

# Data Reference Syntax (DRS) for Regional ReAnalysis (RRA)

Grigory Nikulin<sup>1</sup>, Lars Barring<sup>1</sup>, Martin Juckes<sup>2</sup> and Michael Kolax<sup>1</sup>

<sup>1</sup> Swedish Meteorological and Hydrological Institute (SMHI, Sweden)

<sup>2</sup> Science and Technology Facilities Council (STFC, UK)

11th June 2015

This document provides Data Reference Syntax (DRS) elements necessary for publishing Regional ReAnalysis (RRA) on the Earth System Grid Federation (ESGF). The document includes file and directory naming convention, global attributes and ESGF Search Facets Mappings. The DRS elements can assume values pre-defined by Controlled Vocabularies (CV), free text or free text with build rules.

## 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.

**VariableName** (CV) is the short name of the variable. The name is taken from the [CMIP5](#) or [CORDEX](#) output tables.

**Domain** (CV or free text) is the short name assigned to a region covered by RRA (global attribute **region\_id**) and includes a flag for resolution separated by a '-' (dash). The domain name (**region\_id**) for national RRA is taken from the [ISO 3166 alpha 3 table](#), a three-letter country code i.e. FRA – France, DEU – Germany (CV). The domain name for RRA covering large geographical regions is taken from [the CORDEX archive specifications](#) (see Table 1) if such or similar region is available, i.e. EUR - the Euro-Atlantic sector, NAM - North America (CV) or free text otherwise. The domain name for RRAs remapped onto one of the CORDEX grids is taken from [the CORDEX archive specifications](#) (CV). Additional, more detailed information about the domain is provided by the global attributes “**region**” and “**region\_lexicon**”. For example in case of a region similar to the Euro-CORDEX domain “**region**=Europe (approx)”, **region\_id**=EUR and “**region\_lexicon**=CORDEX”.

The flag for resolution precisely defines resolution as it is in degree or kilometre depending on projection used in RRA. For kilometre the resolution is presented by round numbers (i.e. 10km - 10, 5km - 5) while for degree by digits after “0.” (i.e. 0.44° - 44, 0.05° - 05 and 0.025° - 025). Information about resolution units is provided by the global attribute “**resolution\_units**” (deg or km).

**RADrivingName** (CV) is an identifier of the driving ReAnalysis (RA) which can be global or regional one. The name consists of an institute identifier and a model identifier. The two parts of the name are separated by a '-' (dash), i.e. ECMWF-ERAINT (defined in CORDEX CV) or SMHI-HIRLAM. Dashes in either of the two parts are allowed.

**Product** is “RegRean” in file names, directory structure and the corresponding global attribute while “Regional Reanalysis” in the ESGF search interface.

**RADrivingEnsembleMember** (free text with build rules) identifies the ensemble member and sequence of the driving RA in form of v<N>d<M> where v<N> (or any letter with digits) represents the ensemble member and d<M> represents sequent number of the driving RA (v1d1, v2d1, v1d2, v2d2 etc.). For each subsequent driving RA this DRS element is extended by attaching the subsequent driving RA ensemble member to the previous one separated by '-' (i.e. v1d1-v1d2, v2d1-v1d2). Additional information describing sequence of the driving RAs and how the ensemble members are generated is provided by the global attribute **driving\_ensemble\_description**. If a global driving reanalysis - RA (0.75deg) has two members (v1, v2) and both members are downscaled by the first regional reanalysis RRA1 (0.22deg), which in turn downscaled further by the second RRA2 (0.05deg) file names are:

RRA1:

pr\_EUR-22\_RA\_RegRean\_v1d1\_RRA1\_v1\_day\_20010101-20051231.nc

pr\_EUR-22\_RA\_RegRean\_v2d1\_RRA1\_v1\_day\_20010101-20051231.nc

RRA2

pr\_EUR-05\_RRA1\_RegRean\_v1d1-v1d2\_RRA2\_v1\_20010101-20051231.nc

pr\_EUR-05\_RRA1\_RegRean\_v2d1-v1d2\_RRA2\_v1\_20010101-20051231.nc

Example for the [MESAN regional reanalysis \(the EURO4M project\)](#): two driving reanalyses with one member each: (global ECMWF-ERAINT (0.75deg) -> regional SMHI-HIRLAM (0.22deg) -> regional SMHI-MESAN (0.05deg) )

pr\_EUR-22\_ECMWF-ERAINT\_RegRean\_v1d1\_SMHI-HIRLAM\_v1\_day\_20010101-20051231.nc

pr\_EUR-05\_SMHI-HIRLAM\_RegRean\_v1d1-v1d2\_SMHI-MESAN\_v1\_day\_20010101-20051231.nc

Invariant fields (land sea mask, orography etc. with frequency=*fx*) should have the value *v0d0* (orog\_EUR-05\_SMHI-HIRLAM\_RegRean\_v0d0\_SMHI-MESAN\_v1\_fx.nc)

**RRAModelName** (CV) is an identifier of the RRA modelling system. It consists of the institution identifier and a model (reanalysis) acronym, connected by a dash (e.g. SMHI-HIRLAM or SMHI-MESAN).

