

CORDEX Archive Design

http://is-enes-data.github.io/cordex_archive_specifications.pdf

Version 3.2, 5 May 2020

O.B. Christensen¹, W.J. Gutowski², G. Nikulin³, and S. Legutke⁴

1 Introduction

This document specifies technical aspects of CORDEX archive file and data formats, as well as archive content^a. It includes a common naming system, the Data Reference Syntax (DRS), which allows the identification of data sets wherever they might be located in a distributed CORDEX archive. The file format, the naming system, as well as the metadata build on controlled vocabularies (CVs) as explained in section 5. These CVs are listed in CORDEX_variables_requirement_table.xls (VR)[1] which should be consulted when the CORDEX files are being prepared. Wherever possible, the DRS and metadata are defined as in CMIP5^b. They are described in the following and apply to all three classes of data:

- Core: monthly and seasonal,
- Tier 1: daily,
- Tier 2, higher frequency data.

Tier 2 data will be stored locally at modeling centers and made available on an informal basis upon request^c. Core and Tier 1 data will be published in central CORDEX archives. These may be regional data portals or searchable archives relying on the metadata, like, for example, the CORDEX ESGF archive (see <https://cordex.org/data-access/>). Therefore, at least for the data that are supposed to be published in the CORDEX ESGF archive:

The data must comply 100% with the DRS and CVs.

For detailed instructions on submitting CORDEX data see <https://cordex.org/experiment-guidelines/how-to-submit-data-rcms/> or contact cordex-registration@cordex.org. **Note that changes with regard to earlier versions are written in 'red'.**

¹Danish Meteorological Institute (DMI)

²Iowa State University

³Swedish Meteorological and Hydrological Institute (SMHI)

⁴German Climate Computing Center (DKRZ)

2 File format

Data files have to comply with the NetCDF format, version 4 compressed with zlib deflation, using the NetCDF 4 classic data model^d. The CF convention 1.4 or later has to be followed^e.

Each file may contain only one output field (target variable) from a single simulation. It has to include attributes and coordinate variables as defined herein, and in VR [1]. The entire time series of a target variable has to be distributed over several files as described in section 5.4.

All output fields must be of single precision (type NC_FLOAT), while all coordinate variables (time and space) have to be of double precision (type NC_DOUBLE) in accordance with the CMIP5 specifications.

Note that files produced with the help of the CMOR software are expected to be CORDEX compliant.^f

2.1 NetCDF attributes

The specification of attributes builds on CMIP5 as much as possible (see [5]), as well as on experience gained from the NARCCAP⁵ as well as ENSEMBLES⁶ projects. A list of mandatory and optional global NetCDF attributes^g as well as corresponding CVs can be found in sheets 'Global attributes' and 'CV' of VR [1]. For examples of global attributes please refer to Appendix A. Coordinate or variable attributes depending on the grids or variables are described in the sections that follow.

Examples of a mandatory and an optional NetCDF attribute are the *creation_date* and the *tracking_id* respectively. Please note that services around the ESGF may be based on the assumption that both *tracking_id* and *creation_date* of files with the same name are unique (e.g. annotations may be attached to a file through the *tracking_id* or the *creation_date*). Therefore, whenever a file is provided that shall replace an existing ESGF file, the *tracking_id* and *creation_date* must be different from those in the file to be replaced.

3 Grids

3.1 CORDEX domains

The CORDEX domains are defined as rectangles in rotated-pole coordinates in [2]. A domain has to lie inside the RCM interior computational domain, i.e. in the area left once the relaxation zone is excluded. It is strongly recommended that RCMs using the rotated pole projection exactly follow the CORDEX grid definition provided in detail in Table 1.

⁵http://rcmlab.agron.iastate.edu/narccap/output_requirements.html

⁶<http://ensemblesrt3.dmi.dk>

The domain acronym has to be 'domain'-'resolution', for example 'domain'-44, 'domain'-22, 'domain'-11, etc. corresponding to the chosen grid spacing. The data have to be provided for the CORDEX domain only, i.e. the relaxation zones have to be cut off before the data is delivered.

Names of the CORDEX domains are provided in [Table 1](#). CORDEX data should use the CORDEX domain name that best coincides with the actual domain and resolution. **For native coordinate systems data files must contain required grid information in the variable attribute - *grid_mapping* and in the coordinate variables (*rotated_pole, rlon, rlat* or *Lambert_Conformal, x, y* for example) in accordance with the CF-1.4 or later convention [3].** Examples for the rotated pole and for the Lambert Conformal projections can be found in Appendix A. It is strongly recommended to provide the geographic latitudes and longitudes of the model grid cell positions as well, in order to facilitate the data analyses^h. **Note that longitudes have to have absolute values as small as possible and monotonic.**

3.2 Regular geographic grids

All variables have to be provided on their native computational grid **besides the wind variables which are requested on the mass points of the native computational grid**. In addition, the Core variables have to be provided on a regular geographic latitude/longitude grid. These grids have grid point boundaries (not centers) at whole-number degrees of longitude and latitude, and have roughly the same resolution as the native gridⁱ.

The domain acronyms for the regular grids are the same as those for the corresponding model grid with the letter 'i' appended to the resolution (e.g. "SAM-44i"), as specified in [Table 2](#). **As for the native grid, the files with data on the regular grids have to provide grid information (coordinate variables *lon* and *lat* and their attributes) as required by the CF-1.4 or later convention [3], see examples in Appendix A.**

The regular geographic grid for the Core data has to comply
100% with the CORDEX specifications.

