Project title: Dataset for aerosol and transport research at DLR PA-ESM department. Project acronym for link in /pool/data: DLR-PA-ESM Principal investigator (long-term responsible contact): Mattia Righi (<u>mattia.righi@dlr.de</u>) Applicant (if not the same as above): Allocation period: 01/2024-12/2026

Allocation Period	01/2024 – 12/2026.
Volume	5 TB
Expected Volume Change	5 TB
Total Volume requested	10 TB
License allows usage on DKRZ resources	All open-license datasets will have full read permissions. All other datasets will be open to members of the DKRZ Projects: 80, 853, 1063, 1393, 1395. Members of other DKRZ projects coordinated by DLR can be added if required.

Project overview

This data pool comprises different kinds of datasets required to support the aerosol and transport research at the department of Earth System Modelling for Aeronautics, Space, Transport and Energy of the DLR Institute of Atmospheric Physics (abbreviated DLR-PA-ESM).

The pool currently includes emission datasets from various providers (including DLR-internal ones by the DLR projects TraK and ELK), observational data for model tuning and evaluation purposes, and model-based datasets and look-up tables required for the development and testing of the Python packages developed by the PA-ESM department. The latter, in particular, include the *emacutils* package (a utility package for archiving and processing EMAC model data) and the *TransClim* climate response model (Rieger and Grewe, 2022). A shared conda environment with the required configuration to run those packages will also be included in the data pool, to allow users from different projects to use a common environment and facilitate collaboration and exchange.

Most of the datasets are (u=r,g=r,o=r) and have been collected from the respective sources, while some of them have been generated internally and are therefore (u=r,g=r,o=-). A detailed list is provided in the table at the end of the document.

The data can be accessed via standard tools like noview, cdo, not and the Python packages xarray and pandas. Given the relatively limited size of the data, no particular technical or computational challenges are involved.

Note that although most of the users of this project have also access to the <u>MESSy pool data</u> <u>project</u>, the data envisaged here have a different scope and are managed by different users. Therefore, they should be collected in a dedicated project.

Data content

The data pool is structured as follows:

- EMISSIONS
- OBSERVATIONS
- TESTDATA
- LUT
- □ SOFTWARE

The data are mostly in NetCDF format and have a coverage ranging from 1-year (for emission inventories, test data, and look-up tables) to several years (for observations).

The EMISSIONS sub-directory in the data pool may be extended with newer versions of the current datasets or with the inclusion of additional inventories. The same applies to the look-up tables (LUT) for the development of the *TransClim* package and, to a lesser extent, to the observations. This, however, is expected to remain within the range of a few terabytes.

Range of planned scientific data usage

The datasets are required by the members of the DLR PA-ESM department for software development and data analysis purposes. All datasets that have an open license will be made globally accessible. Access to some datasets, however, will be (u=r,g=r,o=-) because of embargoes or other copyright constraints. Access to these datasets will be granted to members of the DKRZ projects 80 (Climate and Transport), 853 (Earth System Chemistry Integrated Modelling ESCIMo), 1063 (Multiscale simulations of transport impacts on air quality and climate), 1393 (Global aerosol modelling for transport and aviation research) and 1395 (Numerical modelling for transport research). Some of these datasets may need to be made available to the public once the software for which they are required is published (e.g. via DKRZ LTA or Zenodo).

Data Storage Usage Plan

Most of the activities for which the data are required are supported by DLR Projects covering a three-year period. We therefore request an equivalent allocation period, i.e., until 31/12/2026.