**RRAEnsembleMember** (free text with build rules) describes the ensemble member of the RRA in form of v<N> (or any letter with digits). Additional information describing how ensemble members are generated is provided by the global attribute **ensemble\_description**.

**Frequency** (CV) is the output frequency indicator as in CMIP5 or CORDEX: *3hr*=3 hourly, *6hr*=6 hourly, *day*=daily, *mon*=monthly, and *fx*=invariant fields.

**StartTime** and **EndTime** 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. Constant fields (Frequency=*fx*) do not have the *StartTime-EndTime* element in their file names.

**activity** (single value), the same as global attribute **project\_id**

## 2. DRS file naming

*VariableName\_Domain\_RADrivingName\_Product\_RADrivingEnsembleMember\_RRAModelName\_RRAEnsembleMember\_Frequency[\_StartTime-EndTime].nc*

## 3. Directory structure

*<activity>/<product>/<Domain>/<Institute>/<RADrivingName>/<RADrivingEnsembleMember>/<RRAModelName>/<RRAEnsembleMember>/<Frequency>/<VariableName>*

## 4. Global attributes

**Conventions:** 'CF-1.6'

**contact:** contact information (name, e-mail)

**creation\_date:** a string representation of the date when the file was created in the format "YYYY-MM-DD-THH:MM:SSZ"

**domain:** the short name of a region covered by RRA including a flag for resolution (the same as *Domain*)

**region:** provides more information about a region covered by RRA

**region\_id:** an identifier for region covered by RRA

**region\_lexicon** provides information what lexicon is used for the short name assigned to a region covered by RRA (i.e. CORDEX, ISO 3166 alpha 3)

**frequency:** output frequency indicator (the same as *Frequency*)

**driving\_reanalysis:** full name of the driving reanalysis

**driving\_reanalysis\_id:** an identifier of the driving reanalysis (the same as *RADrivingName*)

**driving\_reanalysis\_ensemble\_member:** identifies the ensemble member and sequence of the driving RA (the same as *RADrivingEnsembleMember*);

**driving\_ensemble\_description:** provides detailed information on how the driving RA ensemble members are generated

**institution:** full name of institution that is responsible for RRA

**institute\_id:** a short acronym for the institution responsible for RRA

**reanalysis\_id:** is an identifier of the RRA (the same as *RRAModelName*)

**reanalysis\_ensemble\_member:** identifies the ensemble member of the RRA (the same as *RRAEnsembleMember*)

**ensemble\_description:** provides detailed information on how the RRA ensemble members are generated

**project\_id:** an identifier of project

**resolution\_units:** provides resolution units “deg” or “km”

**product:** type of data - “RegRean”

**realm:** following CMIP5 this attribute indicates the portion of the earth system for which the variable is particularly relevant. For CMIP5, permitted values are: “atmos”, “ocean”, “land”, “landIce”, “seaIce”, “aerosol” “atmosChem”, ocnBgchem (ocean biogeochemical).

**references:** reference to RRA (web pages , publications etc.)

**tracking\_id:** a character string that is almost certainly unique to this file and must be generated using the OSSP utility which supports a number of different DCE 1.1 variant UUID options. Download the software from <http://www.oss.org/pkg/lib/uuid/>. The tracking\_id might look something like: a940beb8-6925-45d6-9178-dfd35df5e5b6.

**mip\_specs:** “CORDEX” (indicates which model intercomparison project output specifications have been followed)

## 5. Optional Global attributes

**Data\_Assimilation:** provides more details on what kind of data assimilation is used

## 6. ESGF Search Facets Mappings

ESGF Search Facet	ESGF THREDDS dataset property	Attribute in netcdf File
<b>Project</b>	<i>activity</i>	project_id
<b>Institute</b>	<i>Institute</i>	institute_id
<b>Product</b>	<i>product</i>	product
<b>Time Frequency</b>	<i>Frequency</i>	frequency
<b>Variable</b>	<i>VariableName</i>	
<b>Reanalysis</b>	<i>RRAModelName</i>	reanalysis_id
<b>Reanalysis Ensemble</b>	<i>RRAEnsembleMember</i>	reanalysis_ensemble_member
<b>Driving Reanalysis</b>	<i>RADrivingName</i>	driving_reanalysis_id

<b>Driving Ensemble</b>	<i>RADrivingEnsembleMember</i>	driving_reanalysis _ensemble_member
<b>Domain</b>	<i>Domain</i>	domain

Acknowledgements: Development of this Data Reference Syntax for regional reanalysis has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 607418 (CLIP-C). We also thank Stephanie Legutke (DKRZ) for valuable comments and suggestions.