3.3 Vertical coordinates

In contrast to CMIP5, all requested variables are two-dimensional. Surface fields do not need a vertical dimension. However, higher-level variables, e.g. near-surface⁷ fields or pressure-level data, must have a single-value vertical coordinate variable

⁷ Near-Surface means at a height between 1.5 to 10.0 m

height or *plev* with the appropriate value and attributes following the CF-1.4 convention or later (see examples in [Appendix A](#) or in [5]). Variables averaged over pressure intervals, *clh*, *clm* and *cll*, need corresponding *plev_bnds*.

4 Time coordinate

The unit of the time coordinate is '*days since 1949-12-01T00:00:00Z*' or equivalent for all files' according to the CF conventions, i.e. the UNIDATA conventions as the relevant part of it. See also the footnote for valid formats. All time dependent variables must have an attribute *cell_methods: time* with values as listed in VR [1].

The time value of the instantaneous data is 0Z, 6Z, 12Z, 18Z and 0Z, 3Z, 6Z, 9Z, 12Z, 15Z, 18Z, 21Z of each day for the 6-hourly and 3-hourly data respectively. The averaged data have the time values 3Z, 9Z, 15Z, 21Z (6-hourly) and 1.5Z, 4.5Z, 7.5Z, 10.5Z, 13.5Z, 16.5Z, 19.5Z, 22.5Z (3-hourly).

Data of lower frequency, if calculated from instantaneous data, have to be based on data sampled at least as frequently as the sampling of the corresponding variable in Tier 2 (see sheet *all* of [1]), but they may equally well use a higher frequency like the model time step.

Variables that are representative for an interval (averages, maxima, minima) must use the midpoint of time intervals as time coordinate values and furthermore must have a *time_bnds* field of dimensions (ntimes,2)^k, where ntimes is the dimension of the time coordinate. The *time_bnds* values for daily, monthly and seasonal data should be on 00:00:00 UTC or multiples of 3 and 6 hours from that time for 3 hourly and 6 hourly data, respectively. Seasonal-mean data are defined with standard seasonal boundaries (DJF, MAM, JJA, and SON).

The *time* variable must have a *calendar* attribute, which corresponds to the calendar of the driving data (normally the one used by the model that produced the data). For the evaluation runs this is [*proleptic-*]gregorian or *standard*.

5 Data Reference System

The Data Reference Syntax (DRS) involves directory names and hierarchies, as well as file name build rules. The names are composed of individual elements which are described in the next section. Most of the elements can assume only values specified in a list of allowed values, the Controlled Vocabulary (CV) of that element, and are not allowed to take any other value. Valid values (CV or free text) that can be assigned to the elements are found in sheets "Global attributes" or "CV" of VR [1].

5.1 DRS elements

The DRS element values have to consist of the characters a-z, A-Z, 0-9 and '-' (dash). No other character is allowed. The terms in brackets following the DRS element names in the list below indicate whether the values are prescribed ('single value'),

have to be taken from a fixed list of values ('CV'), have to be registered with the CORDEX ('CV to register'), or can be chosen freely. Note that most elements must have the same value as a mandatory NetCDF attribute. The attribute name is included in the brackets after the element name in these cases.

The elements are (in alphabetical order):

activity (single value; project_id) has to have the value 'CORDEX'.

CMIP5EnsembleMember (CV; driving_model_ensemble_member) identifies the ensemble member of the CMIP5 experiment that produced the forcing data. It has to have the same value in CORDEX as in CMIP5. For evaluation runs it has to be r1i1p1. Invariant fields (frequency=fx) may have the value r0i0p0 or that of the corresponding *GCMEnsembleMember* attribute.

CMIP5ExperimentName (CV; driving_experiment_name) is either *evaluation*¹ or the value of the CMIP5 *experiment_id* of the data used.

Domain (CV; CORDEX_domain) is the name assigned to each of the CORDEX regions and includes a flag for resolution as listed in [Table 1](#) or [Table 2](#). Please refer also to section 3.1.

GCMModelName (CV; driving_model_id) is an identifier of the driving data. The name consists of an institute identifier and a model identifier. For reanalysis driven runs these are *ECMWF* and a name for the reanalysis data (*ERAINT*). For runs driven by CMIP5 model data these are the associated CMIP5 *institute_id* and the CMIP5 *model_id*. The two parts of the name are separated by a '-' (dash). Note that dashes in either of the two parts are allowed. The CV list of *GCMModelName* can be found at <http://is-enes-data.github.io/GCMModelName.txt>

Institution (CV; institute_id) is an identifier for the institution that is responsible for the scientific aspects of the CORDEX simulation (RCM configuration, experiments ...). The CV for *Institution* has to be coordinated in the worldwide CORDEX community. The actual state of the CV is found at http://is-enes-data.github.io/CORDEX_RCMs_ToU.txt

product (single value) has to have the value *output*.

RCMModelName (CV to register; model_id) is an identifier of the CORDEX RCM. It consists of the *Institution* identifier (see above) and a model acronym, connected by a dash (e.g. DMI-HIRHAM5 or SMHI-RCA4). The CV of the *RCMModelName* has to be coordinated in the worldwide CORDEX community. The actual state of the CV is found at http://is-enes-data.github.io/CORDEX_RCMs_ToU.txt.

