Project: **pd1279 (Pool Data)**

Project title: **MESSy input and boundary conditions**

Principal investigator: **Patrick Jöckel**

Allocation period: **2025-01-01 to 2025-12-31**

The storage size in **/pool** is currently **99 TiB**, i.e. **82.5% of the 120TiB** which have been granted for the above allocation period.

During the allocation period the following data were added:

- the CAMS emission inventory (**23 TiB**): this data set is particularly large, since it provides the data at 0.1° x 0.1° horizontal resolution, monthly resolved for the years 2000 to 2025, split by (emission) sector, and with individual emission fluxes for VOCs (volatile organic compounds) as we require it for atmospheric chemistry research. Moreover
- additional nudging data (2024/25) for various EMAC grid resolutions (**702 GiB / year**),
- additional SST/SIC data for various CMIP6 scenarios and EMAC grid resolutions (**13 GiB**),
- the GRETA emission inventory (**566 GiB**),
- several initial files for tracer initialization in EMAC (**~ 100 GiB**) from previous simulations.

The granted storage in **/arch** (**120TiB**) has not yet been utilized.
Moreover, originally it was planned

- to back-up the **/pool** directory in **/arch**, and
- to establish a work-flow to reduce the required storage size in **/pool**, by moving older data or data that is not used very frequently into **/arch**, and to de-archive the data for larger projects only „on request".

This workflow was **not yet** established, mainly for two reasons:

1. Our data sets comprise both, a large number of small and several very large files. Thus, a simple "copy" or "sync" into **/arch**, although comfortable, is according to DKRZ not feasible in respect of an optimized resource utilization of the tape storage. Yet, we require a simple, **automated** method, since manual packing / archiving and de-archiving / unpacking is not feasible because it is error-prone, time consuming (human resources), and depending on dedicated persons who have to have the write permissions for the file systems.
2. Since a new storage system is planned to be installed soon, we refrained from developing our own "slk-based" hierarchical storage concept, since it will be outdated soon. We hope that the new system will allow for an automated hierarchical storage management to solve the issues mentioned above.