientists via shell or web-system. Therefore, plugged-in tools gain automatically from transparency and reproducibility. Furthermore, when configurations match while starting a evaluation tool, the system suggests to use results already produced by other users—saving CPU time, I/O and disk space. This study presents the different techniques and advantages of such a hybrid evaluation system making use of a Big Data HPC in climate science. website: www-miklip.dkrz.de visitor-login: click on "Guest"