If your group and model are not included in the CORDEX RCM list then you should register by contacting cordex-registration@cordex.org, see for details

Step 1: Registration at:

<https://cordex.org/experiment-guidelines/how-to-submit-data-rcms/>

RCMVersionID (*free string; rcm_version_id*) identifies reruns with perturbed parameters or smaller RCM release upgrades, i.e. equivalent simulations. Major upgrades and improvements should be reflected in the *RCMModelNameⁿ*.

Frequency (*CV; frequency*) is the output frequency indicator: *3hr*=3 hourly, *6hr*=6 hourly, *day*=daily, *mon*=monthly, *sem*=seasonal, and *fx*=invariant fields.

StartTime and **EndTime** (*build rule*) indicate the time span of the file content. The format is YYYY[MM[DD[HH[MM]]]], i.e. the year is represented by 4 digits, while the month, day, hour, and minutes are represented by exactly 2 digits, if they are present at all. In accordance with CMIP5, only those digits have to be included that are necessary to indicate the file content. The *StartTime* and *EndTime* of ~~sub-hourly~~ instantaneous data are based on the time values of the first and last record in the file, while averaged data use the left and right time_bnds values respectively. It is also allowed to use time values of the first and last records in NetCDF files for averaged data, however. The two dates are separated by a dash. All time stamps refer to UTC^o. Examples are found in Appendix C

VariableName (*CV*) is the name of the target variable in the NetCDF files. The CV is found in column B of sheet '*all*' of VR [1]^p.

Most of the DRS elements correspond to a mandatory NetCDF global attribute which must have the same value. These pairs of DRS element and global NetCDF attribute are listed in section C of sheet 'Global attributes' of VR [1].

5.2 DRS File naming

The names of the files in CORDEX archives are made up of the DRS elements described above. The elements are separated by underscores ('_') and must appear in the following order^q:

VariableName_Domain_GCModelName_CMIP5ExperimentName_CMIP5EnsembleMember_
RCMModelName_RCMVersionID_Frequency[_StartTime-EndTime].nc

Constant fields (Frequency=*fx*, e.g., *orog*) do NOT have the *StartTime-EndTime* element in their file names. The suffix 'nc' is mandatory

5.3 DRS Directory structure

The data have to be transferred to the CORDEX data nodes with the following directory structure^r:

```
<activity>/<product>/<Domain>/<Institution>/  
<GCMModelName>/<CMIP5ExperimentName>/<CMIP5EnsembleMember>/  
<RCMModelName>/<RCMVersionID>/<Frequency>/<VariableName>.
```

The elements defining the directory levels are described above in section 5.1. **Note that though the upper 2 levels are always the same (CORDEX/output) they are needed for the ESGF publication process.**

5.4 Time periods for each data file

For each CORDEX experiment, periods have been recommended for which the data should be delivered^s. However, other periods are acceptable as well. Also, downscaling products of other CMIP5 experiments than those defined by CORDEX are welcome (for example rcp26).

The time spans that have to be included into a single file depend on the aggregation, which is 3-hourly, 6-hourly, daily, monthly, seasonal, or invariant:

- 3-hourly or 6-hourly: one year,
- daily: 5 years or less,
- monthly or seasonal: 10 years or less,
- invariant: single file.

Files should always contain full years if the data are available. Exceptions are the seasonal data, which start with the first month of the first season that is available for the first year in the file and ends with the last month of the last season available for the last year in the file, and the StartTime at the beginning of the simulation as well as the EndTime at the end of the simulation.

Files with monthly data start with years that end with '1' or the first year of the experiment; they end with '0' or the last year of the experiment.

Files with seasonal data start with the winter of years that end with '1' and end with autumn of years that end with '0' or of the last year of the experiment.

Season file names therefore match the pattern *_???012-???011 if the content is neither from an initial nor from a final experiment year. The first file chunk with seasonal historical data matches *_194912-195011 or *_195003-195011, the last matches *_200511. The first file chunk with seasonal RCPs data matches *_200512-201011. The last chunk is *_210011.

Daily data files start with years that end with '1' or '6' or the first experiment year; the last year they contain ends with '5' or '0' or is the last experiment year.

Sub-daily (Tier2) data file names with averaged data have 2 options: the StartTime-EndTime element is either based on the first and last time value included in the file (as for instantaneous data) or on the earliest and latest time bounds values.

Examples for filenames are found in Appendix C .

6 Variables to output

The list of requested CORDEX variables is contained in Appendix B as well as in sheet *all* of VR [1]. The 3-hourly and 6-hourly data (Tier 2) have to be delivered as averages or instantaneous data according to the entries in column E of sheet *all* in VR ([1], see also section 4).

The sheet *all* also lists the *standard_name*, *long_name*, and *units* for the variables and the direction of fluxes. These are the same for all aggregations and are therefore not included in subsequent sheets (*3hr to fx*) in [1] which list for each frequency the subset of the variables that are requested with the respective frequency, or as invariant fields. They give as well the *cell_methods:time* attribute values which depend on the aggregation.

Note that there have been some minor changes of some long names since earlier versions of this document in order to increase consistency with CMIP5. These names are written in red in VR [1]. It is recommended to use the new version, however, both are acceptable.

The sheet *all* also lists the values of the CMIP5 attribute *realm*. This attribute is not required for CORDEX data, however, it is recommended to give it the CMIP5 value if used.

For fields with missing values, both attributes *_FillValue* and *missing_value* have to be defined with the value 1.e20.

