How to generate and distribute the CORDEX-EA metadata through ESGF Data-node in APEC Climate Center

• APEC Climate Center Korea, 2019.10.17
• Jeong-Min Han, Jae-Won Choi
1. Background

• Need
  – Experiments are increasing rapidly to generate scenario data for global climate change response
  – Detailed scenarios and regional climate change data capacities are also steadily increasing as the resolution of regional models increase.
  – In order to manage, share and distribute large volumes of data, it is necessary to build a system that complies with international standard.

• Purpose
  – Establish and operate ESGF (Earth System Grid Federation) data node system
  – Federate with ESGF Index node system and CORDEX-EA phase2 metadata distribution
2. History of ESGF data node at APCC

ESGF Technology sharing and East Asia data service cooperation
3. What’s the ESGF

- **Definition of ESGF**
  - ESGF (Earth System Grid Federation)
    - The ESGF is a collaboration peer to peer system
    - The ESGF is software infrastructure for the management and analysis of model output and observational data.
    - ESGF’s primary goal is to facilitate advancements in Earth system science.
  - ESGF federation
4. Requirements for operating ESGF Data node

- ESGF Tier2(data node) sites Requirement

  - Have an uptime more than 90%
  - installation of complete SW stack to support for data node
  - installation of most recent version of SW(within 2Weeks)
  - prompt upgrade in case of detected security breaches (within 7days)
  - be responsible for the node maintenance and operation

- Exclusion of Data nodes:

  There is an ongoing proposition to enable ESGF to exclude a data node that does not satisfy all the ESGF node operation requirements or a data node that will degrade the federation usability

by “Active/Standby” optimal environment

Satisfy the requirement of ESGF Tier 2
5. Dual system for ESGF data node

- ESGF Tier2 Sites configuration

“Active-Standby” system management method applied
6. Network for ESGF data node

- ESGF Tier 2 sites Network system.

ESGF Tier 2 sites Network system.

<table>
<thead>
<tr>
<th>CORDEX Phase-1</th>
<th>ESGF-APCC 1</th>
<th>Storage (Max. 48TB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homepage</td>
<td></td>
<td>Only 2 ~ 3 year Available</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NIMS</th>
<th>A Univ.</th>
<th>B Univ.</th>
<th>C Univ.</th>
<th>Japan</th>
<th>China</th>
</tr>
</thead>
</table>

ESGF service library

Service
(Virtual IP 210.98.xx.xx)
esgf.apcc21.org

Sync (Backup)

ESGF-APCC 2

Temp. Storage (Max. 40TB)

APEC Climate Center

CORDEX Phase-2 Data Quality Control

A Univ. B Univ. C Univ. Japan China

Only 2 ~ 3 year Available
7. ESGF Data Node at APCC

- **URL:** [http://esgf.apcc21.org](http://esgf.apcc21.org) (Datanode Only)

- **Configuration:**
  - APCC’s system is interconnected with NSC system in Sweden (using Peer to peer enterprise system)
  - Distribute CORDEX-EA data on Data node into Index node.
8. Transfer the data using GridFTP

- Interoperating between ESGFAPCC and Globus
  - Sharing the data using Globus
9. Target CORDEX Data

- CORDEX-EA Phase 2 RCMs

Source: NIMS in KOREA
10. Sample data

- Sample of CORDEX Data

<table>
<thead>
<tr>
<th>Sample of CORDEX Data</th>
<th>Supported by NSC(Sweden)</th>
<th>Supported by NIMS(Korea)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESGF</td>
<td>CORDEX</td>
<td>SMHI-RCA4_v1_fx.nc</td>
</tr>
<tr>
<td>tas_EAS-22_ECMWF-ERAINT_evaluation_r1i1p1_NIMS-KMA-HadGEM3-RCA_v1_mon_197901-198012.nc</td>
<td>2019-0</td>
<td></td>
</tr>
<tr>
<td>tas_EAS-22_ECMWF-ERAINT_evaluation_r1i1p1_NIMS-KMA-HadGEM3-RCA_v1_mon_198101-199012.nc</td>
<td>2019-0</td>
<td></td>
</tr>
<tr>
<td>tas_EAS-22_ECMWF-ERAINT_evaluation_r1i1p1_NIMS-KMA-HadGEM3-RCA_v1_mon_198101-199012.nc</td>
<td>2019-0</td>
<td></td>
</tr>
</tbody>
</table>

```
[root@esgf v20130927]# pwd
/esgf/data/datapool/cordexdat/cordex/output/AFR.44/SMHI/CCMa-CanESM2/historical/r0i0p0/SMHI-RCA4_v1/l3/orog/v20130927
[root@esgf v20130927]# ls -al
```

```
root  root  4956 Mar 15 17:30 ...
```

```
root  root  22 Dec 7 23:18 ...
```

```
orog AFR-44 CCCma-CanESM2_historical_r0i0p0_SMHI-RCA4_v1_fx.nc
```

```
tas_EAS-22_ECMWF-ERAINT_evaluation_r1i1p1_NIMS-KMA-HadGEM3-RCA_v1_mon_198101-19851231.nc
```

```
tas_EAS-22_ECMWF-ERAINT_evaluation_r1i1p1_NIMS-KMA-HadGEM3-RCA_v1_mon_197901-198012.nc
```

```
tas_EAS-22_ECMWF-ERAINT_evaluation_r1i1p1_NIMS-KMA-HadGEM3-RCA_v1_mon_198101-199012.nc
```

```
[root@esgf v20130927]# ```
11. Creating Metadata on Local System

- Creating Metadata to APCC System
  - Step1. Mapfile generation
    ```
    (esgf-pub) [root@esgf publication-testing]# esgmapfile --mapfile test.map --project cordex /esg/data/datapool1
    Mapfile(s) generation: 100% | 4/4 file(s)
    Mapfile(s) generated: 1 (in /root/esgfesg/publication-testing/mapfiles)
    Number of file(s) scanned: 4
    Number of error(s): 0
    (esgf-pub) [root@esgf publication-testing]#
    ```
  - Step2. Publication at Data Node local Thredds Server.
    ```
    (esgf-pub) [root@esgf publication-testing]# esgpublish --map mapfiles/test.map --project cordex --noscan --thredds
    INFO 2019-02-27 11:42:07.776 Writing THREDDS catalog /esg/content/thredds/esgcat1/cordex/output/AR44_SMHI.CComa-CanEM2_historical.roi0p0.ROC44.v1.fx.sorg.v20130927.xml
    INFO 2019-03-27 11:43:07.854 Writing THREDDS ESG catalog /esg/content/thredds/esgcat1/catalog.xml
    INFO 2019-03-27 11:42:07.858 Reinitializing THREDDS server
    ```
12. Publication to Index node

- Peer to Index node at Sweden
  - Step 3. Publication at Peer index Node

