

Beyond ESGF - Regional climate model datasets in the cloud on AWS S3

Freitag, 9. Juni 2023 12:00 (15 Minuten)

The [Earth System Grid Federation](#) (ESGF) data nodes are usually the first address for accessing climate model datasets from WCRP-CMIP activities. It is currently hosting different datasets in several projects, e.g., CMIP6, CORDEX, Input4MIPs or Obs4MIPs. Datasets are usually hosted on different data nodes all over the world while data access is managed by any of the ESGF web portals through a web based GUI or the ESGF Search RESTful API. The ESGF data nodes provide different access methods, e.g., https, opendap or globus.

Beyond ESGF, there has been the [Pangeo / ESGF Cloud Data Working Group](#) that coordinates efforts related to storing and cataloging CMIP data in the cloud, e.g., in the [Google cloud](#) and in the [Amazon Web Services](#) Simple Storage Service (S3) where a large part of the WCRP-CMIP6 ensemble of global climate simulations is now available in analysis-ready cloud-optimized (ARCO) zarr format. The [availability in the cloud](#) has significantly lowered the barrier for users with limited resources and no access to an HPC environment to work with CMIP6 datasets and at the same time increases the chance for reproducibility and reusability of scientific results.

We are now in the process of adapting the Pangeo strategy for publishing also regional climate model datasets from the CORDEX initiative on AWS S3 cloud storage. Thanks to similar data formats and meta data conventions in comparison to the global CMIP6 datasets, the workflows require only minor adaptations. In this talk, we will show the strategy and workflow implemented in python and orchestrated in github actions workflows as well as a demonstration of how to access CORDEX datasets in the cloud.

Hauptautor: BUNTEMEYER, Lars (Helmholtz-Zentrum Hereon)

Vortragende(r): BUNTEMEYER, Lars (Helmholtz-Zentrum Hereon)

Sitzung Einordnung: Talks