Acknowledgments

Many people have contributed to these specifications, pointing out numerous inconsistencies and ambiguities. Particularly, Jamie Kettleborough from MOHC helped a lot to make the document consistent with what is produced when using the CMOR library. Many thanks to everyone!

References

- [1] http://is-enes-data.github.io/CORDEX_variables_requirement_table.pdf
- [2] <https://cordex.org/domains/cordex-domain-description/>
- [3] <http://cfconventions.org/latest.html>
- [4] http://is-enes-data.github.io/CORDEX_qc.pdf
- [5] https://pcmdi.llnl.gov/mips/cmip5/CMIP5_output_metadata_requirements.pdf

Tables

Table 1 CORDEX Domains

Definition of CORDEX equidistant longitude/latitude rotated grids with the position of the rotated North Pole given in geographic coordinates in columns 4 and 5. The rotated zero longitude is defined as going through the geographic North Pole. The number of cells is Nlon by Nlat. The last 4 columns define the centers of the outermost cells of the rectangular area in rotated coordinates. Column 3 gives the grid spacing in rotated coordinates.

CORDEX Area	Name	deg ⁸	NP ⁷ lon	NP ⁷ lat	N lon	N lat	West ⁷	East ⁷	South ⁷	North ⁷
South America	SAM-44	0.44	-56.06	70.60	146	167	143.92	207.72	-38.28	34.76
Central America	CAM-44	0.44	113.98	75.74	210	113	-52.80	39.16	-28.60	20.68
North America	NAM-44	0.44	83.00	42.50	155	130	-33.88	33.88	-28.40	28.36
Europe	EUR-44	0.44	-162.00	39.25	106	103	-28.21	17.99	-23.21	21.67
Africa	AFR-44	0.44	N/A	90.00	194	201	-24.64	60.28	-45.76	42.24
South Asia	WAS-44	0.44	-123.34	79.95	193	130	-32.12	52.36	-21.56	35.20
East Asia	EAS-44	0.44	-64.78	77.61	203	167	-40.92	47.96	-26.84	46.20
Central Asia	CAS-44	0.44	-103.39	43.48	153	100	-34.32	32.56	-20.68	22.88
Australasia	AUS-44	0.44	141.38	60.31	200	129	142.16	229.72	-22.88	33.44
Antarctica	ANT-44	0.44	-166.92	6.08	125	97	152.72	207.28	-27.72	14.52
Arctic	ARC-44	0.44	0.00	6.55	116	133	-22.88	27.72	-24.20	33.88
Mediterranean	MED-44	0.44	198.0	39.25	98	63	-23.22	19.46	-21.34	5.94
Middle East and North Africa	MNA-44	0.44	N/A	90.00	232	118	-26.40	75.24	-6.60	44.88
Middle East and North Africa high res.	MNA-22	0.22	N/A	90.00	464	236	-26.51	75.35	-6.71	44.99
Europe high res.	EUR-11	0.11	-162.0	39.25	424	412	-28.375	18.155	-23.375	21.835
South East Asia	SEA-22	0.22	N/A	90.00	264	194	89.26	147.12	-15.18	27.28

⁸ Grid spacing, position of rotated North Pole, and outermost cell centers in rotated coordinates

Table 2 Regular CORDEX Grids

Regular geographic grids corresponding to the CORDEX are as of [Table 1](#) for the additional delivery of the Core variables.

Area	Name	deg ⁹	Nlon	Nlat	West ⁸	East ⁸	South ⁸	North ⁸
South America	SAM-44i	0.5	181	155	-106.25	-16.25	-58.25	18.75
Central America	CAM-44i	0.5	207	111	-124.75	-21.75	-19.75	35.25
North America	NAM-44i	0.5	300	129	-171.75	-22.25	12.25	76.25
Europe	EUR-44i	0.5	221	103	-44.75	65.25	21.75	72.75
Africa	AFR-44i	0.5	173	179	-25.25	60.75	-46.25	42.75
South Asia	WAS-44i	0.5	195	124	19.25	116.25	-15.75	45.75
East Asia	EAS-44i	0.5	227	157	62.75	175.75	-18.75	59.25
Central Asia	CAS-44i	0.5	260	133	10.75	140.25	17.75	69.75
Australasia	AUS-44i	0.5	238	133	88.75	207.25	-53.25	12.75
Antarctica	ANT-44i	0.5	720	70	-179.75	179.75	-89.75	-55.25
Arctic	ARC-44i	0.5	720	83	-179.75	179.75	48.75	89.75
Mediterranean	MED-44i	0.5	144	65	-20.75	51.75	25.25	57.25
Middle East and North Africa	MNA-44i	0.5	206	106	-26.75	75.75	-7.25	45.25
" - high res.	MNA-22i	0.25	410	209	-26.625	75.625	-6.875	45.125
Europe high res.	EUR-11i	0.125	881	408	-44.8125	65.1875	21.8125	72.6875
South East Asia	SEA-22i	0.25	233	172	89.125	147.125	-15.375	27.3750

⁹ Grid spacing, and outermost cell position in geographic coordinates

Appendix A Examples of NetCDF headers

Example with mandatory global attributes only

```
// global attributes:
:contact = "rossby.data@smhi.se" ;
:creation_date = "2012-05-31-T16:01:08Z" ;
:experiment_id = "evaluation" ;
:driving_model_id = "ECMWF-ERAINT" ;
:driving_model_ensemble_member = "r1i1p1" ;
:driving_experiment_name = "evaluation" ;
:frequency = "day" ;
:institute_id = "SMHI" ;
:model_id = "SMHI-RCA4" ;
:rcm_version_id = "v1" ;
:project_id = "CORDEX" ;
:CORDEX_domain = "AFR-44" ;
:product = "output" ;
```

