Bridging Data Lifecycles: The perspective from CESM

Bridging Data Lifecycles: Tracking Data Use via Data Citations Data Workshop

Gary Strand, CESM data manager
strandwg@ucar.edu
“The volume of worldwide climate data is expanding rapidly, creating challenges for both physical archiving and sharing, as well as for ease of access and finding what’s needed, particularly if you are not a climate scientist.”

J T Overpeck et al. Science 2011;331:700-702
Published by AAAS
CESM Data Management

• What counts as CESM data?
• Who is responsible and what are their obligations?
• What gets released and when? For how long are the data stored?
• Standards
General timeline summary

- % data kept
- % restricted

Timeline:
- End of run
- +6 months
- +12 months
- +4 years
- +5 years
- +6 years
- +7 years
- Later
CMIP5 distribution

- Modeling centers (24)
- Gateways (9)
- Nodes (14)
- Users (1000s)
Global attributes as required by CMIP5

```plaintext
:experiment = "historical"
:experiment_id = "historical"
:forcing = "SI GHG VI SS Ds SD BC MD OC Oz AA LU"
:realization = 1
:initialization_method = 1
:physics_version = 1
:parent_experiment = "pre-industrial control"
:parent_experiment_id = "piControl"
:parent_experiment_rip = "r1i1p1"
:branch_time = 937.
:tracking_id = "d33ccf77-a73c-4f55-8f02-3a0734d51151"
:creation_date = "2011-07-22T00:05:32Z"
:history = "2011-07-22T00:05:32Z CMOR rewrote data to comply with CF standards and CMIP5 requirements."
:Conventions = "CF-1.4"
:table_id = "Table Amon (27 April 2011) a5a1c518f52ae340313ba0aada03f862"
:cmor_version = "2.7.1"
```
Data Quality Assurance Procedures

Diagnostics

• Per-component
  AMWG package
  LMWG package
  OMWG package
  IMWG package

• Simple time series
Data QC

Data Published > 10 PB
Data and Metadata QC L1
On globally distributed data nodes

Metadata QC L2 passed?
YES
NO
Discard data

Data QC L2 passed?
YES
NO

To be replicated among ESGF?
YES
NO

Data QC L3 passed?
YES
NO

Replicated:
Copied to PCMDI, BADC, WDCC & elsewhere ~ 1PB

Data NOT formally citable
Modelling group control access manually.

Data NOT formally citable
Automatic access granted after filling in ESGF registration page.

DOI Assigned:
Data formally citable
Data can appear in IPCC-DDC
Automatic access granted after filling in ESGF registration page

(Informal citation still requested where formal citation not available)
## CMIP5 Data QC

<table>
<thead>
<tr>
<th>QC Level 1</th>
<th>QC Level 2</th>
<th>QC Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>CMOR2 and ESG publisher conformance checks</td>
<td>Data consistency checks</td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td>preliminary; no user notification about changes; performed for all data; metadata may not be complete</td>
<td>no user notification about changes; performed for CMIP5 requested metadata and data</td>
</tr>
<tr>
<td><strong>Access</strong></td>
<td>constrained to CMIP5 modeling centers</td>
<td>constrained to non-commercial research and educational purposes</td>
</tr>
<tr>
<td><strong>Access Control</strong></td>
<td>PCMDI on behalf of WMO/WGCM</td>
<td>PCMDI, BADC, WDCC/DKRZ core data archives on behalf of WMO/WGCM</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>no citation reference</td>
<td>informal citation reference</td>
</tr>
<tr>
<td><strong>Quality Flag</strong></td>
<td>&quot;automated conformance checks passed&quot;</td>
<td>&quot;subjective quality control passed&quot;</td>
</tr>
</tbody>
</table>
# QC Results for CMIP5 Experiments

The listed CMIP5 ([http://cmip-pcmdi.llnl.gov/cmip5/](http://cmip-pcmdi.llnl.gov/cmip5/)) experiments are in the quality control process ([http://cmip5qc.wdc-climate.de/](http://cmip5qc.wdc-climate.de/)) They appear in the list as soon as the QC Level 2 checks have started. Click on an experiment to get a detailed list of available QC information.

Status Explanation:
- QCL2_started: QC Level 2 checks are ongoing.
- QCL2_assigned: QC Level 2 is assigned to the experiment.
- QCL3_started: QC Level 3 checks are ongoing.
- QCL3_assigned: QC Level 3 is assigned to the experiment and the data is DOI published according to the DataCite regulations ([http://datacite.org/](http://datacite.org/)). This data, its basic metadata and its DOI are persistent.
- QCL1: The data of the experiment is withdrawn and waiting for a QC Level 2 check restart.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>cmip5/output1/CCMa/CanESM2/historical</td>
<td>QCL2_started</td>
</tr>
<tr>
<td>cmip5/output1/CCMa/CanESM2/piControl</td>
<td>QCL2_started</td>
</tr>
<tr>
<td>cmip5/output1/CCMa/CanESM2/rcp26</td>
<td>QCL2_started</td>
</tr>
<tr>
<td>cmip5/output1/CCMa/CanESM2/rcp45</td>
<td>QCL2_started</td>
</tr>
<tr>
<td>cmip5/output1/CCMa/CanESM2/rcp85</td>
<td>QCL2_started</td>
</tr>
<tr>
<td>cmip5/output1/CCMa/CanESM2/piCO2</td>
<td>QCL2_started</td>
</tr>
<tr>
<td>cmip5/output1/CNRM-CERFACS/CNRM-CM5/decadal1960</td>
<td>QCL2_started</td>
</tr>
<tr>
<td>cmip5/output1/CNRM-CERFACS/CNRM-CM5/decadal2005</td>
<td>QCL2_started</td>
</tr>
<tr>
<td>cmip5/output1/CNRM-CERFACS/CNRM-CM5/historical</td>
<td>QCL2_started</td>
</tr>
<tr>
<td>cmip5/output1/CNRM-CERFACS/CNRM-CM5/piControl</td>
<td>QCL2_started</td>
</tr>
</tbody>
</table>
And, finally...
Websites

CESM website
http://cesm.ucar.edu

CESM Data Management Plan

CMIP5 website
http://cmip.llnl.gov/cmip5