

Simulated temperature change compared to 1986-2005 for the scenario RCP 8.5 using the earth system model MPI-ESM.

Data Management for CMIP5 at DKRZ

The Climate Model Inter-comparison Project (CMIP) is a framework for global climate simulations under the auspices of the World Climate Research Programme (WCRP).

In CMIP5, the 5th phase of CMIP, a set of standard model experiments is promoted in order to

- evaluate how realistically the models can reproduce the recent past,
- provide estimates of future climate change,
- understand the reasons for differences in model results.

The purpose of the simulations is to investigate scientific questions that have been identified by the Intergovernmental Panel on Climate Change (IPCC) in its recent report AR4, i.e. to improve the understanding of climate, and to provide future climate projections useful to adapt to or mitigate climate change.



German Climate Computing Center (DKRZ)

The mission of DKRZ is to provide high performance computing (HPC) platforms, sophisticated and high capacity data management and services for premium climate science.

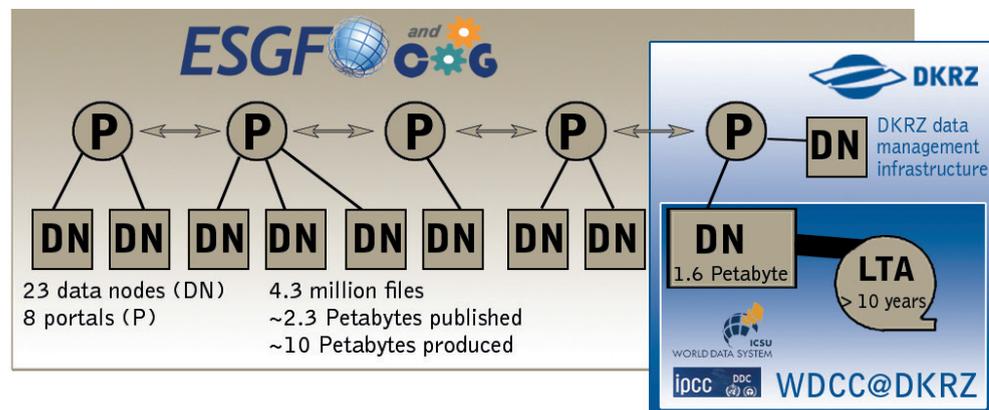
Besides providing HPC services, DKRZ supports projects in all aspects relevant to data management. This includes preparation, quality assessment, distribution, and long-term archiving (LTA) of data.

www.dkrz.de

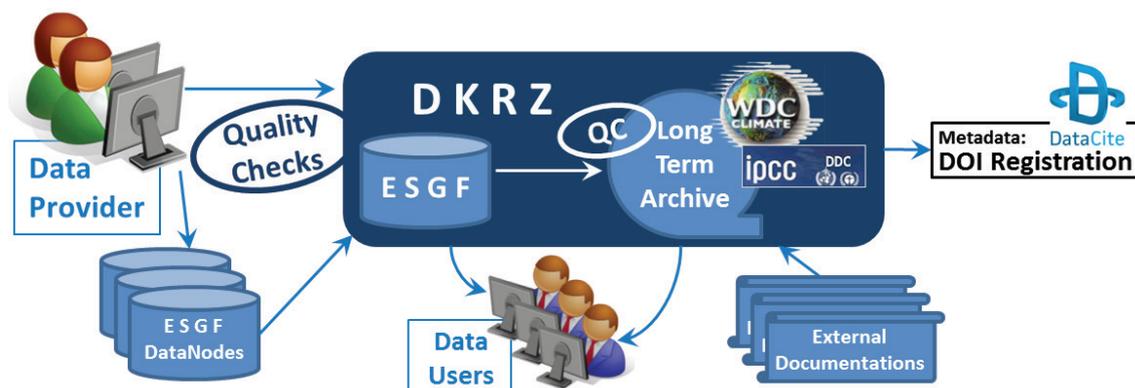


Challenges in data management for CMIP5:

- Big data volumes
- Huge number of files
- Numerous modeling groups providing data of varying quality
- International cooperation / global archive federation
- Long production cycles (3 to 5 years)
- Cross-discipline data usage beyond the climate modeling community requires
 - fast and reliable data access
 - extensive data documentation
 - ensuring data quality
- Long-term availability



The CMIP5 data are published within the Earth System Grid Federation (ESGF). As part of ESGF, DKRZ maintains a data node (DN) and a portal (P) for data search and selection. With its long-term archive (LTA), DKRZ provides the basis for the IPCC Data Distribution Centre.