A file with the above attributes would have the name *VariableName_AFR-44_ECMWF-ERAINT_evaluation_r1i1p1_SMHI-RCA4_v1_Frequency_StartTime-EndTime.nc*

Example with mandatory and optional global attributes

```
// global attributes:
:Conventions = "CF-1.4" ;
:contact = "rossby.data@smhi.se" ;
:creation_date = "2012-05-31-T16:01:08Z" ;
:experiment = "Evaluation run with reanalysis forcing" ;
:experiment_id = "evaluation" ;
:driving_experiment = "ECMWF-ERAINT, evaluation, r1i1p1" ;
:driving_model_id = "ECMWF-ERAINT" ;
:driving_model_ensemble_member = "r1i1p1" ;
:driving_experiment_name = "evaluation" ;
:frequency = "day" ;
:institution = "Swedish Meteorological and Hydrological
Institute" ;
:institute_id = "SMHI" ;
:model_id = "SMHI-RCA4" ;
:rcm_version_id = "v1" ;
:project_id = "CORDEX" ;
:CORDEX_domain = "AFR-44" ;
:product = "output" ;
:references = "http://www.smhi.se/en/Research/Research-departments" ;
:tracking_id = "d3afe777-5792-4e07-8781-ca469c0078ec" ;
:rossby_run_id = "201106" ;
:rossby_grib_path = "/rossby/joint_exp/cordex/201106/raw/" ;
```

A file with the above attributes would have the name *VariableName_AFR-44_ECMWF-ERAINT_evaluation_r1i1p1_SMHI-RCA4_v1_Frequency_StartTime-EndTime.nc*

Coordinate attributes for pressure

```
double plev;
    plev:long_name = "pressure";
    plev:units = "Pa" ;
    plev:axis = "Z" ;
    plev:positive = "down" ;
    plev:standard_name = "air_pressure" ;
```

Coordinate attributes for height

```
double height ;
    height:long_name = "height" ;
    height:standard_name = "height" ;
    height:units = "m" ;
    height:positive = "up" ;
    height:axis = "Z" ;
```

Coordinate attributes for cloud amounts within pressure layers

```
double plev
    plev:long_name = "pressure" ;
    plev:standard_name = "air_pressure";
    plev:units = "Pa" ;
    plev:axis = "Z" ;
    plev:positive = "down" ;
    plev:bounds = "plev_bnds";

double plev_bnds(2)
```

where plev has the values 22000., 56000., or 84000., with plev_bnds [44000., 0.],[68000., 44000.], and [100000., 68000.] respectively.

Example with 'monthly mean of daily min'

```
float tasmin(time, rlat, rlon);
    tasmin:standard_name= "air_temperature";
    tasmin:long_name = "Daily Minimum Near-Surface Air Temperature";
    tasmin:units = "K" ;
    tasmin:cell_methods = "time: minimum within days time: mean over days";
    tasmin:grid_mapping = "rotated_pole";
    tasmin:coordinates = "lon lat height";
    tasmin:_FillValue = 1.e+20f
    tasmin:missing_value = 1.e+20f;
```

Example with Rotated Pole coordinates for AFR-XX

```
char rotated_pole ;
    rotated_pole:grid_mapping_name = "rotated_latitude_longitude";
```

```

    rotated_pole:grid_north_pole_latitude = 90.;
    rotated_pole:grid_north_pole_longitude = -180.;

double rlon(rlon) ;
    rlon:standard_name = "grid_longitude" ;
    rlon:long_name = "longitude in rotated pole grid" ;
    rlon:units = "degrees" ;
    rlon:axis = "X" ;

double rlat(rlat) ;
    rlat:standard_name = "grid_latitude";
    rlat:long_name = "latitude in rotated pole grid";
    rlat:units = "degrees";
    rlat:axis = "Y";

double lon(y, x) ;
    lon:standard_name = "longitude" ;
    lon:long_name = "longitude" ;
    lon:units = "degrees_east" ;

double lat(y, x) ;
    lat:standard_name = "latitude" ;
    lat:long_name = "latitude" ;
    lat:units = "degrees_north" ;

```

Example with Lambert Conformal projection for EUR-XX

```

charLambert_Conformal;
    Lambert_Conformal:grid_mapping_name = "lambert_conformal_conic";
    Lambert_Conformal:longitude_of_central_meridian = "11.5";
    Lambert_Conformal:latitude_of_projection_origin = "50.4";
    Lambert_Conformal:standard_parallel = "11.5";

double x(x);
    x:units = "km";
    x:long_name = "x coordinate of projection";
    x:standard_name = "projection_x_coordinate";
    x:axis = "X" ;

double y(y);
    y:units = "km";
    y:long_name = "y coordinate of projection";
    y:standard_name = "projection_y_coordinate";
    y:axis = "Y";

double lon(y, x) ;
    lon:standard_name = "longitude" ;
    lon:long_name = "longitude" ;
    lon:units = "degrees_east" ;

double lat(y, x) ;
    lat:standard_name = "latitude" ;
    lat:long_name = "latitude" ;
    lat:units = "degrees_north" ;

