Table of Contents

You find more information about cordex publishing data in this manual:
?https://github.com/snic-nsc/datanode-mgr-doc/raw/master/ro/Datanodemgr-doc.pdf and general publishing in :
?http://www2-pcmdi.llnl.gov/Members/bdrach/.personal/esg-publisher-configuration/

Data Publishing

Configuring CORDEX project for ESGF publication

In order to publish, you have to configure a text file, `/esg/config/esgcet/esg.ini`.

We must modify default project in esg.ini

```
[DEFAULT]
checksum = md5sum | MD5
dburl = postgresql://esgcet:Xubuntu@localhost:5432/esgcet
gateway_options = ESG-PCMDI, ESG-NCAR, ESG-ORNL, ESG-BADC, ESG-NCI
hessian_service_certfile = %(home)s/.globus/certificate-file
hessian_service_debug = false
hessian_service_keyfile = %(home)s/.globus/certificate-file
hessian_service_polling_delay = 3
hessian_service_polling_iterations = 10
hessian_service_port = 443
hessian_service_remote_metadata_url = http://host/esgcet/remote/hessian/guest/remoteMetadataService
hessian_service_url = https://esgf-node.ipsl.fr/esg-search/remote/secure/client-cert/hessian/publishingService
log_format = %(levelname)-10s %(asctime)s %(message)s
log_level = DEBUG
offline_lister =
       HRMatPCMDI | hsi
project_options =
       cordex | CORDEX Output data | 1
rest_service_url = https://esgf-node.ipsl.fr/esg-search/ws
root_id = unican
solr_search_service_url = http://esgf-node.ipsl.fr/esg-search/search
thredds_aggregation_services =
       OpenDAP | /thredds/dodsC/ | gridded
thredds_authentication_realm = THREDDS Data Server
thredds_catalog_basename = %(dataset_id)s.v%(version)s.xml
thredds_dataset_roots =
       esg_cordexnoncommercial | /datasets/cordex-noncommercial
thredds_error_pattern = Catalog init
thredds_fatal_error_pattern = **Fatal
thredds_file_services =
       HTTPServer | /thredds/fileServer/ | HTTPServer | fileservice
       GridFTP | gsiftp://data.meteo.unican.es:2811/ | GRIDFTP | fileservice
       OpenDAP | /thredds/dodsC/ | OpenDAP | fileservice
thredds_master_catalog_name = Earth System Grid catalog
thredds_max_catalogs_per_directory = 500
thredds_offline_services =
       SRM | srm://host.sample.gov:6288/srm/v2/server?SFN=/archive.sample.gov | HRMatPCMDI
thredds_password = Xubuntu
thredds_reinit_error_url = https://localhost:443/thredds/admin/content/logs/catalogInit.log
thredds_reinit_success_pattern = reinit ok
thredds_reinit_url = https://localhost:443/thredds/admin/debug?catalogs/reinit
thredds_restrict_access = esg-user
thredds_root = /esg/content/thredds/esgcet
thredds_root_catalog_name = Earth System Root catalog
thredds_url = http://data.meteo.unican.es/thredds/esgcet
thredds_username = dnode_user
```

