Geophysical Research Abstracts
Vol. 14, EGU2012-2019, 2012
EGU General Assembly 2012

Federated Quality Control Procedure for CMIP5 / IPCC-AR5 Data

M. Stockhause, H. Höck, M. Kurtz, M. Lautenschlager, and F. Toussaint
WDC Climate / DKRZ, Hamburg, Germany (stockhause@dkrz.de)

The International Panel on Climate Change (IPCC) aims to advance the knowledge of climate change and climate variability. The results collected within the Climate Model Intercomparison Project No. 5 (CMIP5) are intended to underlie the coming fifth assessment report (IPCC-AR5). In comparison with the CMIP3 (IPCC-AR4) three main improvements have been implemented in the data infrastructure:

- Decentral data storage on local data nodes of the Earth System Grid (ESG, http://pcmdi3.llnl.gov/esgcet) and replication of the most important data (relevant for IPCC-AR5) among the three primary CMIP5 archive centers, PCMDI (Program for Climate Model Diagnosis and Intercomparison), BADC (British Atmospheric Data Centre), and WDCC (World Data Center for Climate) at DKRZ.
- Detailed descriptions of numerical climate models and the simulations using the CIM (Common Information Model) developed by METAFOR (http://metaforclimate.eu).
- Data curation was improved by introducing a versioning concept and a quality assessment process providing a uniform identification of datasets as well as a persistent identifier DOI (Digital Object Identifier) for data citation in scientific publications (http://cmip5qc.wdc-climate.de).

The quality control (QC) concept was developed on the background of the existing federated data infrastructure of the ESG and the external metadata source. For this reason as well as for sharing the work load of the quality checks a federated / distributed quality control procedure was developed, consisting of:

- a QC repository for QC result and information storage and exchange within the QC process,
- a QC checker tool, and
- a QC service package to support QC repository storage, QC result analyses, QC information access for QC managers as well as for data users.

For CMIP5 the quality control procedure consists of three quality levels. With increasing quality level the checks are performed more centralized:

QC level 1 Separate technical QC checks on data (CMOR2, ESG conformance) and metadata (CIM conformance) are performed, locally at the data nodes and the CIM repository.

QC level 2 Consistency checks on data to meet project standards are carried out under supervision of the three primary CMIP5 archive centers.

QC level 3 DOI data publications of CMIP5 simulations are performed by the DOI publication agency WDCC together with the scientific data author according to DataCite regulations (http://datacite.org).

In this contribution the federated concept of the QC procedure, its pilot implementation for CMIP5, and first experiences within CMIP5 are presented.