Dataset / Path	Restriction	Purpose
EMISSIONS/		
CAMS-GLOB	(u=r,g=r,o=r)	Transport research
CEDS/CEDS-2021-04-21	(u=r,g=r,o=r)	Transport research
CEDS/CEDS-2023-04-18	(u=r,g=r,o=r)	Transport research
CEDS-GBD-MAPS	(u=r,g=r,o=r)	Transport research
EDGAR 6	(u=r,g=r,o=r)	Transport research
ELK	(u=r,g=r,o=-)	Transport research
TraK	(u=r,g=r,o=-)	Transport research
OBSERVATIONS/		
ACCESS	(u=r,g=r,o=-)	Model tuning / evaluation
ACCESS-2	(u=r,g=r,o=-)	Model tuning / evaluation
AERONET	(u=r,g=r,o=r)	Model tuning / evaluation
AirBase	(u=r,g=r,o=r)	Model tuning / evaluation
AsmiACP2011	(u=r,g=r,o=r)	Model tuning / evaluation

Overview of the data pool.

ATom	(u=r,g=r,o=r)	Model tuning / evaluation
BennartzACP2017	(u=r,g=r,o=r)	Model tuning / evaluation
CARSNET	(u=r,g=r,o=r)	Model tuning / evaluation
CASTNET	(u=r,g=r,o=r)	Model tuning / evaluation
CERES-EBAF	(u=r,g=r,o=r)	Model tuning / evaluation
CIRRUS	(u=r,g=r,o=-)	Model tuning / evaluation
CONCERT	(u=r,g=r,o=-)	Model tuning / evaluation
CR-AVE	(u=r,g=r,o=-)	Model tuning / evaluation
DC3	(u=r,g=r,o=-)	Model tuning / evaluation
EANET	(u=r,g=r,o=r)	Model tuning / evaluation
EMEP	(u=r,g=r,o=r)	Model tuning / evaluation
EMeRGe-Asia	(u=r,g=r,o=-)	Model tuning / evaluation
EMeRGe-EU	(u=r,g=r,o=-)	Model tuning / evaluation
ERA5	(u=r,g=r,o=r)	Model tuning / evaluation
ERA-Interim	(u=r,g=r,o=r)	Model tuning / evaluation
ESACCI-CLOUD	(u=r,g=r,o=r)	Model tuning / evaluation
GPCP-SG	(u=r,g=r,o=r)	Model tuning / evaluation
GrosvenorRG2018	(u=r,g=r,o=r)	Model tuning / evaluation
HIPPO	(u=r,g=r,o=r)	Model tuning / evaluation
HuneeusACP2011	(u=r,g=r,o=r)	Model tuning / evaluation
IMPROVE	(u=r,g=r,o=r)	Model tuning / evaluation
INCA	(u=r,g=r,o=-)	Model tuning / evaluation
KarydisACP2017	(u=r,g=r,o=r)	Model tuning / evaluation
KoehlerPCCP2009	(u=r,g=r,o=r)	Model tuning / evaluation
KraemerACP2009	(u=r,g=r,o=-)	Model tuning / evaluation
LACE	(u=r,g=r,o=-)	Model tuning / evaluation
MAC	(u=r,g=r,o=r)	Model tuning / evaluation
MahrtACP2018	(u=r,g=r,o=r)	Model tuning / evaluation
Melpitz	(u=r,g=r,o=r)	Model tuning / evaluation
ML-CIRRUS	(u=r,g=r,o=-)	Model tuning / evaluation
Oct-AVE	(u=r,g=r,o=-)	Model tuning / evaluation
Putaud	(u=r,g=r,o=r)	Model tuning / evaluation
SALTRACE	(u=r,g=r,o=-)	Model tuning / evaluation
TC4	(u=r,g=r,o=-)	Model tuning / evaluation
UCN-Pacific	(u=r,g=r,o=r)	Model tuning / evaluation
TESTDATA/	(u=r,g=r,o=r)	Test suite of the <i>emacutils</i> package
LUT/	(u=r,g=r,o=r)	Look-up tables for the <i>TransClim</i> package
SOFTWARE/	(u=r,g=r,o=-)	Shared conda environment

We provide a BibJSON bibliography for all datasets, which also includes license information. Generally, all datasets that have an open license will be made globally accessible. Access to some datasets, however, will be restricted because of embargoes or other copyright constraints.