```

Example with Regular Geographic coordinates

```

double lon(lon);
    lon:standard_name = "longitude";
    lon:long_name = "longitude";
    lon:units = "degrees_east";
    lon:axis = "X";

double lat(lat);
    lat:standard_name = "latitude";
    lat:long_name = "latitude";
    lat:units = "degrees_north";
    lat:axis = "Y";

```

Example with Scalar Dimension (height or plev)

```

dimensions:
    rlon = 424 ;
    rlat = 412 ;
    time = UNLIMITED ; // (1826 currently)
    bnds = 2 ;
variables:
    double height;
        height:axis = "Z";
        height:long_name = "height";
        height:positive = "up";
        height:standard_name = "height";
        height:units = "m";
    float tas(time, rlat, rlon);
        tas:grid_mapping = "rotated_pole";
        tas:_FillValue = 1.e+20f;
        tas:missing_value = 1.e+20f;
        tas:standard_name = "air_temperature";
        tas:long_name = "Near-Surface Air Temperature";
        tas:units = "K";
        tas:coordinates = "lon lat height";
        tas:cell_methods = "time: mean";

```

Appendix B List of CORDEX variables

<i>Description</i>	<i>Short name</i>
Near-Surface ¹⁰ Air Temperature	<i>tas</i>
Daily-Maximum Near-Surface ¹⁰ Air Temperature	<i>tasmax</i>
Daily-Minimum Near-Surface ¹⁰ Air Temperature	<i>tasmin</i>
Precipitation	<i>pr</i>
Surface Pressure	<i>ps</i>
Mean Sea Level Pressure	<i>psl</i>
Near-Surface ¹⁰ Relativ Humidity (variable added later)	<i>hurs</i>
Near-Surface ¹⁰ Specific Humidity	<i>huss</i>
Near-Surface ¹⁰ Wind Speed	<i>sfcWind</i>
Daily-Maximum Near-Surface ¹⁰ Wind Speed	<i>sfcWindmax</i>
Total Cloud Cover	<i>clt</i>
Sunshine Hours (duration when surface solar radiation flux exceeds 120 W/ m ²)	<i>sund</i>
Surface Downwelling Shortwave Radiation	<i>rsds</i>
Surface Downwelling Longwave Radiation	<i>rlds</i>
Surface Latent Heat Flux	<i>hfls</i>
Surface Sensible Heat Flux	<i>hfss</i>
Upwelling Surface Shortwave Radiation	<i>rsus</i>
Upwelling Longwave radiation	<i>rlus</i>
Surface Evaporation	<i>evspsbl</i>
Potential Evapotranspiration	<i>evspsblpot</i>
Soil Frozen Water Content	<i>mrfs0</i>
Surface Runoff	<i>mrros</i>
Total Runoff	<i>mrro</i>
Total Soil Moisture Content	<i>mrso</i>
Surface Snow Amount	<i>snw</i>
Snow Melt	<i>snm</i>
Maximum 1-hour Precipitation Rate within 24 hour period	<i>prhmax</i>

Convective Precipitation	<i>prc</i>
TOA Outgoing Longwave Radiation	<i>rlut</i>
TOA Incident Shortwave Radiation	<i>rsdt</i>
TOA Outgoing Shortwave Radiation	<i>rsut</i>
Eastward Near-Surface¹⁰ Wind	<i>uas</i>
Northward Near-Surface¹⁰ Wind	<i>vas</i>
Maximum Near-Surface¹⁰ Gust Wind Speed	<i>wsgsmax</i>
Surface Downward Eastward Wind Stress	<i>tauu</i>
Surface Downward Northward Wind Stress	<i>tauv</i>
Surface (Skin) Temperature	<i>ts</i>
Atmospheric Boundary Layer Thickness	<i>zmla</i>
Column Water Vapour	<i>prw</i>
Column Condensed (liquid+ice) Water Content	<i>clwvi</i>
Column Ice Water Content	<i>clivi</i>
Zonal (eastward) Wind at 850 hPa	<i>ua850</i>
Meridional (northward) Wind at 850 hPa	<i>va850</i>
Temperature at 850 hPa	<i>ta850</i>
Specific Humidity at 850 hPa	<i>hus850</i>
Zonal (eastward) Wind at 500 hPa	<i>ua500</i>
Meridional (northward) Wind at 500 hPa	<i>va500</i>
Geopotential Height at 500 hPa	<i>zg500</i>
Temperature at 500 hPa	<i>ta500</i>
Zonal (eastward) Wind at 200 hPa	<i>ua200</i>
Meridional (northward) Wind at 200 hPa	<i>va200</i>
Temperature at 200 hPa	<i>ta200</i>
Geopotential Height at 200 hPa	<i>zg200</i>
High Clouds (p<440hPa)	<i>clh</i>
Medium Clouds (680 hPa > p > 440 hPa)	<i>clm</i>

¹⁰ Near-Surface means at a height between 1.5 to 10.0 m.

