

Project: 1099

Project title: ERA-5-Extract (E5E)

Principal investigator: Hannes Thiemann

Report period: 2022-11-01 to 2023-10-31

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-22 to Oct-23, the retrieval and post-processing of the ERA5 time series covering the years 1940 to 1979 was finalized. This means that the entire ERA5 data (1940 to present) are now available to DKRZ users on Levante's /pool/data/ERA5. The DKRZ news entry to this achievement can be found at https://www.dkrz.de/en/communication/news-archive/era5-on-levante?set_language=en. In addition to the retrieval of the ERA5 data, the entire ERA5.1 data were added to /pool/data/ERA5, and an automated daily retrieval of the early-release ERA5-T data was implemented. During the reporting period, all data located at /pool/data/ERA5 have been reorganized and standardized according to a harmonized Data Reference Syntax, which includes file naming, directory names, and overall data structure. Details are described in the updated documentation https://docs.dkrz.de/doc/dataservices/finding_and_accessing_data/era_data/index.html#era-data. Users of /pool/data/ERA5 were notified about significant changes through a post on the ERA mailing list, which has 81 subscribers. The retrieval and reorganization of ERA5-Land data is in progress.

As of October 2023, around 1500 TiB of the granted 1600 TiB Levante disk storage are being used. For the postprocessing and restructuring of all /pool/data/ERA5 data, 347 CPU node hours were consumed.