Geophysical Research Abstracts
Vol. 20, EGU2018-10685, 2018
EGU General Assembly 2018

Earth System Simulation Toolkit

Dirk Fleischer (1), Thilo Paul-Stueve (2), Florian Lier (3), and Sven Wachsmuth (3)

(1) Christian-Albrechts University, Kiel Marine Science, Data and Information Services, Kiel, Germany (dfleischer@kms.uni-kiel.de), (2) Christian-Albrechts University, Kiel, Germany, (3) Excellenc Cluster Cognitive Interaktion Technologie (CITEC) University Bielefeld, D-33501 Bielefeld

Climate and oceanographic modelling relies on massive computational resources and may produce a vast amount of data in a short period of time. In this context, the relationship between model source code, model parameters, and model output are essential for the interpretation, comparison, evaluation and eventually for the reproducibility of model results. High Performance Cluster (HPC) operator experience indicates that researchers tend to manage and handle model source code, executables, and input parameters, and the corresponding output individually different.

Hence, on the one hand, a lack of documentation and modularisation, as well as intransparent relationships between research artefacts often hinder traceability and — more importantly — comprehensibility and reproducibility. On the other hand, the central HPC services have a hard time to provide all researchers with the same quality of support required to maintain a state-of-the-art reproducibility track of their simulations. Learning from the robotics community, this reproducibility track is essential for the continuous development and advancement of complex systems, and ultimately required for publications complying with the FAIR Data Principles.

The data management of Kiel Marine Science and the central data management of Kiel University (CAU) together with the Cluster of Excellence – Cognitive Interaction Technology (CITEC, Bielefeld University) evaluate synergy effects at the CAU HPC environment to establish and maintain a strong connection between software (model) code, parameters, and the output in the storage infrastructure. CITEC develops the Cognitive Interaction Toolkit (CITK), a software tool chain embedded in a newly proposed development process, to target reproducible research in robotics; in an initial workshop, we exemplarily set up the CITK at the CAU HPC.

We will present results from this workshop, in particular the deployment at the CAU HPC, the required customisation to the climate modelling domain, and the lessons learned from adaptation of earth simulations to state-of-the-art computer science procedures.