Project: **1017**

Project title: **Konsortialdatenprojekt CMIP-DKRZ-Datenpool**

Principal investigator: **Stephan Kindermann**

Report period: **2021-11-01 to 2022-10-31**

The storage resources allocated to the CMIP-DKRZ-Datenpool (4.75 PByte) are fully booked with the national CMIP6 contributions and replicated core data collections from the overall global CMIP6 data holding (currently ~12 PByte).

The data pool content on one hand is builds up the DKRZ storage space accessible via the global ESGF data federation based on ESGF data nodes and the ESGF portal deployed at DKRZ (https://esgf-data.dkrz.de) and provides the largest CMIP data space integrated into the European ENES data infrastructure. On the other hand the pool is accessible to all DKRZ users (via /pool/data). The direct exploitation of this data pool based on the DKRZ HPC resources is supported by:

- Automatic generation and provisioning of intake data catalogs, directly usable in jupyter notebooks (e.g. by using the DKRZ jupyter hub)
- Provisioning of detailed documentation which is continuously updated, major updates of the pool are communicated via DKRZ documentation[1] and the DKRZ Blog[2]

During the reporting period the management of the data pool content (replication, update of data collections, removal of data collections) was driven by different requirements:

- Satisfying the requests of researchers to include additional data replicas into the data pool to support their analysis activity.
- Updating the data pool by removing retracted data and updating data with new versions of CMIP6 data.

As the DKRZ CMIP data pool content is directly accessible to all DKRZ users, no detailed overview can be given on the different analysis activities which were supported during the reporting period. For non DKRZ users and to support coordinated exploitation of the data pool at the European level (supported by the IS-ENES3 project) the DKRZ project 1088 was heavily used and showed interest from various user groups:

- IPCC working group members, preparing figures for the IPCC report
- Climate model evaluation activities (e.g. using the ESMValTool)
- Climate impact researchers interested generating derived data products
- Climate service center members and commercial climate service providers exploring the possibilities of having a direct access to CMIP6 via jupyterhub.
- Interdisciplinary user groups in the context of projects related to the European Open Science cloud.
- Other DKRZ projects currently making use of the CDP include ClimExtreme, DAKI RegIKlim, CLINT and CLICCs etc.

[1] DKRZ CDP documentation: https://cmip-data-pool.dkrz.de

[2] DKRZ CDP blog: z.B. https://blog.dkrz.de/dkrz-cdp-updates-july-21.html

The following figure illustrates the current status (a dataset corresponds to a complete time series of a specific variable, thus datasets correspond to a set of netcdf files):

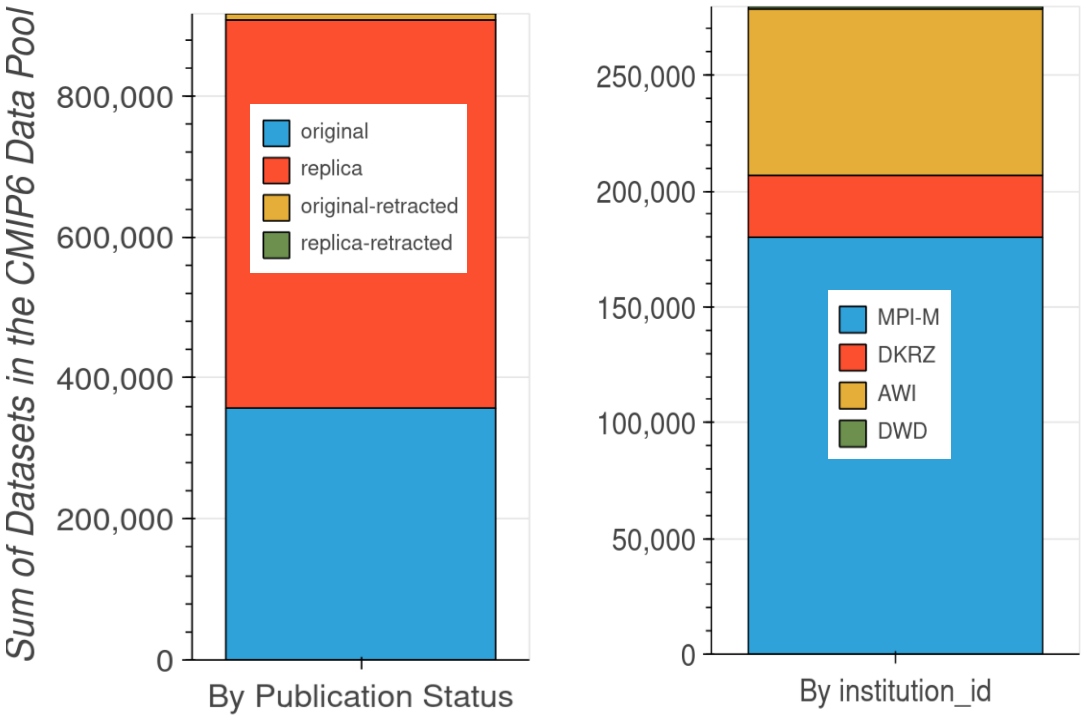


**Fig 1:** *Left*: An internationally agreed and communicated subset of CMIP6 is re-published into ESGF as “replica”. Modeling groups continue to update and correct data, outdated data is “retracted” and after some time removed from the pool. Currently the datasets sum up to more than 4 PBytes. *Right*: The CMIP6 pool contains all German CMIP6 contributions. This figure illustrates the number of datasets made available by the individual German CMIP6 modeling groups. The 180.000 datasets correspond to around 1.5 PByte.

**CovidMIP and Ensembles**

Emission-driven experiments based on CMIP experiments were performed in CovidMIP (Jones et. al, 2021) which became part of CMIP and the CDP. Within CovidMIP, such ESM simulations were used to assess the impact on climate of reduced emissions of $CO_2$, other greenhouse gases and ozone induced by lockdown restrictions.

Because CMIP provides the community-wide accepted protocol and the data standard for many basic evaluation experiments, cross-project activities adapt the CMIP workflow and increase the value of the CDP by extending single model simulation ensembles (SMSE). In addition to the CMIP model ensemble, the pool contains SMSE for a range of CMIP and ScenarioMIP experiments. As an example: for each of MPI-ESM1-2-LR (paper in preparation), EC-EARTH3 (Döscher et. al, 2021) or CanESM5 (Swart et. al, 2019), the CDP provides more than 40 simulations for the historical experiment. In contrast to the CMIP model ensemble, such SMSE helps to better estimate the internal variability and to address uncertainties in model evaluation by enabling more refined statistical analysis. This is only feasible when all ensemble members are kept available at one destination on the pool.

Jones, C.D., et al. (2021) The climate response to emissions reductions due to COVID-19: Initial results from CovidMIP. *Geophys. Res. Letters*, https://doi.org/10.1029/2020GL091883

Döscher, R., et al. (2022) The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6, Geosci. Model Dev., 15, 2973–3020, https://doi.org/10.5194/gmd-15-2973-2022 , 2022.

Swart, N., et al. (2019): The Canadian Earth System Model version 5 (CanESM5.0.3), Geosci. Model Dev., 12, 4823–4873, https://doi.org/10.5194/gmd-12-4823-2019 .