

The WGCM Infrastructure Panel (WIP) Update on CMIP6 Status

Karl E. Taylor and V. Balaji

on behalf of the WIP

Presented at the twenty-first session of the
WCRP Working Group on Coupled Modeling (WGCM)

Exeter, UK
10 October 2017

CMIP infrastructure coordination

- The WGCM Infrastructure Panel (WIP) appointed by the WGCM
- The WIP
 - Manages and coordinates infrastructure development and implementation.
 - Maintains a website hosting "Position Papers" and some of the specifications for CMIP6:
 - <https://www.earthsystemcog.org/projects/wip/>
 - Oversees the CMIP Data Node Operations Team (CDNOT)
- Major contributions to the infrastructure come from ESGF, ES-DOC, PCMDI, BADC, IPSL, DKRZ, and others.

This talk is structured following the CMIP6 Guide

See <https://pcmdi.llnl.gov/CMIP6/Guide> (really 3 guides):

►►► Part 1: For participating climate modelers

- Requirements and expectations
- Experiment design
- Forcing data sets
- Model output fields
- Model output requirements
- Software for preparing/checking output
- Archiving/publishing output
- Documentation process
- CMIP6 organization and governance

►►► Part 2: For ESGF data node managers and operators (being drafted)

- CDNOT
- Tracking and replication

►►► Part 3: For Users of CMIP6 model output (being drafted)

- Registering published work based on CMIP6
- obs4MIPs

This talk is structured following the CMIP6 Guide

See <https://pcmdi.llnl.gov/CMIP6/Guide>

►►► Part 1: For participating climate modelers

- **Requirements and expectations**
- Experiment design
- Forcing data sets
- Model output fields
- Model output requirements
- Software for preparing/checking output
- Archiving/publishing output
- Documentation process
- CMIP6 organization and governance

►►► Part 2: For ESGF data node managers and operators (being drafted)

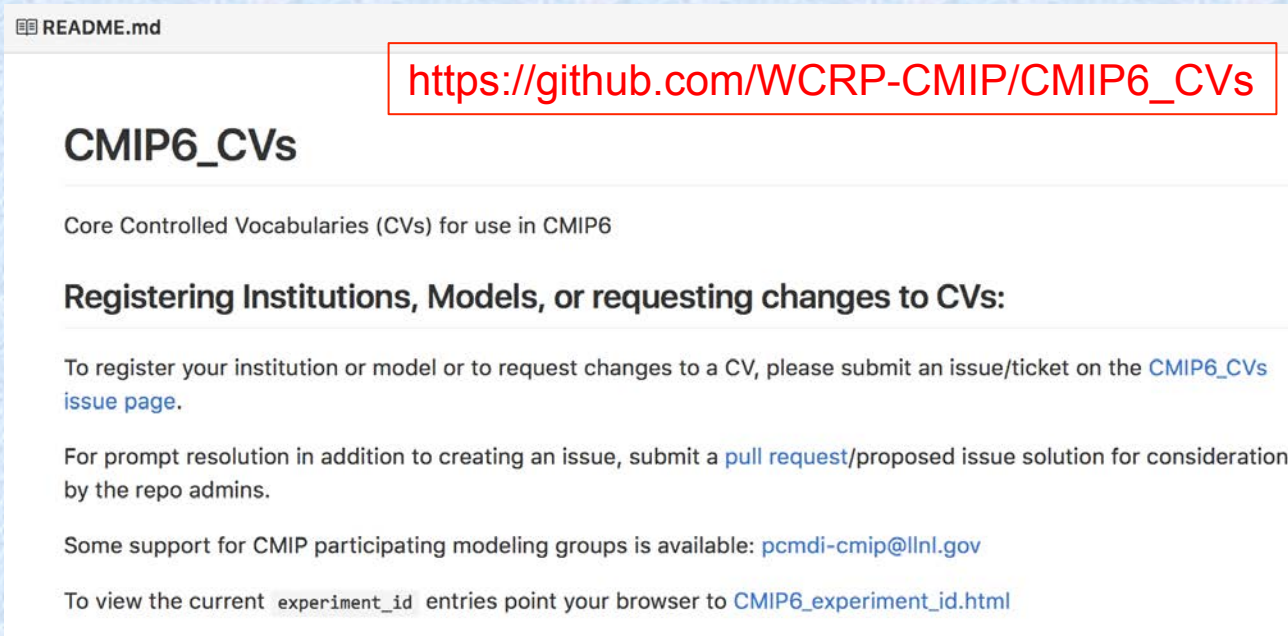
- CDNOT
- Tracking and replication

►►► Part 3: For Users of CMIP6 model output (being drafted)

- Registering published work based on CMIP6
- obs4MIPs

CMIP6 Guide: Requirements and expectations

- Summary of your major obligations, including a **high-priority request**:
 - You must register your institution and model(s) *before* preparing and publishing model output.
 - Registration enables you to:
 - Rely on software tools to write and check your output for compliance with the data specifications.
 - Publish your model output on ESGF



README.md

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

CMIP6_CVs

Core Controlled Vocabularies (CVs) for use in CMIP6

Registering Institutions, Models, or requesting changes to CVs:

To register your institution or model or to request changes to a CV, please submit an issue/ticket on the [CMIP6_CVs issue page](#).

