

Naming of climate simulations in the CLM Community

C. Steger¹, S. Brienien¹, A. Dobler², H. Feldmann³, H. Hagemann⁴, S. Kotlarski⁵, M. Schupfner⁶,
K. G6rger⁷

¹Deutscher Wetterdienst

²Norwegian Meteorological Institute

³Karlsruhe Institute of Technology

⁴Helmholtz-Zentrum Hereon

⁵MeteoSchweiz

⁶Deutsches Klimarechenzentrum

⁷Forschungszentrum J6ulich

21 November 2024

Background

The information about the institutions performing the simulations within CORDEX should be visible in the CORDEX directory structure and filenames. This was not the case for most of the simulations done in CORDEX – CMIP5 where CLMcom was used as institute_id. However, these credits are important for the institutions in order to defend the usage of fundings. Therefore, the following naming convention should be used for simulations created and published by members of the CLM Community.

Naming conventions

General comments:

- No changes for published simulations
- The **modeling groups which performed the simulation should be acknowledged** (e.g. as CLMcom-ETH) in publications! The [CORDEX Terms of Use](#) states: “Include in publications an acknowledgment with language similar to: “*We acknowledge the World Climate Research Programme, the CORDEX Science Advisory Team (SAT) – coordinating body of CORDEX, and the Working Group on Coupled Modelling (WGCM) – responsible panel for CMIP6. We thank the CORDEX climate modeling groups for producing and making available their model output, CMIP6 for providing the driving data, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CORDEX, CMIP6 and ESGF.*”
- Thus, the institution_id should include the name of the institution that performed the simulation. We suggest extending the acknowledgements for existing simulations (CORDEX-CMIP5) where CLMcom is used as institution_id, by “This simulation has been produced by <institution> as part of the contribution of the CLM-Community (CLMcom) to CORDEX.” and/or mention the appropriate institution in the table (e.g. as CLMcom-ETH, CLMcom-BTU, ...). The necessary information on the institutes performing the specific CLMcom simulations is available at [CORDEX section on the CLM-Community webpage](#).

(1) DRS elements

Table 1: Overview of the DRS elements used in the CORDEX directory structure and the filenames. Please note, that the value of each DRS element must meet the requirements defined in the CORDEX archive specifications (especially registration of controlled vocabulary). ETH is used as a representative for the member institution.

DRS element	COSMO-CLM	ICON-CLM
institution_id	<p>Identifier for the institution that is responsible for generating and providing CORDEX simulations. All CORDEX institutions must be registered to publish their simulations on ESGF (see CORDEX-CMIP6 archiving specifications for further information).</p> <p>Example: CLMcom-ETH</p>	<p>Identifier for the institution that is responsible for generating and providing CORDEX simulations. All CORDEX institutions must be registered to publish their simulations on ESGF (see CORDEX-CMIP6 archiving specifications for further information).</p> <p>Example: CLMcom-ETH</p>
source_id	<p>Identifier of the CORDEX RCM (max. 25 characters). Different configurations of the same RCM such as different combinations of parameterization schemes or changes in parameters for existing schemes must be reflected in source_id by a free text suffix (e.g. RCM123, RCM123A for Africa or RCM123T for the tropics). The same applies to coupled model versions. The same applies to coupled model versions.</p> <p>RCM simulations with spectral nudging must use the “SN” suffix in source_id (e.g. RCM123-SN) while RCM simulations with Newtonian/dynamical nudging must use the “NN” suffix (e.g. RCM123-NN).</p> <p>Example: CCLM6-0-1</p>	<p>Identifier of the CORDEX RCM (max. 25 characters). Different configurations of the same RCM such as different combinations of parameterization schemes or changes in parameters for existing schemes must be reflected in source_id by a free text suffix (e.g. RCM123, RCM123A for Africa or RCM123T for the tropics). The same applies to coupled model versions.</p> <p>RCM simulations with spectral nudging must use the “SN” suffix in source_id (e.g. RCM123-SN) while RCM simulations with Newtonian/dynamical nudging must use the “NN” suffix (e.g. RCM123-NN).</p> <p>For ICON-CLM we will use the following naming convention: ICON-CLM-YYYYMM-X-Y Example: ICON-CLM-202407-1-1</p> <p>YYYYMM are the year and month of the open source release on which the version is based. X is a number to indicate potential sub releases (e.g. for hotfixes) of the open source release. Y is a number that can be used for potential changes or different setups of ICON-CLM from the CLM Community (i.e. ICON-CLM subversions)</p>

Example path	/CORDEX/CMIP6/DD/EUR-12/CLMcom-ETH/GCM/historical/r1i1p1f1/CCLM6-0-1/v1-r1/mon/tas/v20240319	/CORDEX/CMIP6/DD/EUR-12/CLMcom-ETH/GCM/historical/r1i1p1f1/ICON-CLM-202407-1-1/v1-r1/mon/tas/v20240319
Example file	tas_EUR-12_GCM_historical_r1i1p1f1_CLMcom-ETH_CCLM6-0-1_v1-r1_mon_201101-201412.nc	tas_EUR-12_GCM_historical_r1i1p1f1_CLMcom-ETH_ICON-CLM-202407-1-1_v1-r1_mon_201101-201412.nc

If the source_id cannot hold the complete model information, e.g., in case of coupled models with several component models, an abbreviated source_id needs to be found; the “label_extended” may be used to specify the coupled model’s details in the CORDEX-CMIP6 controlled vocabulary source_id registration and the optional “comment” in the global attributes of the ESGF compliant simulation data netCDF file may contain this information too.