Then, we are going to create a new project called `cordex` in esg.ini

```
[project:cordex]
categories =
       project          | enum | true | true | 0
       domain           | enum | true | true | 1
       institute        | enum | true | true | 2
       product          | enum | true | true | 3
       driving_model    | enum | false | true | 4
       experiment       | enum | false | true | 5
       ensemble         | enum | false | true | 6
       model            | enum | false | true | 7
       time_frequency   | enum | false | true | 8
       version          | enum | false | true | 9
       rcm_model        | enum | false | true | 10
       rcm_version      | enum | false | true | 11
       description      | text | false | false | 99
category_defaults =
       domain | EUR-44
       institute | UCAN
       driving_model | ECMWF-ERAINT
       ensemble | r1i1p1
       time_frequency| mon
       experiment | evaluation
       model| WRF331G
       product | output
model_map = map(project,rcm_model : model)
       cordex| AUTH-LHTEE-WRF321B| WRF321B
       cordex| AUTH-Met-WRF331A| WRF331A
       cordex| AWI-HIRHAM5| HIRHAM5
       cordex| BCCR-WRF331C| WRF331C
       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| KNMI-RACMO22E| RACMO22E
       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
domain_map = map(project_id,domain : domain_description)
```

```
       cordex | SAM-44 | South America
       cordex | CAM-44 | Central America
       cordex | NAM-44 | North America
       cordex | EUR-44 | Europe
       cordex | EUR-22 | Europe
       cordex | AFR-44 | Africa
       cordex | WAS-44 | West Asia
       cordex | EAS-44 | East Asia
       cordex | CAS-44 | Central Asia
       cordex | AUS-44 | Australasia
       cordex | ANT-44 | Antarctica
       cordex | ARC-44 | The Arctic
       cordex | MED-44 | HYMEX Mediterranean
       cordex | EUR-11 | High-res. Europe
       cordex | SAM-44i | South America
       cordex | CAM-44i | Central America
       cordex | NAM-44i | North America
       cordex | EUR-44i | Europe
       cordex | AFR-44i | Africa
       cordex | WAS-44i | West Asia
       cordex | EAS-44i | East Asia
       cordex | CAS-44i | Central Asia
       cordex | AUS-44i | Australasia
       cordex | ANT-44i | Antarctica
       cordex | ARC-44i | The Arctic
       cordex | MED-44i | HYMEX Mediterranean
       cordex | EUR-11i | High-res. Europe
       cordex | MNA-44  | Middle East and North Africa
       cordex | MNA-44i | Middle East and North Africa
       cordex | MNA-22  | Middle East and North Africa
       cordex | MNA-22i | Middle East and North Africa
domain_options = SAM-44,CAM-44,NAM-44,EUR-44,EUR-22,EUR-44i,AFR-44,AFR-44i,WAS-44,EAS-44,CAS-44,AUS-44,ANT-44,ARC-44,MED-4
driving_model_options = ERAINT, ECMWF-ERAINT, CCCma-CanESM2, CNRM-CERFACS-CNRM-CM5, ICHEC-EC-EARTH, MIROC-MIROC5, MOHC-Had
ensemble_options = r1i1p1, r12i1p1, r0i0p0
product_options = output1, output2, output
experiment_options =
       cordex | evaluation | no description
       cordex | historical | no description
       cordex | rcp4 | no description
       cordex | rcp26 | no description
       cordex | rcp45 | no description
       cordex | rcp85 | no description
institute_map = map(project_id,model : institute)
       cordex | WRF331G | UCAN
institute_options = UCAN
las_configure = false
las_time_delta_map = map(time_frequency : las_time_delta)
       mon     | 1 month
       day     | 1 day
       fx      | fixed
       sem     | semi
maps = institute_map, las_time_delta_map, domain_map
model_options = WRF331G
parent_id = wdcc2.cordex
project_handler_name = basic_builtin
rcm_model_options = UCAN-WRF331G
rcm_version_options = v01,v02
thredds_exclude_variables = a, a_bnds, alev1, alevel, alevhalf, alt40, b, b_bnds, basin, bnds, bounds_lat, bounds_lon, dbz
time_frequency_options = day,fx,mon,sem,3hr,6hr
variable_locate = clivi, clivi_|  clt, clt_|  evspsbl, evspsbl_ | hfls , hfls_ | hfss , hfss_ | hus850 , hus850_|  huss ,h
variable_per_file = true
```

```
version_options = 20140328
model= WRF331G

dataset_id = cordex.%(product)s.%(domain)s.%(institute)s.%(driving_model)s.%(experiment)s.%(ensemble)s.%(model)s.%(rcm_ver
directory_format = /datasets/cordex-noncommercial/cordex/%(product)s/%(domain)s/%(institute)s/%(driving_model)s/%(experime
```

Therefore, if you use the above configuration file, you will have to create a tree directory like this:

```
[root@data ~]# tree /datasets/
/datasets/
`-- cordex-noncommercial
   `-- cordex
       `-- output
           |-- EUR-22
           |   `-- UCAN
           |       `-- ECMWF-ERAINT
           |           `-- evaluation
           |               `-- r1i1p1
           |                   `-- UCAN-WRF331G
           |                       `-- v02
           |                           |-- 3hr
           |                           |   |-- clivi
           |                           |   |   `-- v20140328
           |                           |   |-- clt
           |                           |   |   `-- v20140328
           |                           |   |       |-- clt_EUR-22_ECMWF-ERAINT_evaluation_r1i1p1_UCAN-WRF331G_v02_3hr_1979
           |                           |   |       |-- clt_EUR-22_ECMWF-ERAINT_evaluation_r1i1p1_UCAN-WRF331G_v02_3hr_1980
           |                           |   |       |-- clt_EUR-22_ECMWF-ERAINT_evaluation_r1i1p1_UCAN-WRF331G_v02_3hr_1981
           |                           |   |       |-- clt_EUR-22_ECMWF-ERAINT_evaluation_r1i1p1_UCAN-WRF331G_v02_3hr_1982
           |                           |   |       |-- clt_EUR-22_ECMWF-ERAINT_evaluation_r1i1p1_UCAN-WRF331G_v02_3hr_1983
           |                           |   |       |-- clt_EUR-22_ECMWF-ERAINT_evaluation_r1i1p1_UCAN-WRF331G_v02_3hr_1984
           |                           |   |       |-- clt_EUR-22_ECMWF-ERAINT_evaluation_r1i1p1_UCAN-WRF331G_v02_3hr_1985
           |                           |   |       |-- clt_EUR-22_ECMWF-ERAINT_evaluation_r1i1p1_UCAN-WRF331G_v02_3hr_1986
           |                           |   |       |-- clt_EUR-22_ECMWF-ERAINT_evaluation_r1i1p1_UCAN-WRF331G_v02_3hr_1987
           |                           |   |       |-- clt_EUR-22_ECMWF-ERAINT_evaluation_r1i1p1_UCAN-WRF331G_v02_3hr_1988
           |                           |   |       |-- clt_EUR-22_ECMWF-ERAINT_evaluation_r1i1p1_UCAN-WRF331G_v02_3hr_1989
           |                           |   |       |-- clt_EUR-22_ECMWF-ERAINT_evaluation_r1i1p1_UCAN-WRF331G_v02_3hr_1990
           |                           |   |       |-- clt_EUR-22_ECMWF-ERAINT_evaluation_r1i1p1_UCAN-WRF331G_v02_3hr_1991
           |                           |   |       |-- clt_EUR-22_ECMWF-ERAINT_evaluation_r1i1p1_UCAN-WRF331G_v02_3hr_1992
```

