

EGU24-18256, updated on 24 Mar 2025
<https://doi.org/10.5194/egusphere-egu24-18256>
EGU General Assembly 2024
© Author(s) 2025. This work is distributed under
the Creative Commons Attribution 4.0 License.



Data access for km-scale resolution models

Florian Ziemer¹, Tobias Kölling², and Lukas Klufft²

¹Deutsches Klimarechenzentrum, Applications, Hamburg, Germany (ziemer@dkrz.de)

²Max-Planck-Institut für Meteorologie, Climate Physics, Hamburg, Germany

With the transition to global, km-scale simulations, model outputs have grown in size, and efficient ways of accessing data have become more important than ever. This implies that the data storage has to be optimized for efficient read access to small sub-sets of the data, and multiple resolutions of the same data need to be provided for efficient analysis on coarse as well as fine-grained scales.

In this high-level overview presentation, we present an approach based on datasets. Each dataset represents a coherent subset of a model output (e.g. all model variables stored at daily resolution). Aiming for a minimum number of datasets makes us enforce consistency in the model output and thus eases analysis. Each dataset is served to the user as one zarr store, independent of the actual file layout on disks or other storage media. Multiple datasets are grouped in catalogs for findability.

By serving the data via https, we can implement a middle layer between the user and the storage systems, allowing to combine different storage backends behind a unifying frontend. At the same time, this approach allows us to largely build the system on existing technologies such as web servers and caches, and efficiently serve data to users outside the compute center where the data is stored.

The approach we present is currently under development in the BMBF project WarmWorld with contributions by the H2020 project nextGEMS, and we expect it to be useful for many other projects as well.