Project: 1099

Project title: ERA-5-Extract (E5E)

Principal investigator: Hannes Thiemann

Report period: 2025-11-01 to 2026-10-31

The aim of the E5E project is to provide German climate research with easy and fast access to comprehensive ERA5 climate reanalysis data on the disk of DKRZ's HPC Levante (directory /pool/data/ERA5). ERA5 data are produced by the Copernicus C3S at ECMWF. The entire ERA5 data are stored in ECMWF's MARS archive. Retrieval of ERA5 data from this archive is slow and manpower intensive. Having direct access to the data on Levante is therefore of great benefit to many climate users.

Within the report period Nov-25 to Oct-26, the most recent ERA5 data were continuously added. This means that the ERA5 data (1940 to present) and ERA5-Land data (1950 to present) are now available to DKRZ users on Levante's /pool/data/ERA5. Overall, the ERA5 data holdings in /pool/data/ERA5 consist of around six million GRIB files.

To facilitate efficient data access, index files were created for every GRIB file and stored in the same directory as the corresponding GRIB files. These index files serve as metadata files that provide a structured overview of the data contained in GRIB files. When combined with the DKRZ ERA5 intake catalog, this enables users to efficiently identify and access specific data entries without needing to read the entire file.

Users of /pool/data/ERA5 were informed about significant changes through announcements via the ERA mailing list, updates to the ERA5 user documentation in the DKRZ documentation portal, and meetings with key users. These meetings, including discussions with MPI users, confirmed the suitability of the data provision approach.

The ERA5 time series (in a compressed form) were additionally archived in the tape archive for backup.

As of the end of October 2025, around 1516 TiB of the granted 1700 TiB Levante disk storage are being used. For the retrieval, post-processing and archiving of ERA data, all of the granted 53 CPU node hours were consumed.