Then you have to add the project name to the `esgcet_models_table.txt` file

```
$ echo "cordex | WRF331G | UCAN | http://meteo.unican.es" >> /esg/config/esgcet/esgcet_models_table.txt
```

After modifying `esgcet_models_table.txt` and `esg.ini` files, you have to update the data base by executing :

```
$ cd /usr/local/uvcdat/1.4.0/bin/
$ ./esginitialize -i /esg/config/esgcet/esg.ini -c
```

To remove all tables : `esginitialize -d 0`

Using the ESGF Publisher

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

This takes place in three steps:

- Scan each file for metadata and save the metadata in the node database. (This is in contrast to running `esgscan_directory`, which just scans the directory structure.)
- Generate a THREDDS catalog based on the scanned information. THREDDS is a data and metadata server used by ESGF.
- Notify the idx that one or more catalogs have been generated.

File Scan Phase

In order to scan the cordex files for metadata, run `esgscan_directory` to generate a mapfile and after that run `esgpublish` with input from a mapfile:

```
$ whoami
root
$ cd /usr/local/uvcdat/1.4.0/bin
$ ./esgscan_directory -i /esg/config/esgcet/esg.ini --project cordex -o ~/cordex_v20140328.txt /datasets --service fileser
```

Generate a THREDDS catalog

You can generate the THREDDS catalog with :

```
$ cd /usr/local/uvcdat/1.4.0/bin
$ ./esgpublish -i /esg/config/esgcet/esg.ini --project cordex --map ~/cordex_v20140328.txt --service fileservice --new-ver
```

In order to remove the catalogs from the THREDDS :

```
$ cd /usr/local/uvcdat/1.4.0/bin
$ ./esgunpublish -i /esg/config/esgcet/esg.ini --map ~/cordex_v20140328.txt --skip-gateway
```

idx notification

First, obtain a digital certificate from an ESGF trusted MyProxy server, and rename it to whatever path you have defined in esg.ini.

> Remember, you have to log in a Federation to do it.

```
$ /usr/local/globus/bin/myproxy-logon -s esgf-node.ipsl.fr -l josecarlosblanco -o ~/.globus/certificate-file
```

Then you can publish the cordex catalog by executing :

```
$ cd /usr/local/uvcdat/1.4.0/bin
$ ./esgpublish -i /esg/config/esgcet/esg.ini --project cordex --map ~/cordex_v20140328.txt --service fileservice --new-ver
INFO       2013-11-19 20:01:24,817 Publishing: cordex.EUR-22.UCAN.ECMWF-ERAINT.evaluation.r1i1p1.WRF331G_v02.3hr.hfls
INFO       2013-11-19 20:01:28,678   Result: SUCCESSFUL
INFO       2013-11-19 20:01:28,678 Publishing: cordex.EUR-22.UCAN.ECMWF-ERAINT.evaluation.r1i1p1.WRF331G_v02.3hr.hfss
INFO       2013-11-19 20:01:32,416   Result: SUCCESSFUL
INFO       2013-11-19 20:01:32,417 Publishing: cordex.EUR-22.UCAN.ECMWF-ERAINT.evaluation.r1i1p1.WRF331G_v02.3hr.huss
INFO       2013-11-19 20:01:36,125   Result: SUCCESSFUL
INFO       2013-11-19 20:01:36,125 Publishing: cordex.EUR-22.UCAN.ECMWF-ERAINT.evaluation.r1i1p1.WRF331G_v02.3hr.pr
INFO       2013-11-19 20:01:39,964   Result: SUCCESSFUL
INFO       2013-11-19 20:01:39,965 Publishing: cordex.EUR-22.UCAN.ECMWF-ERAINT.evaluation.r1i1p1.WRF331G_v02.3hr.prc
```

