



Contribution ID: 6

Type: **not specified**

Big Data Research at DKRZ

Tuesday 7 October 2014 14:30 (30 minutes)

DKRZ is running one the largest climate data archives. In August 2014 the mass storage archive (HPSS) contains about 36 PB. This includes the long-term archive Word Data Center Climate (WDCC) with more than 4 PB climate model reference data. Emphasis at DKRZ is on production, analysis, storage, curation and dissemination of climate model data and related observations. At the current HPC system at DKRZ (HLRE-2) annual growth rates are observed of 8 PB for HPSS and 0.5 PB up to 1.5 PB for WDCC. The next generation of the HPC system (HLRE-3) implies the estimation of annual data growth rates of 75 PB for HPSS and 8 PB for WDCC. Important for HLRE-3 will a seamless end-to-end workflow which covers all steps in the data life cycle. This means that not only services around data processing and data storage have to be optimized but also parallelization of climate model code and improvement of I/O processes. The optimization of the end-to-end workflow is key to make optimal use existing HPC resources. The aim is not to produce pure numbers but to generate climate information.

Presenter: LAUTENSCHLAGER, Michael