CMIP6 Standards Enabling Management, Search, and Interpretation of Model Output

Karl E. Taylor, Paul Durack, Denis Nadeau, and Sasha Ames

Presented at the 2016 Earth System Grid Federation (ESGF) Face-to-Face Conference

Washington D.C.
5-9 December 2016
Major components of CMIP6 modeling and data archive infrastructure must communicate and be coordinated

- Prepare “forcing” datasets
- Define experiments
- Run models, produce output
- Prepare data for CMIP6 archive
- Archive data and provide data services
- Record model and experiment documentation
- Analyze data
- Publish research
- Register publication
- Assign and serve PID’s & DOI’s
- Analyze data
- Prepare “forcing” datasets
- Define experiments
- Run models, produce output
- Prepare data for CMIP6 archive
- Archive data and provide data services
- Record model and experiment documentation
- Analyze data
- Publish research
- Register publication
- Assign and serve PID’s & DOI’s
- Analyze data
- Prepare “forcing” datasets
- Define experiments
- Run models, produce output
- Prepare data for CMIP6 archive
- Archive data and provide data services
- Record model and experiment documentation
- Analyze data
- Publish research
- Register publication
- Assign and serve PID’s & DOI’s
- Analyze data
- Prepare “forcing” datasets
- Define experiments
- Run models, produce output
- Prepare data for CMIP6 archive
- Archive data and provide data services
- Record model and experiment documentation
- Analyze data
- Publish research
- Register publication
- Assign and serve PID’s & DOI’s
- Analyze data
- Prepare “forcing” datasets
- Define experiments
- Run models, produce output
- Prepare data for CMIP6 archive
- Archive data and provide data services
- Record model and experiment documentation
- Analyze data
- Publish research
- Register publication
- Assign and serve PID’s & DOI’s
- Analyze data
- Prepare “forcing” datasets
- Define experiments
- Run models, produce output
- Prepare data for CMIP6 archive
- Archive data and provide data services
- Record model and experiment documentation
- Analyze data
- Publish research
- Register publication
- Assign and serve PID’s & DOI’s
- Analyze data
- Prepare “forcing” datasets
- Define experiments
- Run models, produce output
- Prepare data for CMIP6 archive
- Archive data and provide data services
- Record model and experiment documentation
- Analyze data
- Publish research
- Register publication
- Assign and serve PID’s & DOI’s
- Analyze data
- Prepare “forcing” datasets
- Define experiments
- Run models, produce output
- Prepare data for CMIP6 archive
- Archive data and provide data services
- Record model and experiment documentation
- Analyze data
- Publish research
- Register publication
- Assign and serve PID’s & DOI’s
- Analyze data
- Prepare “forcing” datasets
- Define experiments
- Run models, produce output
- Prepare data for CMIP6 archive
- Archive data and provide data services
- Record model and experiment documentation
- Analyze data
- Publish research
- Register publication
- Assign and serve PID’s & DOI’s
- Analyze data
- Prepare “forcing” datasets
- Define experiments
- Run models, produce output
- Prepare data for CMIP6 archive
- Archive data and provide data services
- Record model and experiment documentation
- Analyze data
- Publish research
- Register publication
- Assign and serve PID’s & DOI’s
- Analyze data
- Prepare “forcing” datasets
- Define experiments
- Run models, produce output
- Prepare data for CMIP6 archive
- Archive data and provide data services
- Record model and experiment documentation
- Analyze data
- Publish research
- Register publication
- Assign and serve PID’s & DOI’s
- Analyze data
- Prepare “forcing” datasets
- Define experiments
- Run models, produce output
- Prepare data for CMIP6 archive
- Archive data and provide data services
- Record model and experiment documentation
- Analyze data
- Publish research
- Register publication
- Assign and serve PID’s & DOI’s

ESGF F2F Conference
7 December 2016
K. E. Taylor
PCMDI
Standardized output, metadata and CVs are relied on in managing and interpreting CMIP6 results.
Data preparation and publication steps.

1. Define CV’s, technical specifications and data requirements
2. Prepare data for CMIP6 archive
3. Archive data and provide data services

CF Conventions
Expanded view of data preparation and publication steps

Define CV’s, technical specifications and data requirements

CF Conventions

“CMORize” data

Populate directory structure with files

Perform metadata checks (PrePARE)

Publish data and generate Solr index

Data request specs

Model output
Recommend that PrePARE be run at time of file creation, so errors can be corrected immediately.

