# IS-ENES2 Datanode Administrator's Reference Manual

Editor: Prashanth Dwarakanath, NSC, Sweden

Version: 2014-1.03

# Contents

# 1 Purpose and limitations

The purpose of this document is to serve as an unambigious single resource for reference by administrators of IS-ENES2 ESGF datanodes, to configure their datanodes and publish data in compliance with regulations discussed and adopted by all datanode managers. This document aggregates information from sources such as the Trieste meeting notes [2], Martin Juckes' 'CORDEX: ESGF Search Facet Mappings' document [1] and other discussions which have led to collective consensus. This document only contains information from the perspective of publishing/maintaining data on the ESGF datanode and may not be refered to for any other purpose.

# 2 Latest Version

The latest version of this document will always be available at:
https://github.com/snic-nsc/datanode-mgr-doc/raw/master/ro/Datanodemgr-doc.pdf
The entire repository, which includes the LaTeX source file can be cloned from:
https://github.com/snic-nsc/datanode-mgr-doc.git

# 3 IS-ENES2 ESGF datanode Search Facet Configuration

IS-ENES2 ESGF datanodes have some additional search facets pertaining to CORDEX. Here below are the entire list of facets used, in an IS-ENES2 ESGF datanode

Filename: `facets.properties`
Standard location: `/esg/config/facets.properties`

```
project=0:Project:optional_project_description
institute=1:Institute:optional_institute_description
model=2:Model:optional_model_description
source_id=3:Instrument:optional_instrument_description
experiment_family=4:Experiment Family:optional_experiment_family_description
experiment=5:Experiment:optional_experiment_description
time_frequency=6:Time Frequency:optional_time_frequency_description
product=7:Product:optional_product_description
realm=8:Realm:optional_realm_description
variable=9:Variable:optional_variable_description
variable_long_name=10:Variable Long Name:optional_variable_long_name_description
cmor_table=11:CMIP Table:optional_cmor_table_description
cf_standard_name=12:CF Standard Name:optional_cf_standard_name_description
ensemble=13:Ensemble:optional_ensemble_description
domain=14:Domain:optional_domain_description
driving_model=15:Driving Model:optional_driving_model_description
rcm_version=16:Downscaling realisation:optional_ds_description
data_node=17:Data Node:optional_data_node_description
```

# 4 ESGF Attribute Services

File name: `esgf_ats_static.xml`
Standard location: `/esg/config/esgf_ats_static.xml`

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<!-- File that may contain custom attribute service and registration endpoints
     in addition to those contained in the file esgf_ats.xml.
     This file is supposed to be maintained by the local system administrators,
     while the file esgf_ats.xml is dynamically generated by the node manager. -->
<ats_whitelist xmlns="http://www.esgf.org/whitelist">

    <!-- pcmdi9 Attribute and Registration services: it is included here to allow registration for "esgf-test" nodes,
         otherwise it should be contained in file esgf_ats.xml for nodes in the "esgf-prod" group -->
    <attribute type="CMIP5 Research"
               attributeService="https://pcmdi9.llnl.gov/esgf-idp/saml/soap/secure/attributeService.htm"
               description="Users of CMIP5 data for non-commercial research purposes only"
               registrationService="https://pcmdi9.llnl.gov/esgf-idp/secure/registrationService.htm"/>
    <attribute type="CMIP5 Commercial"
               attributeService="https://pcmdi9.llnl.gov/esgf-idp/saml/soap/secure/attributeService.htm"
               description="Users of CMIP5 data for commercial purposes"
               registrationService="https://pcmdi9.llnl.gov/esgf-idp/secure/registrationService.htm"/>