Simulations with two nesting steps

Information about standardization of convection-permitting simulations can be found in the document https://www.hymex.org/cordexfps-convection/wiki/lib/exe/fetch.php?media=fps_convection_filenaming_2020_update.pdf. The filename and directory structure for simulations with two nesting steps is the same as for “normal” CORDEX simulations. For the convection-permitting simulations with CORDEX-CMIP5 boundary conditions, the information about the nesting steps were implemented into the CORDEX Data Reference Syntax (DRS) element rcm_version_id (corresponds to version_realization in CORDEX-CMIP6). However, the DRS element version_realization has a structured form in CORDEX-CMIP6 and is not a free string anymore. The procedure for simulations with two nesting steps is not clear at the moment. An official standard for this kind of simulation is necessary.

Summary of nesting info for CORDEX-CMIP5 (description of the DRS element RCM_VERSION_ID in the specific cases; see FPS convection document for more details):

https://www.hymex.org/cordexfps-convection/wiki/lib/exe/fetch.php?media=fps_convection_filenaming_2020_update.pdf,

<project_acronym> stands for the short name of the “project” the simulation belongs to, e.g. fpsconv for FPS Convection)

- No nest: <project_acronym>-x0n1-v1
- One nest, no differences between nest and initial run: <project_acronym>-x1n2-v1
- One nest, differences between nest and initial run: <project_acronym>-x2n2-v1
- One nest, additional differences between nests: <project_acronym>-x2yzn2

However, we are not aware of a possibility to publish anything on ESGF under CORDEX on other domains than the official ones listed here: https://github.com/WCRP-CORDEX/cordex-cmip6-cv/blob/main/CORDEX-CMIP6_domain_id.json



For CORDEX-FPSCONV, these are the currently available options that more or less define the naming conventions: <https://github.com/ESGF/config/blob/master/publisher-configs/ini/esg.cordex-fpsconv.ini>

(2) Attributes in NetCDF files

Table 2: Overview of the global attributes used in the header information of the NetCDF files. Please note, that the value of each attribute must meet the requirements defined in the CORDEX archive specifications (especially registration of controlled vocabulary). ETH is used as a representative for the member institution. Examples from other modeling centers for `source_id` are available here https://github.com/WCRP-CORDEX/cordex-cmip6-cv/blob/main/CORDEX-CMIP6_source_id.json. The fields `rcm_config_cclm/rcm_config_icon-clm` and `rcm_config_int2lm` are only for internal information/documentation and are not part of the official CORDEX standard. The data producer should include these additional global attributes in the header of the NetCDF files and provide the configuration files in the CLM Community Gitlab repository at DKRZ ([CLM-Community / COSMO-CLM Configurations · GitLab \(dkrz.de\)](https://gitlab.dkrz.de/CLM-Community/COSMO-CLM-Configurations)) for COSMO-CLM/INT2LM and [CLM-Community / ICON-CLM Configurations · GitLab \(dkrz.de\)](https://gitlab.dkrz.de/CLM-Community/ICON-CLM-Configurations) for ICON-CLM.

Attribute	COSMO-CLM	ICON-CLM
institution	Climate Limited-area Modelling Community (CLMcom) partner: ETH Zurich, Zurich, Switzerland	Climate Limited-area Modelling Community (CLMcom) partner: ETH Zurich, Zurich, Switzerland
institution_id	CLMcom-ETH	CLMcom-ETH
source	COSMO model in climate mode (COSMO-CLM), Version 6.0 clm 1	ICON model in Climate Limited-area Mode (ICON-CLM), Version 202407-1-1
source_id	CCLM-6-0-1	ICON-CLM-202407-1-1
further_info_url	https://www.clm-community.eu/COSMO-CLM/Overview	https://www.clm-community.eu/ICON-CLM/
label	= source_id	= source_id
label_extended	= source	= source
license	<p>Creative Commons Attribution 4.0 International License (CC BY 4.0); https://creativecommons.org/licenses/by/4.0</p> <p>Comment: To our knowledge the license for the data must be given here. Only one license per source_id (= model version) can be registered. We recommend CC BY 4.0. If an institution cannot use this license, it is likely necessary to register another model version.</p>	<p>Creative Commons Attribution 4.0 International License (CC BY 4.0); https://creativecommons.org/licenses/by/4.0</p> <p>Comment: To our knowledge the license for the data must be given here. Only one license per source_id (= model version) can be registered. We recommend CC BY 4.0. If an institution cannot use this license, it is likely necessary to register another model version.</p>
release_year	The year in which the model version was released.	The year in which the model version was released. For ICON this corresponds to the year in source_id
source_type	ARCM or AORCM Explanation: ARCM = "atmospheric regional climate model, including a land model"	ARCM or AORCM Explanation: ARCM = "atmospheric regional climate model, including a land model"