```
(esgf-pub) [root@esgf ~]# myproxy-logon -s esg-dn1.nsc.
  connect to
  connect to esg-dn1.nsc.liu.se
Connection timed out
```

- Step 4. Publish

```
A credential has been received for user esgfapcc in /root/.globus/certificate-file.
(esgf-pub) [root@esgf publication-testing]# esgpublish --map mapfiles/test.map --project cordex --noscan --publish
INFO 2018-12-17 17:00:28,233 Publishing: cordex.output.AFR-44.SMH1.CCMa-CanESM2.historical.r010p0.RCA4.v1.fx.org
INFO 2018-12-17 17:00:34,192 Result: SUCCESSFUL
(esgf-pub) [root@esgf publication-testing]# echo
```
13. Data processing for ESGF distribution

- 1. check the properties for CORDEX attributes
- 2. Target data check
- 3. Extract Metadata
- 4. Make thredds server
- 5. Publish to index
14. Data Searching on Index node

- CORDEX-EA Data on Index Node in Sweden
15. Search method Dataset on ESGF

- Data search on ESGF System
  - Default search: project, variable, keyword
  - Optional search: default search + file name, period, dataset version
16. Search results on ESGF

- ESGF Data results
  - XML file formats
  - JSON file formats
17. Metadata view on ESGF

- View the metadata about Datasets
18. Data download

- CORDEX-EA Phase 2

- NSC’s Index Node
- APCC’s Data Node
- NIMS’s CORDEX-EA Data Center
19. Status of data service

- CORDEX-EA Phase-II data in Korea

<table>
<thead>
<tr>
<th></th>
<th>HadGEM3-RA (NIMS)</th>
<th>SNU-RCM (UNIST)</th>
<th>CCLM (POSTECH)</th>
<th>WRF (PNU)</th>
<th>RegCM4 (KNU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERA-Interim</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>GCM core</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Data QC</td>
<td>Complete</td>
<td>Complete</td>
<td>Complete</td>
<td>Complete</td>
<td>Progress</td>
</tr>
</tbody>
</table>

- CORDEX-EA Phase-II in Asia Countries

<table>
<thead>
<tr>
<th></th>
<th>Japan (MRI-JMA)</th>
<th>China (IAP-CAS)</th>
<th>China (Nanjing-UNI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Priority Data</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>2nd Priority Data</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Under discussion

*MRI-JMA: Meteorological Research Institute- Japan Meteorological Agency
*IAP-CAS: Institute of Atmospheric Physics-Chinese Academy of Sciences
20. Cooperate with East Asia countries

- Data transfer diagram between Korea and other countries

- NIMS, Korea
  (Daily 14TB 이내)

- KMA Supercomputer Center
  1) QA-DKRZ : QC run
  2) Deliver QC results to S5/S6
  3) Download using ScienceDMZ

- POSTECH
  (Daily 14TB 이내)

- UNIST
  (Daily 14TB 이내)

- PNU
  (Daily 14TB 이내)

- KNU
  (Daily 14TB 이내)

- KISTI
  Download Speed: 110 MB/s

- APEC Climate Center
  Complete the CORDEX-EA Phase II QC

- ScienceDMZ
  Download Speed: 110 MB/s

- China
- Japan
- Taiwan

Complete the CORDEX-EA Phase II QC
Data distribution is only possible on APCC System (at Data Node).
Thank you.

- APEC Climate Center, Korea
- Email: goal@apcc21.org