Low Clouds (p>680hPa); the lower plev_bnds value of <i>cfl</i> has to be 100000Pa; actually the cloud fraction between 68000Pa and the surface is requested	<i>cfl</i>
Snow Area Fraction	<i>snc</i>
Snow Depth	<i>snd</i>
Sea Ice Area Fraction	<i>sic</i>
Snowfall Flux	<i>prsn</i>
Atmosphere Grid-Cell Area	<i>areacella</i>
Surface Altitude	<i>orog</i>
Land Area Fraction	<i>sftlf</i>
Fraction of Grid Cell Covered with Glacier	<i>sftgif</i>
Capacity of Soil to Store Water	<i>mrsofc</i>
Maximum Root Depth	<i>rootd</i>

Appendix C Examples of file names

File names:

orog_EUR-44_ECMWF-ERAINT_evaluation_r0i0p0_SMHI-RCA4_v1_fx.nc
orog_EUR-44_ECMWF-ERAINT_evaluation_r1i1p1_SMHI-RCA4_v1_fx.nc
tas_EUR-44_ECMWF-ERAINT_evaluation_r1i1p1_SMHI-RCA4_v1_sem_198812-199011.nc
tas_AFR-44_CMCC-CMCC-CM_historical_r1i1p1_CLMcom-CCLM4-8-17_v1_sem_194912-195011.nc
tas_AFR-44_CMCC-CMCC-CM_historical_r1i1p1_CLMcom-CCLM4-8-17_v1_sem_195003-196011.nc
tas_AFR-44_CMCC-CMCC-CM_historical_r1i1p1_CLMcom-CCLM4-8-17_v1_sem_200012-200511.nc
tas_AFR-22_MPI-MPI-ESM-LR_historical_r1i1p1_CLMcom-CCLM4-8-17_v1_mon_200101-200512.nc
tas_EUR-11_CNRM-CERFACS-CNRM-CM5_rcp45_r1i1p1_DMI-HIRHAM5_v1_day_20060101-20101231.nc
mrro_EUR-11_MPI-MPI-ESM-LR_rcp26_r2i1p1_MPI-CSC-REMO2009_radv_6hr_2091010100-2092010100.nc
snw_AFR-44_ICHEC-EC-EARTH_rcp85_r1i1p1_SMHI-RCAO-SN_v1_6hr_2091010100-2091123118.nc
psl_EUR-11_MPI-MPI-ESM-LR_rcp26_r2i1p1_MPI-CSC-REMO2009_radv_3hr_2091010100-2091123121.nc
pr_EUR-11_MPI-MPI-ESM-LR_rcp26_r2i1p1_MPI-CSC-REMO2009_radv_3hr_2091010100-2092010100.nc
pr_EUR-11_MPI-MPI-ESM-LR_rcp26_r2i1p1_MPI-CSC-REMO2009_radv_3hr_209101010130-209101012230.nc

The *pr* data file name containing averaged data may have one of the two names given above. The reason is that, when using CMOR, the StartTime-EndTime element is based on the first and last time value included in the file (as for instantaneous data) while earlier versions of the specifications said that it should be the earliest and latest time bound values.

Endnotes

^aThere are two tools available for automated quality assurance. They check the data for their compliance with the specifications outlined in this document.

One is available from an SVN server at BADC

(<http://proj.badc.rl.ac.uk/exarch/wiki/PackageFCC>,

<http://proj.badc.rl.ac.uk/svn/>). Revisions older than 47, however, check only a part of the specifications.

Another tool is available upon request from a DKRZ SVN server (contact Heinz-Dieter hollweg@dkrz.de). A switch is being implemented that activates exactly the compliance rules laid down here. This tool can be configured to check additional data aspects. In particular it can check the consistency of time records (strict monotonicity, correct and equal time steps, no gaps; suspicion of duplicated data; etc.) and will point out unexpected features in the data values.

^b When the CORDEX archive specifications were first written, it had not yet been decided that the central archive will be an ESGF archive. Some CMIP5 specifications were not considered relevant for CORDEX (e.g. the *realm* attribute), and therefore the adoption of CMIP5 specifications is not always fully and correctly realized.

^c In order to get access to Tier-2 data contact please the data providers to get to know access conditions and methods. Contact points may be found in the NetCDF attribute *contact* of the *Core* or *Tier1* files.

^dThe NetCDF4 'classic data model' allows reading the data using the same subroutine calls as for NetCDF 3 files. Note that this means that the files must not be compressed afterwards, e.g. with gzip. NetCDF 3 files can be converted to NetCDF 4 compressed with "nccopy -k 4 -d 1 -s infile outfile".

^e The CF-1.4 convention or later has to be followed not only with regard to global attributes, but also with regard to the coordinate variables and their attributes. The specific version number has to be supplied if the *conventions* attribute is given. For more details on CF conventions see <http://cf-pcmdi.llnl.gov>.

^f This statement applies with the reservation that the full set of CORDEX files has not yet been produced with CMOR. The case could be that modifications of the CMOR software are required.

^g If any doubt remains about CORDEX file contents additional information can be provided in the global or variable attributes *comment*. Also, it is possible to use any global attribute that is not a CORDEX attribute. In order to avoid confusion among the data users it is recommended to only use CMIP5 global attributes if the meaning corresponds to the one in CMIP5. Therefore in CORDEX an attribute *realm*, for example, should only be used with the values as defined in CMIP5.

^h The computational grids have to be quasi-uniform. The grid spacing should be at least about 50-km for all domains, which corresponds to that of the rotated-pole grid with a regular 0.44° grid spacing in both directions. It is recommended to perform runs with this resolution first. For Africa 50-km simulations is indeed requested. Additional simulations with refined grids are welcome.

The grid spacing for refined grids is recommended to be for all domains any power of 2 (i.e. 50-km, 25-km, 12.5-km, 6.25-km, etc.) in both horizontal directions. If the rotated-pole grid is used this corresponds to a regular 0.44°, 0.22°, 0.11°, etc. grid.