    <attribute type="CORDEX_Commercial" attributeService="https://esg-dn1.nsc.liu.se/esgf-idp/saml/soap/secure/\
    attributeService.htm" description="User group for possible commercial users of CORDEX data" registrationService="\
    https://esg-dn1.nsc.liu.se/esgf-idp/secure/registrationService.htm"/>
    <attribute type="CORDEX_Research" attributeService="https://esg-dn1.nsc.liu.se/esgf-idp/saml/soap/secure/\
    attributeService.htm" description="User group for non-commercial users of CORDEX data only" registrationService="\
    https://esg-dn1.nsc.liu.se/esgf-idp/secure/registrationService.htm"/>
</ats_whitelist>
```

# 5 ESGF IDP Whitelist settings

File name: `esgf_idp_static.xml`
Standard location: `/esg/config/esgf_idp_static.xml`

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<idp_whitelist xmlns="http://www.esgf.org/whitelist">
    <value>https://cmip-gw.badc.rl.ac.uk/openid/openidserver</value>
    <value>https://ceda.ac.uk/OpenID/Provider/server</value>
    <value>https://ceda.ac.uk/openid/Provider/server</value>
    <value>https://ipcc-ar5.dkrz.de/openid/provider.htm</value>
    <value>https://albedo2.dkrz.de/esgcet/openid/provider.htm</value>
    <value>https://esg-gateway.jpl.nasa.gov/openid/provider.htm</value>
    <value>https://www.earthsystemgrid.org/openid/provider.htm</value>
    <value>https://esg.ucar.edu/openid/provider.htm</value>
    <value>https://esg.nci.org.au/esgcet/openid/provider.htm</value>
    <value>https://esg.nersc.gov/esgcet/openid/provider.htm</value>
    <value>https://esg2-gw.ccs.ornl.gov/esgcet/openid/provider.htm</value>
    <value>https://esg-gw.ornl.teragrid.org/openid/provider.htm</value>
    <value>https://hydra.fsl.noaa.gov/esgf-idp/idp/openidServer.htm</value>
    <value>https://dev-hydra.esrl.svc/esgf-idp/idp/openidServer.htm</value>
    <value>https://pcmdi9.llnl.gov/esgf-idp/idp/openidServer.htm</value>
    <value>https://pcmdi11.llnl.gov/esgf-idp/idp/openidServer.htm</value>
    <value>https://esg-dn1.nsc.liu.se/esgf-idp/idp/openidServer.htm</value>
    <value>https://esg-datanode.jpl.nasa.gov/esgf-idp/idp/openidServer.htm</value>
    <value>https://esgf-data.dkrz.de/esgf-idp/idp/openidServer.htm</value>
    <value>https://esgf-node.ipsl.fr/esgf-idp/idp/openidServer.htm</value>
    <value>https://noresg.norstore.uio.no/esgf-idp/idp/openidServer.htm</value>
    <value>https://esg.bnu.edu.cn/esgf-idp/idp/openidServer.htm</value>
    <value>https://cordexesg.dmi.dk/esgf-idp/idp/openidServer.htm</value>
</idp_whitelist>
```

# 6 Publication Version

It was decided at the Trieste meet that all data published on IS-ENES2 datanodes will clearly specify the version number which is the date of the publication, expressed in the format `yyyymmdd`. This requires the creation of directory with that name, in the physical directory structure. This directory has to be created after the 'Variable name' directory. Examples:

`/datapool1/cordexdata/cordex/output/MNA-22/SMHI/ECMWF-ERAINT/evaluation/r0i0p0/SMHI-RCA4/v1/fx/orog/20131101`

`/datapool1/cordexdata/cordex/output/ARC-44/SMHI/NCC-NorESM1-M/historical/r0i0p0/SMHI-RCA4/v1/fx/sftlf/20140123`

To get this version number correctly, the procedure is to append a `--new-version <versionnum>` to the `esgpublish` command.

# 7 Directory Structure

The path to the directory tree containing the data shall have `Project/Product` followed by the directory tree containing the data.
Given below are examples of valid and invalid directory structures.

`/cordex/output/...` ✓
`/localfs/localpath/cordex/output/...` ✓[1][2]
`/corddata/output/...` ✗ //non-standard name corresponding to 'Project'.
`/cordex/AFR-44/...` ✗ //there is no directory corresponding to 'Product'.

# 8 Variables to be excluded during publish: CORDEX

The following declaration inside `/esg/esgcet/esg.ini` should be used to exclude certain variables from the THREDDS catalogues generated by `esgpublish`. Note that this differs from the default value created by previous versions of `esgsetup`; in particular managers should ensure that the variable `basin` is NOT excluded.

```
thredds_exclude_variables = a, a_bnds, alev1, alevel, alevhalf, alt40, b, \
b_bnds, bnds, bounds_lat, bounds_lon, dbze, depth, depth0m, depth100m, \
depth_bnds, geo_region, height, height10m, height2m, Lambert_Conformal, lat,\
lat_bnds, lat_bounds, latitude, latitude_bnds, layer, lev, lev_bnds, location,\
lon, lon_bnds, lon_bounds, longitude, longitude_bnds, olayer100m, olevel, oline,\
p0, p220, p500, p560, p700, p840, plev, plev3, plev7, plev8, plev_bnds, plevs, \
pressure1, region, rho, rlat, rotated_pole, rlon, scatratio, sdepth, sdepth1, \
sza5, tau, tau_bnds, time, time1, time2, time_bnds, vegtype, x, y
```

[1] Some sites use the lower-case 'cordex' while some use 'CORDEX'; While there is no rule, the lower-case 'cordex' may be considered as the prefered option.

[2] 'output' is the value of the 'Product' facet option here. It may take other values that are applicable to the 'Product' facet in the future.

# 9 Value for the 'Model' facet

It was decided that the value of the 'Model' facet should NOT contain the institute information, as this information is already captured and presented by the 'Institute' facet. However, the directory corresponding to the 'Model' contains the name of the institute too, along with the model name, as stipulated by the CORDEX archive specifications [3]. This results in the requirement for some special handling.

The easiest way to handle this is by creating a substitution map for the variable.