Thorough quality checks are performed before data is published in ESGF and long-term archived in WDCC. In order to allow proper citation, DKRZ assigns DataCite DOIs to reasonable data entities.

Implementation of the CMIP5 project at DKRZ

Quality assessed database

In the framework of CMIP5, DKRZ established a database hosting CMIP5 data of researchers in Germany and worldwide. Due to the high responsibility of the project and the envisaged long-term cross-disciplinary usage of the data, a three-step quality ensuring workflow for data and metadata was developed and implemented at DKRZ. Data standards set up by CMIP5 provide the basis for the quality assessment. The information provided, on one hand allows climate modelers to assess whether their data is ready for publication and, on the other hand, enables data users to evaluate the appropriateness of the data for their purpose.

ESGF data publication

After having passed the quality assessment, the data is made available through the Earth System Grid Federation (ESGF). ESGF is a gateway to scientific data collections hosted at several globally distributed data centers. Together with the PCMDI (USA) and the BADC (GB), DKRZ runs one of the major ESGF-nodes for CMIP5, which are responsible for data replication. By using ESGF, researchers have comprehensive and consistent access to all data sets available within the project.

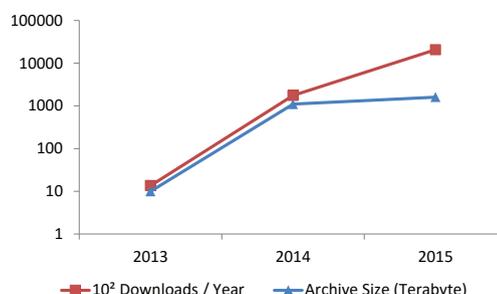
Data citation

In order to make the data citable, i.e. help data users to properly give credit to the data creators, DKRZ established an agreed workflow that procures a uniform citation of CMIP5 data and culminates in the assignment and registration of DataCite digital object identifiers (DOI) for transparent data access. These DOIs are unique, persistent, and globally resolvable.

Long-term archiving in WDCC

Generally accepted science policies request the data to be preserved for at least ten years. Therefore the German and core parts of international CMIP5 data – together with additional documentations on models, related scientific articles and contributor names – are long-term archived within the certified World Data Centre for Climate (WDCC) run by DKRZ. As an additional service, DKRZ hosts the reference data archive of the IPCC Data Distribution Centre (IPCC-DDC). The CMIP5 data archived in ESGF,

WDCC and IPCC-DDC is long-term curated and disseminated via different portals beyond the lifetime of the CMIP5 project. In order to avoid data duplication, a gateway between ESGF and the DKRZ long-term archive was implemented. The guaranteed long-term availability not only grants a reference data collection to the scientific community for later reuse, it also allows a reproducible scientific workflow within the project.



Size of CMIP5 data within the WDCC archive and number of downloads per year.

Benefits for CMIP6

The successful data management within this project was only possible through close cooperation with local partners such as the Max Planck Institute for Meteorology, as well as international partners within the framework of the European Network for Earth Sciences (ENES), ESGF, and others. The next phase of CMIP, CMIP6, has already started. It will benefit from the experiences gained during CMIP5 as basic aspects of the workflows and infrastructure will be the same. A tenfold increase of the data volume, however, is expected. Data management for CMIP5 at DKRZ was supported by BMBF grants and the DKRZ shareholders. To learn more about data management services at DKRZ, please visit:

www.dkrz.de/daten/data-services

Key benefits of DKRZ data services

For data users:

- Consistent data and metadata structure
- Quality controlled and well-documented data collection
- Stable and uniform access interface

For data producers:

- Long-term data archiving, curation, and dissemination in a certified repository
- Quality controlled data publication including DOI registration
- Consistent citations by data users with the DOI citation reference