Geographic latitude/longitude coordinates are specified by assigning an attribute *coordinate* to the target variable that points to coordinate variables *lat* and *lon* (`<variableName>: coordinates=" lon lat"`), and including coordinate variables named *lat* and *lon* with the same dimensions as the grid, that contain the latitudes and longitudes respectively of the native grid cell centers.

ⁱ The constraint that data on a quasi-uniform grid of approximately 50 km grid spacing (e.g. the 0.44-degree rotated pole grid) have to be interpolated to a geographic latitude/longitude grid of which all whole-number latitudes or longitudes align with cell boundaries, means that the grid is a 0.5 degree grid with grid point centers at -0.75, -0.25, +0.25 etc. in order to make 0 deg. etc. be exactly at a cell boundary. Accordingly, the regular lat/lon grid corresponding to a 12.5-km quasi-uniform grid (e.g. a 0.11-degree rotated pole grid) has its cell centers at -0.1875, -0.0625, 0.0625, 0.1875, etc.

^j Using the time coordinate unit '*days since 1949-12-01T00:00:00Z*' for all data means that the time values of sub-daily data are fractional. Because *1949-12-01T00:00:00Z* is the earliest starting time among the simulations, all time values will be positive. The values of the variable attribute *cell_methods* depend on whether the data are representative for the period spanned by *time_bnds* (e.g. `<variable>:cell_methods="time: minimum"`), or whether the variable values represent instantaneous output (`<variable>:cell_methods="time: point"`).

The basetime specification (after 'since') must have a format according to the UNIDATA conventions which is part of the CF conventions. E.g. the time (after the T) may be omitted. If it exists it must be zero, has to have 2, 5, or 8 digits, and has to use the colon as separator. The date separator has to be a dash. Date and time have to be separated by either a 'T' or a blank. The Z may be omitted.

Please take care with monthly means of daily maxima or minima: They should have the attribute *cell_methods="time: maximum within days time: mean over days"* and the corresponding attributes for minima. Sunshine duration averaged over longer periods should have *cell_methods="time: sum within days time: mean over days"*.

^k The number of bounds (2 in the case of time coordinate) is always the one that varies the fastest (CF-convention). The order of dimensions that should be displayed by `ncdump -h file.ncis "time_bnds(ntimes,2)"`.

^l Note that the *experiment_id* of experiments forced by reanalysis data is 'evaluation' and NOT 'historical', because the latter term is reserved for GCM-driven simulations forced with data from CMIP5 *historical* experiments.

ⁿ The discrimination between small and major upgrades is left to the individual institution. Most institutions use consecutive numbering like *v1*, *v2*, ... , or use *v1* in case of a single *RCMVersionID*. Generally, all strings are equivalent, the value is free text, besides that, as for all other DRS elements, only valid DRS characters are allowed. Note that for each *RCMModelName*, a model description will be requested. Differences between equal *RCMModelName* models with different *RCMVersionID* should be noted in a global NetCDF attribute *comment* or in the *source* attribute.

^o A *StartTime* value of e.g. '1979010100' indicates '00 UTC on 1 Jan 1979'. For monthly and seasonal data, the specification of the day is not necessary. Consequently, *StartTime* and *EndTime* have the form YYYYMM. The day or hour indicators are not allowed. In the case of daily data, the format is YYYYMMDD, while for sub-daily data the hour indicator and possibly even the minutes have to be included (YYYYMMDDHH[mm]). Instantaneous 3-hourly data also require the specification of the minutes (YYYYMMDDHHMM). *EndTime* at a day limit is denoted by YYYYMMDP00[00] where YYYYMMDP is the day following YYYYMMDD.

^p If applicable the CMIP5 standard_output.xls variable name is used for the target variable name of a file. In contrast to CMIP5, all CORDEX variables are two-dimensional. Variables that are requested on more than one level have the level value attached to their names (e.g. ua500 is eastward wind at 500 hPa).

^q It has to be noted that the CMIP5 DRS file names have fewer levels than those of CORDEX. The CMIP5 file name build rule is:
<variable name>_<MIP table>_<model>_<experiment>_<ensemble member>[_<temporal subset>][_<geographical info>].nc

^r The upper two directory levels are included here for compliance with the CMIP5 specification. They are needed by the ESGF in the case that there are data from more than one *activity* in the archive. For comparison we include here the CMIP5 DRS directory structure (data delivery):

<activity>/<product>/<institute>/<model>/<experiment>/<frequency>/<modeling realm>/<MIP table>/<ensemble member>/<version number>/<variable name>

^s The advised CORDEX downscaling periods are 1989-2008 for the evaluation runs, 1950-2005 for the historical CMIP5 experiment, and 2006-2100 for the RCP 4.5 and RCP 8.5 projections. All periods are inclusive. It will not be a hindrance to ESGF publication, though, if the advised periods have not always been adhered to.

However, before the data is submitted, detailed information about any deviation should be sent to the ESGF data node management. In order that the seasonal mean data start with a winter season, the evaluation run and the historical experiment may start in December of the previous year.

If CORDEX groups do not have enough computational resources to simulate the full advised RCP period from 2005 to 2100, it is advised to simulate 30-year time slices in the following order of priority 1981-2010, 2041-2070, 2011-2040, 2071-2100, and 1951-1980.