- Define CV’s, technical specifications and data requirements
- CF Conventions
- Populate directory structure with files
- “CMORize” and run PrePARE
- Data request specs
- Model output
- Publish data and generate Solr index
- Perform metadata checks (PrePARE)
- Redundant check
Define global attributes that are used in identifying datasets, constructing filenames, and defining search facet

<table>
<thead>
<tr>
<th>DRS Element</th>
<th>Examples</th>
<th>Controlled Vocabulary?</th>
</tr>
</thead>
<tbody>
<tr>
<td>activity_id</td>
<td>PMIP, CFMIP, ScenarioMIP</td>
<td>CV</td>
</tr>
<tr>
<td>product</td>
<td>=output</td>
<td>only 1 option</td>
</tr>
<tr>
<td>institution_id</td>
<td>IPSL, CCCma</td>
<td>CV</td>
</tr>
<tr>
<td>source_id</td>
<td>EC-Earth-3-LR, NorESM2-LM</td>
<td>CV</td>
</tr>
<tr>
<td>experiment_id</td>
<td>piControl, historical, 1pctCO2</td>
<td>CV</td>
</tr>
<tr>
<td>variable_id</td>
<td>tas, pr, hur</td>
<td>data request</td>
</tr>
<tr>
<td>table_id</td>
<td>Amon, 3hr, Oday</td>
<td>CV</td>
</tr>
<tr>
<td>variant_label</td>
<td>r2i1p1f1, r1i1p1f2</td>
<td>template</td>
</tr>
<tr>
<td>version</td>
<td>v20160218, v20170821</td>
<td>template</td>
</tr>
<tr>
<td>sub_experiment_id</td>
<td>1980, 1981, 2001</td>
<td>CV</td>
</tr>
<tr>
<td>grid_label</td>
<td>gn, gr, gr1, gr2</td>
<td>CV</td>
</tr>
<tr>
<td>mip_era</td>
<td>=CMIP6</td>
<td>CV</td>
</tr>
<tr>
<td>frequency</td>
<td>mon, day, 6hr</td>
<td>CV</td>
</tr>
<tr>
<td>realm</td>
<td>atmos, ocean, land, seaice</td>
<td>CV</td>
</tr>
</tbody>
</table>

doc hosted via WIP CoG site: [https://goo.gl/Cqbd6i](https://goo.gl/Cqbd6i)
46 Global attributes are defined in a table (with notes)

The attributes provide critical information needed to interpret the model output and are key attributes are relied on by the infrastructure.

<table>
<thead>
<tr>
<th>CMIP6 global attribute</th>
<th>description</th>
<th>examples</th>
<th>corresponding attribute in CMIP5</th>
<th>form</th>
<th>when required?</th>
<th>further information and rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>activity_id</td>
<td>activity identifier(s)</td>
<td>“CMIP”, “PMIP”, “LS3MIP LUMIP” see note 3</td>
<td>project_id</td>
<td>CV</td>
<td>always</td>
<td>renamed more generically, since not all activities are projects; also multiple activities may now be listed separated by single spaces.</td>
</tr>
<tr>
<td>branch_method</td>
<td>branching procedure</td>
<td>“standard”, “none provided”, “no parent” see note 4</td>
<td>-</td>
<td>free form</td>
<td>whenever parent exists</td>
<td>in CMIP6 some branching methods will involve short spin-up periods or other non-standard procedures which need to be described. See note 4. If no parent, omit or set to &quot;no parent&quot;</td>
</tr>
<tr>
<td>branch_time_in_child</td>
<td>branch time with respect to child’s time axis</td>
<td>365.0D0, 0.0D0 see note 5</td>
<td>-</td>
<td>double precision float</td>
<td>whenever parent exists</td>
<td>aids in interpreting branch times; units are the same as the units used for the child’s time axis. If no parent, omit (preferred) or set to start time of the run.</td>
</tr>
<tr>
<td>branch_time_in_parent</td>
<td>branch time with respect to parent time axis</td>
<td>3650.0D0 see note 5</td>
<td>branch_time</td>
<td>double precision float</td>
<td>whenever parent exists</td>
<td>changed name to explicitly distinguish it from branch_time_in_child; units are specified in the attribute: parent_time_units. If no parent, omit or set to 0.0D0.</td>
</tr>
</tbody>
</table>

ESGF F2F Conference
7 December 2016

doc hosted via WIP CoG site: [https://goo.gl/Cqbd6i](https://goo.gl/Cqbd6i)

K. E. Taylor
PCMDI
ESGF search facets are largely based on DRS elements governed by CVs.

**DRS elements**
- activity_id
- experiment_id
- subexperiment_id
- source_id
- run_variant_id
- frequency
- realm
- table_id
- variable_id
- grid_label
- institution_id
- product

**Other**
- Model cohort
- Source_type
- CF standard_name
- long_name
- grid_resolution
- Data_node

See CMIP6 search requirements: [https://goo.gl/rAvXIB](https://goo.gl/rAvXIB)
Controlled vocabularies are used to define filenames and directory structure and to reach documentation

- The CVs are defined by JSON files on github.