## Examples of netcdf files

```
netcdf tasmax_EUR-05_SMHI-HIRLAM_RegRean_v1d1-v1d2_SMHI-MESAN_v1_day_19890101-19890102
{
dimensions:
  bnds = 2 ;
  time = UNLIMITED ; // (2 currently)
  rlon = 1051 ;
  rlat = 1000 ;

variables:
  double height ;
    height:axis = "Z" ;
    height:long_name = "height" ;
    height:positive = "up" ;
    height:standard_name = "height" ;
    height:units = "m" ;
  double time_bnds(time, bnds) ;
  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" ;
  char rotated_pole ;
    rotated_pole:grid_mapping_name = "rotated_latitude_longitude" ;
    rotated_pole:grid_north_pole_latitude = 29. ;
    rotated_pole:grid_north_pole_longitude = -165. ;
  double time(time) ;
    time:standard_name = "time" ;
    time:units = "days since 1949-12-01 00:00:00" ;
    time:calendar = "standard" ;
    time:long_name = "time" ;
    time:bounds = "time_bnds" ;
    time:axis = "T" ;
  float tasmax(time, rlat, rlon) ;
    tasmax:grid_mapping = "rotated_pole" ;
    tasmax:_FillValue = 1.e+20f ;
    tasmax:missing_value = 1.e+20f ;
    tasmax:standard_name = "air_temperature" ;
    tasmax:long_name = "Daily Maximum Near-Surface Air Temperature" ;
    tasmax:units = "K" ;
    tasmax:coordinates = "lon lat height" ;
    tasmax:cell_methods = "time: maximum" ;
    tasmax:comment = "12-hour day-time maximum temperature: 06Z - 18Z" ;
  double lon(rlat, rlon) ;
    lon:standard_name = "longitude" ;
    lon:long_name = "longitude" ;
    lon:units = "degrees_east" ;
  double lat(rlat, rlon) ;
    lat:standard_name = "latitude" ;
    lat:long_name = "latitude" ;
    lat:units = "degrees_north" ;

// global attributes:
  :Conventions = "CF-1.6" ;
  :contact = "This is a pilot release and no support is provided" ;
  :comment = "The regional reanalysis MESAN has been generated in the EURO4M FP7
project (http://www.euro4m.eu/ Grant agreement no. 242093). This MESAN netcdf release
has been prepared in the CLIPC FP7 Copernicus precursor project (http://www.clipc.eu/
Grant agreement no: 607418)." ;
  :creation_date = "2015-03-24-T18:48:00Z" ;
  :domain = "EUR-05" ;
  :frequency = "day" ;
  :driving_reanalysis = "High Resolution Limited Area Model (HIRLAM) 3D-VAR
dynamical downscaling regional reanalysis; http://www.euro4m.eu/downloads/D2.3\_HIRLAM-
based\_reanalysis\_dataset\_at\_25\_km\_for\_most\_of\_the\_past\_20\_yr\_period.pdf" ;
  :driving_reanalysis_id = "SMHI-HIRLAM" ;
```

```
:driving_reanalysis_ensemble_member = "v1d1-v1d2" ;
:driving_ensemble_description = "The ERA-Interim reanalysis (v1d1) is downscaled
by SMHI-HIRLAM to 0.22deg (v1d2) and then SMHI-HIRLAM is downscaled by SMHI-MESAN to
0.05deg." ;
:institution = "Swedish Meteorological and Hydrological Institute" ;
:institute_id = "SMHI" ;
:reanalysis_id = "SMHI-MESAN" ;
:reanalysis_ensemble_members = "v1" ;
:ensemble_description = "N/A" ;
:project_id = "CLIPC" ;
:region = "Europe (approx)" ;
:region_id = "EUR" ;
:region_lexicon = "CORDEX" ;
:resolution_units = "degree" ;
:references =
"http://www.euro4m.eu/downloads/D2.4_Improved_MESAN_downscaled_dataset_at_3-
12_km_resolution_for_most_of_the_past_20_yr_period.pdf" ;
:realm = "atmos" ;
:mip_specs = "CORDEX" ;
:product = "RegRean " ;
:tracking_id = "44f80420-fd24-4314-a914-8abacfd3e801" ;
```