Use `esgunpublish` to delete idx datasets:

```
$ ./esgunpublish -i /esg/config/esgcet/esg.ini --map ~/cordex_v20140328.txt --skip-thredds
```

Running all publication steps

For convenience, the full publication can be performed with one command. Also, if the arguments are directories rather than a mapfile, the directories will be scanned as if esgscan_directory were run:

```
$ esgpublish -i /esg/config/esgcet/esg.ini --project cordex --map ~/cordex_v20140328.txt --service fileservice --new-versi
INFO       2013-11-19 19:46:30,642 Writing THREDDS catalog /esg/content/thredds/esgcet/1/cordex.EUR-22.UCAN.ECMWF-ERAINT.e
INFO       2013-11-19 19:46:30,837 Writing THREDDS catalog /esg/content/thredds/esgcet/1/cordex.EUR-22.UCAN.ECMWF-ERAINT.e
INFO       2013-11-19 19:46:31,019 Writing THREDDS catalog /esg/content/thredds/esgcet/1/cordex.EUR-22.UCAN.ECMWF-ERAINT.e
```

`esgunpublish` will remove the datasets from the idx, THREDDS, and node database in that order:

```
$ esgpublish -i /esg/config/esgcet/esg.ini --database-delete --map ~/cordex_v20140328.txt
```

Access files

Finally, in order to grant access to the files you need to add the lines below :

```
 <policy resource=".*cordexnoncommercial.*" attribute_type="CORDEX_Research" attribute_value="user" action="Read"/>
   <policy resource=".*view.*" attribute_type="CORDEX_Research" attribute_value="user" action="Read"/>
   <policy resource=".*cordex.*" attribute_type="wheel" attribute_value="super" action="Write"/>
```

in your `esgf_policies_local.xml` file :

```
cat /esg/config/esgf_policies_local.xml
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>

<!-- This file is intended to be edited and maintained by the local Node admnistrators.
    It should contain only policies specific to data served by the local Node.
    It will not be overridden by a software update. -->
<policies xmlns="http://www.esgf.org/security">

   <!-- The following statements allow all members of group "CMIP5 Research" or "CMIP5 Commercial" to read any local URL t
        Note that the groups "CMIP5 Research" and "CMIP5 Commercial" are administered by PCMDI -->
   <!-- <policy resource=".*cmip5.*" attribute_type="CMIP5 Research" attribute_value="user" action="Read"/> -->
   <!-- <policy resource=".*cmip5.*" attribute_type="CMIP5 Commercial" attribute_value="user" action="Read"/> -->
   <!-- These statements provide CMIP5 Read access for members of the old gateways -->
   <!-- <policy resource=".*cmip5.*" attribute_type="CMIP5 Research" attribute_value="default" action="Read"/> -->
   <!-- <policy resource=".*cmip5.*" attribute_type="CMIP5 Commercial" attribute_value="default" action="Read"/> -->


   <!-- The following statement allows all members of group "MY GROUP" to read any local URL that contains "my_data"
   <policy resource=".*my_data.*" attribute_type="MY GROUP" attribute_value="user" action="Read"/> -->

   <!-- The following statement allows members of group "MY GROUP" with role="publisher" to publish local datasets with id
   <policy resource=".*my_data.*" attribute_type="MY GROUP" attribute_value="publisher" action="Write"/> -->

   <!-- The following statements makes resources that contain '.*test.*' freely available for download -->
   <!-- <policy resource=".*test.*" attribute_type="ANY" attribute_value="" action="Read"/> -->

   <!-- The following statements makes all resources freely available for download -->
   <!-- <policy resource=".*" attribute_type="ANY" attribute_value="" action="Read"/> -->
   <policy resource=".*cordexnoncommercial.*" attribute_type="CORDEX_Research" attribute_value="user" action="Read"/>
   <policy resource=".*view.*" attribute_type="CORDEX_Research" attribute_value="user" action="Read"/>
   <policy resource=".*cordex.*" attribute_type="wheel" attribute_value="super" action="Write"/>
</policies>
```

Finally , you have to restart the services :

```
$ esg-node --restart
```

See also

- ESGF Data Visibility API
- ESGF Node Installation
- ESGF-Security