Geophysical Research Abstracts
Vol. 13, EGU2011-2859-1, 2011
EGU General Assembly 2011
© Author(s) 2011

Data Publication and Quality Control Procedure for CMIP5 / IPCC-AR5 Data

Martina Stockhause, Michael Lautenschlager, Heinke Höck, and Frank Toussaint
WDC Climate / DKRZ, Hamburg, Germany (martina.stockhause@zmaw.de)

Climate data and especially data of climate projections of Earth System Models play a key role for the investigation of climate change impacts and for the derivation of mitigation and adaptation strategies. The user community grows from the original climate research community into a heterogeneous climate community with diverse requirements for data and data related services.

The next assessment report of the Intergovernmental Panel on Climate Change (IPCC-AR5) will provide a large dataset of high quality for earth system and impact researchers as well as for decision makers. The climate projections are defined within CMIP5 (Coupled Model Intercomparison Project Phase 5, cmip-pcmdi.llnl.gov/cmip5/). Data is served via the distributed data infrastructure of the Earth System Grid (ESG, www.earthsystemgrid.org). Thus data has to be described by metadata and fulfill reliable and uniform quality standards.

Data passes different types of Quality Controls (QC) connected with different access policies. Three different QC data levels are distinguished:

- QC level 1: CMOR2, ESG conformance of data and CIM conformance of metadata,
- QC level 2: WDCC conformance and subjective controls, and
- QC level 3: STD-DOI data publication.

For QC level 1 the completeness of metadata and the well-formattedness of data are checked. Data assigned QC level 1 is visible for all. The access is restricted to the data owner and specific scientists announced by him.

Within the quality checks for level 2 metadata and data are checked still separately. Metadata is controlled by a scientist for plausibility and consistency. The expense of the data checks is shared between the three ESG Federation partners: PCMDI (www-pcmdi.llnl.gov), BADC (British Atmospheric Data Centre, badc.nerc.ac.uk), and WDCC (World Data Center Climate, wdc-climate.de). The Quality Assurance (QA) procedures includes consistency between data header and data and of statistical quantities against pre-defined ranges. Data of QC level 2 is accessible under the risk of possible data changes.

The QC procedure of level 3 follows the rules of STD-DOI data publication (www.std-doi.de), based on which an international regulation is currently derived within the DataCite consortium (datacite.org). Data checks consist of double-checks (using results of performed QCs level 1 and 2) and cross-checks between data and metadata. In addition to this technical QA (TQA) supervised by the WDCC, the scientific QA (SQA) is performed by the data author. After receiving the final author approval for the content of the core metadata and the description of the SQA procedures, the data is published, i.e. assigned the persistent identifier DOI. Data of QC level 3 / assigned a DOI is no longer a matter of change and accessible to all registered data users. Data is then uniquely citable in scientific publications by its DOI.

The granularity of the STD-DOI data publication has to be suitable for citation within model comparison studies. This is the experiment, i.e. the aggregation of all data belonging to all ensemble runs performed for a specific climate projection. Optionally, a peer-reviewed publication in a data journal like ESSD (Earth System Science Data, www.earth-system-science-data.net/) or alike for a larger data aggregation (coarser DOI) is supported by the WDCC, providing the opportunity for the scientists to receive additional scientific credits for their data publication.