- Filenames template:
  - `<variable_id>_<table_id>_<source_id>_<experiment_id>_<member_id>_`<grid_label>[_<time_range>]`.nc`

- Directory structure template:
  - `<mip_era>/<activity_id>/<institution_id>/<source_id>/<experiment_id>/`<member_id>/`<table_id>/`<variable_id>/`<grid_label>/`<version>`

- ES-DOC documentation reachable via the further_info_url global attribute constructed from a template:
  - `http://furtherinfo.es-doc.org/`<mip_era>._<institution_id>._<source_id>._<experiment_id>._<sub_experiment_id>._<variant_label>`
JSON files are hosted by GitHub

https://github.com/WCRP-CMIP/CMIP6_CVs

Issues

CMIP6_activity_id.json

CMIP6_institution_id.json

ESGF F2F Conference
7 December 2016
Some JSON file contents are viewable through web browsers

`experiment_id.json` rendering:


- `experiment_id`
- `activity_id`
- `parent_experiment_id`
- `model_components` (description)
CMIP data request software and requirements

• Through an API, you can determine what variables to save by specifying
  ➞ An experiment
  ➞ A year of the simulation
  ➞ The experiment suite planned for your model

• Metadata associated with each variable are retrievable:
  ➞ e.g., standard_name, units, cell_methods
  ➞ CMOR tables are generated based on the metadata recorded by the data request
CMIP data request tools and documentation

- Primary source found at the WIP CoG site:
  https://www.earthsystemcog.org/projects/wip/CMIP6DataRequest

CMIP6 Data Request

The CMIP6 experimental design and organization has been agreed at the WGCM 18th Session in October 2014, see details on the CMIP Panel website at http://www.wcrp-climate.org/index.php/wgcm-cmip/about-cmip. Part of this covers the creation and timeline of the CMIP6 Data Request.

The data request is available through a repository, and the latest version is available here (updated October 21st, 2016):

http://proj.badc.rl.ac.uk/svn/exarch/CMIP6dreq/tags/latest

An overview of the pressure levels proposed for atmospheric diagnostics is available for discussion (here).

Key documents describing the request (in the "docs" directory of the repository) are:

- Examples
- Python Library (dreqPy)
- The Request XML document and Schema
- Spreadsheet view of the variable definitions
- A searchable list of variables in the request, linking to
- A browsable HTML view of the request
- Overview tables for tier 1, priority 1 and all tiers and priorities
- Discussion of issues: old forum, new github pages
- Registration for email list: CMIP6-DATAREQUEST@JISCMAIL.AC.UK
- Installation and usage of the python package

See Version 01.beta.38 Release Notes for more details

When problems are found, raise an issue! “CMIP6_DataRequest_VariableDefinitions”
What is your status/timeline?

• Agreement on
  -> Global attributes
  -> Filename template and directory structure
  -> Search facets

• Global attribute reference CVs and supporting CVs established
  -> Sources and institutions being added
  -> Decadal prediction sub-experiments need to be defined

• Data request first (non-beta) release is imminent
  -> Some additional revision of variable lists
  -> Need to correct lots of details

• CMOR 3.2 released

• Summary: Information and CVs are largely in place for
  -> Preparing model output for the archive
  -> Checking output for compliance with CMIP6 requirement
  -> Publishing output to ESGF
  -> Enabling faceted search of archive
Prospects for ongoing stable funding? Uncertain

- Development of a rational organization of experiments and model results requires a comprehensive understanding of climate models and climate science
  - Has been generously supported by DOE (largely through PCMDI)
  - Requires active leadership by climate scientists, but isn’t a “research” activity and funding is difficult to defend.

- Development of metadata requirements and CV’s requires input from an expanding variety of perspectives.
  - Before CMIP6, PCMDI consulted with individual experts and was largely responsible for this.
  - Now PCMDI relies on the WIP volunteers to work toward consensus
    - Process more open and inclusive
    - Process is less efficient
Summary of resources for CVs, metadata, data requirements

- Document defining global attributes, the “Data Reference Syntax” (DRS), the filename template and directory structure: [https://goo.gl/Cqbd6ii](https://goo.gl/Cqbd6ii) (or accessible from [https://www.earthsystemcog.org/projects/wip/position_papers](https://www.earthsystemcog.org/projects/wip/position_papers)

- CMIP6 reference controlled vocabularies (CVs): [https://github.com/WCRP-CMIP/CMIP6_CVs](https://github.com/WCRP-CMIP/CMIP6_CVs)

- Document defining search facets for CMIP6: [https://goo.gl/rAvXIB](https://goo.gl/rAvXIB)

- What variables should be archived and what attributes should be recorded? (“Data Request” information): [https://www.earthsystemcog.org/projects/wip/CMIP6DataRequest](https://www.earthsystemcog.org/projects/wip/CMIP6DataRequest)
Summary of resources for CVs, metadata, data requirements

- **CF Conventions** [http://cfconventions.org/cf-conventions](http://cfconventions.org/cf-conventions)
  - Standard names
  - Conventions

- **CMIP5 output requirements**

- **CMOR3 available to meet metadata requirements**
  - Code available from [https://github.com/PCMDI/cmor](https://github.com/PCMDI/cmor)