Project: MESSy input and boundary conditions

Project acronym for link in /pool/data: MESSY

Principal investigator (long-term responsible contact): Patrick Jöckel (Patrick.Joeckel@dlr.de)

Applicant (if not the same as above):

Allocation period: 2022-01-01 to 2026-12-31

Allocation Period	2022-01-01 to 2026-12-31
Volume	143 TByte
Expected Volume Change	Yes
License allows usage on DKRZ resources	Yes (for MESSy license holders)

## Project overview

This project will enable sustainable access to initial and boundary conditions needed for simulations with the Modular Earth Submodel System (MESSy) in connection with various legacy base models, such as

- EMAC (ECHAM/MESSy Atmospheric Chemistry)
- COSMO-CLM/MESSy and MECO(n) (MESSy-fied ECHAM and COSMO models nested n-times)
- CESM1/MESSy
- ICON/MESSy

and for additional 'build-in' components (MPIOM (Max Planck Institute Ocean Model), CLM (community land model), OASIS3-MCT (coupler)). The links/paths to these datasets are hard-coded in the universal MESSy run-script and are therefore required to be persistent over time.

These data are needed by anybody wanting to perform simulations using the Modular Earth Submodel System with one (or more) of the legacy basemodels listed above on DKRZ resources.

The basis of the data are several external data sources (ECMWF, CCMI & CMIP, COSMO-CLM communities, scenario providers, emission inventory providers, etc.). In many cases the originators and creation methods of the data files are mentioned in the netcdf-file headers. Usage of data is documented in various MESSy-related publications (https://www.messy-interface.org -> MESSy Publications). For MESSy-submodel related files, a file naming convention applies.

Nudging data for EMAC "specified dynamics" simulations are created from ECMWF raw data with specific cdo-based tools in two steps, of which the first one is target-model-resolution independent (see also below).

The datasets were/are created to be specifically used by MESSy and are thus compatible with that model ecosystem. Access to the data is limited to holders of a MESSy license

(https://www.messy-interface.org -> MESSy Licence).

Moreover, access to ECMWF based forecast data (for MECO(n)) is restricted due to the very restrictive corresponding ECMWF license.

## Data content

Data is stored in a hierarchy of subdirectories:

- 1. Some directories contain data which are specific for various legacy base-models, which are connected to MESSy:
  - CESM1: initial and boundary data for the CESM1 Earth System Model
  - COSMO: grid, initial, and boundary data for the COSMO model
  - ECHAM5/
    - version>/init/<resolution> : original ECHAM5 initial files
    - <version>/add\_spec\*/<resolution> : specific (date/time) ECHAM5 initial files NOTES:

<version> = echam5.3.01, echam5.3.02

<resolution> = T106L191MA, T106L41DLR, T106L90MA, T21L19, T31L31ECMWF, T31L49MA, T42L31ECMWF, T42L47MA, T63L31ECMWF, T63L47MA, T63L90MA, T85L39MA, T85L49MA, T85L90MA, T106L31ECMWF, T106L47MA, T106L95MA, T21L39MA, T31L39MA, T31L90MA, T42L39MA, T42L87MA, T63L39MA, T63L49MA, T63L95MA, T85L41DLR, T85L60ECMWF, T85L95MA, T106L39MA, T106L87MA, T10L39MA, T255L47MA, T31L47MA, T42L19, T42L41DLR, T42L90MA, T63L41DLR, T63L87MA, T85L31ECMWF, T85L47MA, T85L87MA The specific initial files are required for "realistic" initialization at a given date and time (e.g. nudged hind-cast simulations).

 ICON/<version>/<setup>: grid and initial data for the ICON model in various setups NOTES:

<version> = icon2.0 <setup> = HDCP2, NWP : specific setups <setup> = shared : data shared between various setups

- 2. Some directories contain data for legacy components, which are used with MESSy:
  - MPIOM (is used as MESSy submodel): initial data for the ocean model
  - CLM (coupled to COSMO/MESSy via OASIS3-MCT): initial and boundary data for the community land model
  - OASIS/<setup>: weight data for various coupled COSMO/MESSy CLM and MECO(n) – CLM setups
- 3. These directories contain initial and boundary conditions for the various (mostly atmospheric chemistry related) MESSy submodels:
  - MESSy1/<setup> : development cycle 1 of MESSy, old specific setups
  - MESSy2/<setup> : development cycle 2 of MESSy, old specific setups

- MESSy2/raw/<submodel> : submodel specific initial and boundary data, e.g. emissions of trace gases and aerosols, specific scenarios, etc.
- 4. These directories contain "nudging" data for EMAC (ECHAM/MESSy) in specified dynamics ("nudged") mode:
  - NUDGING/ECMWF/[ANALY,ERAI]/<resolution> : old IEEE file format based on

ECMWF ERA-interim or operational ANALYsis data

- NUDGING\_NC/ECMWF/<type>/<resolution> : netcdf format
  <type> = ERA05.0, ERA05.1, ERA05 : ECMWF ERA5-data
  = ERAI : ECMWF ERA-Interim data
  = ANALY : ECMWF operatoinal ANALYsis data
  - = ANA, FCpre, FC : ECMWF forecast data [!!! License !!!]
- 5. Sea surface temperature / sea ice data for performing EMAC (ECHAM/MESSy) chemistry climate simulations:
  - SST
- 6. Auxiliary data to create specific boundary conditions or initial files:
  - EMISSIONS
  - TOOLS

Data are mostly in netcdf, some in grib, and some in ASCII format.

## Range of planned scientific data usage

Anybody running the Modular Earth Submodel System requires access to this data.

## Data Storage Usage Plan

We currently require **143 TByte\*** of storage in order to sustainably provide access to MESSy initial and boundary data.

The data volume is expected to increase, more specifically:

- recent ECMWF ERA5 based "nudging" data, shortly after the corresponding raw data become available
- ECMWF operational analysis and forecast data for MECO(n)-based hind- and forecast simulations accompanying measurement campaigns
- additional emission inventories and scenario data (CCMI, CMIP, other projects) for atmospheric chemistry applications
- additional grids (COSMO and ICON)
- \* Special note:

90 TByte of the 143 TByte data are currently occupied by the data in subdirectories

MESSY/DATA/NUDGING\_NC/ECMWF/ERA05[.1]/PRE/

These "PRE" data are required to generate the *model resolution-specific* EMAC (ECHAM) nudging data for "specified dynamics" (SD) simulations based on ECMWF ERA5 data, and are the result of a first, preparatory and *model resolution independent* pre-processing step based on ECMWF raw data.

Nudging data for specific EMAC resolutions will be generated from these "PRE" data and provided on request for the MESSy user community. For the cdo-based processing (both steps!) the I/O is the limiting factor, due to the high spatial and temporal (hourly) resolution of the ECMWF ERA5 data. Thus, data processing needs to be parallelised, which is only possible "over time", meaning that several monthly datasets are processed in parallel. For this, however, the high resolution "PRE" data need to be stored on disc, since repeated archiving and retrieval from tape would impose another bottleneck.

The required disc space has been applied for (/work) in the application of project 853 (ESCiMo) for the year 2021, however, it was not granted by the WLA. Thus, it was required to move these data (at least temporarily) to /pool, in order to be able to continue project 853. The disc space in /work was again requested by project 853 for the year 2022. In case the WLA will approve it this time, the data could also be moved back to the /work space of 853, although it is better suited and preferred to be stored in /pool, since it does not solely serve project 853, but all other projects with EMAC (SD simulation) involvement as well.