	AORCM = "atmosphere-ocean regional climate model"	AORCM = "atmosphere-ocean regional climate model"
contact	xxx@env.ethz.ch (one <i>generic email address or website with contact information</i> for each institution)	xxx@env.ethz.ch (one <i>generic email address or website with contact information</i> for each institution)
references	www.clm-community.eu and www.cosmo-model.org	https://www.icon-model.org/publications/reference-publications
comment	Add something like: "Please use the following reference for this climate data: Climate projection performed by institution or name or both...."	Add something like: "Please use the following reference for this climate data: Climate projection performed by institution or name or both...."
grid	Rotated-pole latitude-longitude with 0.11 degree grid spacing.	Rotated-pole latitude-longitude with 0.11 degree grid spacing, interpolated by nearest neighbor remapping from the original unstructured icosahedral ICON grid R13B05 (~12.1 km).
config_cclm / config_icon-clm	<p><project_id>_<mip_era>_<activity_id>_<domain_id>_<institution_id>_<driving_source_id>_<driving_experiment_id>_<driving_variant_label>_<source_id>_<version_realization>_<version></p> <p>Attribute describing the configuration of CCLM and INT2LM used for this simulation. The configuration files for CCLM (YUSPECIF, INPUT_ASS, INPUT_DIA, INPUT_DYN, INPUT_INI, INPUT_IO, INPUT_ORG, INPUT_PHY, optional README with further information) and INT2LM (OUTPUT, INPUT, optional README with further information) must be provided in a separate directory in the Gitlab project https://gitlab.dkrz.de/clm-community/cosmo-clm-configurations. The name of the directory must match the attribute exactly.</p> <p>Example attribute: CORDEX_CMIP6_DD_EUR-12_CLMcom-ETH_GCM_historical_r1i1p1f1_CCLM6-0-1_v1-r1_v20240319</p>	<p><project_id>_<mip_era>_<activity_id>_<domain_id>_<institution_id>_<driving_source_id>_<driving_experiment_id>_<driving_variant_label>_<source_id>_<version_realization>_<version></p> <p>Attribute describing the configuration of ICON-CLM used for this simulation. The configuration files must be provided in a separate directory in the Gitlab project https://gitlab.dkrz.de/clm-community/icon-clm-configurations. The name of the directory must match the attribute exactly.</p> <p>Example attribute: CORDEX_CMIP6_DD_EUR-12_CLMcom-ETH_GCM_historical_r1i1p1f1_ICON-CLM-202407-1-1_v1-r1_v20240319</p>

Simulations with two nesting steps

For CORDEX-CMIP5 simulations with more than one nesting step, the information in RCM_VERSION_ID provides a first overview of the nesting steps of the simulation. Due to the fixed set of mandatory global attributes in CORDEX, this information cannot be described in more detail in the mandatory global attributes. For CORDEX-CMIP5, it is recommended to include the information about nests and differences between them by using the following optional global attributes in the NetCDF files:

- :nesting_levels: the number of nests
- :comment_nesting: general info
- :comment_1nest: info referring to outer nest-run
- :comment_2nest: info referring to inner nest-run

Example: (from orog_ALP-3_ECMWF-ERAINT_evaluation_r0i0p0_CLMcom-ETH-COSMO-crCLIM_fpsconv-x2yn2-v1_fx.nc)

```
:nesting_levels = "2" ;
```

```
:comment_nesting = "these are results of the 2nd nest of a 2-step nesting COSMO-crCLIM simulation (1st nest at 0.11° spatial resolution over EUR-11 domain, 2nd nest at 0.02° spatial resolution over ALP-3 domain)" ;
```

```
comment_1stNest = "convection parametrized ERA-Interim driven simulation (direct downscaling) for EUR-11 domain" ;
```

```
:comment_2ndNest = "convection permitting FPS simulation; forcing derived from 1st nest output" ;
```

For CORDEX-CMIP6 simulations the procedure is not clear at the moment. An official standard for this kind of simulation is necessary.

(3) Use of DOI

We recommend that each data set (each simulation) should get a DOI. With a publication in the long term archive WDCC (which will be done after the ESGF publication) a DOI would be attributed to each simulation. Otherwise, the institution which produced the data set should take care of this.

References:

- CORDEX-CMIP6 archiving specifications: [CORDEX-CMIP6 Archiving Specifications for Dynamical Downscaling \(zenodo.org\)](https://zenodo.org/record/3908127/files/CORDEX-CMIP6_Archiving_Specifications_for_Dynamical_Downscaling.pdf)
- CORDEX Terms of Use: <https://cordex.org/data-access/cordex-cmip6-data/cordex-cmip6-terms-of-use/>
- CORDEX FPS “Convective phenomena at high resolution over Europe and the Mediterranean” file naming protocol: https://www.hymex.org/cordexfps-convection/wiki/lib/exe/fetch.php?media=fps_convection_filenaming_2020_update.pdf