Towards increasing the reusability of atmospheric model data: adapting metadata standards and introducing quality criteria

Karsten Peters

Deutsches Klimarechenzentrum (DKRZ), Hamburg, Germany

Daniel Neumann

Deutsches Klimarechenzentrum (DKRZ), Hamburg, Germany

Hannes Thiemann

Deutsches Klimarechenzentrum (DKRZ), Hamburg, Germany

Abstract

We present current and future work at the German Climate Computing Center (DKRZ) aimed at increasing the long-term reusability of atmospheric and climate model data. Specifically the recently funded project AtMoDat (project start: June 2019, duration of 3 years) will be in the focus of the presentation.

DKRZ plays a central role for the German and international Earth System Science (ESS) community by hosting the IPCC (Intergovernmental Panel on Climate Change) Data Distribution Center (IPCC DDC) Reference Data Archive for global climate model output (World Data Center for Climate, WDCC) as well as a Tier 1 data node of the ESGF (Earth System Grid Federation, an infrastructure enabling the global dissemination of large-volume climate model data). Further, DKRZ offers a variety of data management services covering all aspects of the research data life cycle.

In particular, DKRZ services play a vital role for the long-term curation and reusability of numerical model output gathered in the context of large and internationally coordinated coupled model intercomparison projects (CMIPs). In order for CMIP to work, data and metadata standards as well as long-term data curation mechanisms have been thoroughly defined and are continuously adapted to allow for effective data sharing, reuse and preservation of the scientific record. Once a new CMIP-standard is established by the international community in an iterative process, the post-processing of numerical model output is often very cumbersome because i) the process behind the definition of the standard appears intransparent, ii) the standard is elaborate and requires meticulous attention to details, e.g. in the creation of metadata and iii) CMIP-compliance of model output is often not taken into account from the beginning of data production. Once the model output data fullfill the standard, their further processing and usage is simplified.

This high degree of standardisation and long-term reusability of climate model data is desirable, but the application of the CMIP-standard does not lend itself to most sub-communities of atmospheric science and smaller model intercomparison projects.

The project AtMoDat (www.atmodat.de), lead by DKRZ, builds on the existing expertise at DKRZ and its project partners to focus on increasing the reusability of atmospheric model data. Data standards as well as curation criteria are developed for meteorological sub-communities for which i) the application of the full CMIP standard is not feasible and ii) no agreed data standards for ensuring (inter-)disciplinary data reuse even exist. Specific use-cases are the urban-climate science community and small-scale climate model intercomparison projects. The developments in AtMoDat start from the full CMIP-standard, which is adapted to meet the demands of these application domains. Further, AtMoDat will develop and test the application of atmospheric model data quality criteria to perform an „AtMoDat DOI-Branding" by developing a specific adaptation of the DataCite metadata model for atmospheric model data.

First results of AtMoDat will be illustrated in the presentation. AtMoDat is funded by the German Federal Ministry for Education and Research (BMBF).