1. Under the options for `[project:cordex]`, find the configuration line that says `maps`
2. Edit the line to say the following:
   `maps = model_map,institute_map, las_time_delta_map, domain_map`
3. Create a new map 'model_map' and populate it with entries that correspond to the models that you handle, leaving out the institute part in the last field.
4. Look at the example below for reference:

   ```
   model_map = map(project,rcm_model : model)
       cordex |SMHI-RCA4| RCA4
       cordex |SMHI-RCA4-SN| RCA4-SN
   ```

5. Use the regex 'model' in the place of the directory corresponding to the model directory, in the `dataset_id` string.

   ```
   dataset_id = cordex.%(product)s.%(domain)s.%(institute)s.%(driving_model)s.\
   %(experiment)s.%(ensemble)s.%(model)s.%(rcm_version)s.%(time_frequency)s.\
   %(variable)s
   ```

## 9.1 Complete `model_map`

The current and comprehensive list of CORDEX models may be obtained from:
http://cordex.dmi.dk/joomla/images/CORDEX/RCMModelName.txt.
Given below is a script that can generate a complete `model_map` table, that could then be pasted into the ini file. Also present is the output of the script.

```
#!/bin/bash

curl -s 'http://cordex.dmi.dk/joomla/images/CORDEX/RCMModelName.txt'|grep -v '#'|grep 'confirmed'|tr -s ' ' >modelnames.txt
nl=`cat modelnames.txt|wc -l`;
echo "model_map = map(project,rcm_model : model)" >modelmap.txt
for (( i=1; i<=nl; i++ )); do
        ln=`cat modelnames.txt|head -$i|tail -1`;
        inst=`echo $ln|cut -d' ' -f2`;
        res=`echo $ln| sed "s/$inst-\(.*\) \(.*\) confirmed/\tcordex| $inst-\1| \1/"`
        echo -e "$res" >>modelmap.txt;
done
```

[3] "RCMModelName is an alphanumeric identifier chosen by the modeling group; it should consist of an institute acronym and a model acronym, connected by a dash, e.g., DMI-HIRHAM5 or SMHI-RCA3."[3]

```
model_map = map(project,rcm_model : model)
        cordex| AUTH-LHTEE-WRF321B| WRF321B
        cordex| AUTH-Met-WRF331A| WRF331A
        cordex| AWI-HIRHAM5| HIRHAM5
        cordex| CCCma-CanRCM4| CanRCM4
        cordex| CHMI-ALADIN52| ALADIN52
        cordex| CLMcom-CCLM4-8-17| CCLM4-8-17
        cordex| CNRM-ALADIN52| ALADIN52
        cordex| CNRM-ARPEGE52| ARPEGE52
        cordex| CRP-GL-WRF331A| WRF331A
        cordex| CUNI-RegCM4-2| RegCM4-2
        cordex| DHMZ-RegCM4-2| RegCM4-2
        cordex| DMI-HIRHAM5| HIRHAM5
        cordex| ENEA-RegCM4-3| RegCM4-3
        cordex| HMS-ALADIN52| ALADIN52
        cordex| ICTP-RegCM4-3| RegCM4-3
        cordex| IDL-WRF331D| WRF331D
        cordex| IPSL-INERIS-WRF331F| WRF331F
        cordex| KNMI-RACMO21P| RACMO21P
        cordex| KNMI-RACMO22T| RACMO22T
        cordex| MIUB-WRF331A| WRF331A
        cordex| MOHC-HadGEM3-RA| HadGEM3-RA
        cordex| MOHC-HadRM3P| HadRM3P
        cordex| MPI-CSC-REMO2009| REMO2009
        cordex| NUIM-WRF331F| WRF331F
        cordex| SMHI-RCA4| RCA4
        cordex| SMHI-RCA4-SN| RCA4-SN
        cordex| SMHI-RCAO| RCAO
        cordex| SMHI-RCAO-SN| RCAO-SN
        cordex| UCAN-WRF331G| WRF331G
        cordex| UCAN-WRF350I| WRF350I
        cordex| UCLM-PROMES| PROMES
        cordex| UHOH-WRF331H| WRF331H
        cordex| UQAM-CRCM5| CRCM5
```

# 10 Acknowledgments

Many people have contributed to this document, pointing out errors and suggesting improvements. Thanks in particular to Katharina Berger and Stephanie Legutke of the DKRZ and Stephen Pascoe of the BADC for their suggestions. Together, we hope to make the task of datanode administration a bit less of a hopeless task, with this reference manual!

# References


[1] Martin Juckes. CORDEX: ESGF Search Facet Mappings. URL: https://github.com/snic-nsc/datanode-mgr-doc/raw/master/ro/cordexSearchFacets_v5_20140210.doc.

[2] Martin Juckes. Trieste Meeting Notes. URL: https://docs.google.com/document/d/1rRXn4pyOlb95K9mYqxpxoZrIlc9XlWxn4HA.

[3] O.B. Christensen et al. CORDEX Archive Design. URL: https://verc.enes.org/data/projects/documents/cordex-archive-design.