For prompt resolution in addition to creating an issue, submit a [pull request](#)/proposed issue solution for consideration by the repo admins.

Some support for CMIP participating modeling groups is available: pcmdi-cmip@llnl.gov

To view the current `experiment_id` entries point your browser to [CMIP6_experiment_id.html](#)

CMIP6 Guide: Requirements and expectations

- Summary of your major obligations, including a **high-priority request**:
 - You must register your institution and model(s) **before** preparing and publishing model output.
 - Registration enables you to:
 - Rely on software tools to write and check your output for compliance with the data specifications.
 - Publish your model output on ESGF

README.md

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

CMIP6_CVs

Core Controlled Vocabularies (CVs) for use in CMIP6

Registering Institutions, Models, or requesting changes to CVs:

To register your institution or model or to request changes to a CV, please submit an issue/ticket on the [CMIP6_CVs issue page](#).

For prompt resolution in addition to creating an issue, submit a [pull request](#)/proposed issue solution for consideration by the repo admins.

Some support for CMIP participating modeling groups is available: pcmdi-cmip@llnl.gov

To view the current `experiment_id` entries point your browser to [CMIP6_experiment_id.html](#)

Registration status

- 31 institutions/consortia have officially registered for CMIP6
- 65 models/source_id's are registered
- List of groups (with expressed intent) who have not yet registered (**missing in red**)

ACCESS-ESM (Australia)	CESS-THU (China)	HadGEM3 (UK)	MRI-AGCM3 (Japan)
AWI-CM (Germany)	CMCC (Italy)	INM (Russia)	NICAM (Japan)
BCC (China)	CNRM (France)	IITM-ESM (India)	NorESM (Norway)
BESM (Brazil)	EC-Earth3 (Europe)	IPSL-CM6 (France)	NUIST (China)
BNU (China)	EMAC (Germany)	K-ACE (Republic of Korea)	TaiESM (Taiwan)
CAMS-CMS (China)	FGOALS (China)	MIROC-CGCM (Japan)	UKESM (UK)
CanESM (Canada)	FIO (China)	MIROC-ESM (Japan)	VRESM (South Africa/Australia)
CasESM (China)	GFDL (USA)	MPI-ESM (Germany)	
CESM2 (USA)	GISS (USA)	MRI-ESM2 (Japan)	

Citation services in support of CMIP6 also require registration

- Citation services provide those using CMIP model output with DOI's and other information needed to properly cite and provide credit to modeling centers.
- Registration of at least one individual from each modeling center is required to record names, affiliations, etc. of those deserving credit for producing the CMIP6 data.
- The citation information will be accessible through URL's recorded as global attributes in each netCDF output file.

This talk is structured following the CMIP6 Guide

See <https://pcmdi.llnl.gov/CMIP6/Guide>

► Part 1: For participating climate modelers

- Requirements and expectations
- Experiment design
- **Forcing data sets**
 - Prepared by various expert groups
 - Encouraged to be contributed to a central archive (inputMIPS)
- Model output fields
- Model output requirements
- Software for preparing/checking output
- Archiving/publishing output
- Documentation process
- CMIP6 organization and governance

► Part 2: For ESGF data node managers and operators (being drafted)

- CDNOT
- Tracking and replication

► Part 3: For Users of CMIP6 model output (being drafted)

- Registering published work based on CMIP6
- obs4MIPs

Forcing datasets for CMIP6: Input4MIPs status

- Project initiated April 2016
- Purpose
 - ▶ To collect, version-control, and archive CMIP6 forcing data sets
 - ▶ To impose data and metadata standards facilitating use
- Forcing datasets description/status
 - <https://esgf-node.llnl.gov/projects/input4mips/>
- input4MIPs holdings to-date
 - ▶ 1647 files & 559 Gb of data
 - ▶ 3 CMIP panel releases: v6.0.0, v6.1.1 and v6.2.0
 - ▶ Data footprint expected to ~double+ over coming months (satellite MIP data)
- input4MIPs project summary being prepared for GMD CMIP6 special issue

Input4MIPs DECK/historical forcing data status

Forcing Dataset	Status	Temporal Coverage	Latest Data Version(s)	Contact
SLCF Emissions	Available	1750-01 to 2014-12	2017-05-18, 2017-08-30 (Aircraft; updated)	Steven Smith ssmith@pnnl.gov
Biomass Burning	Available	1750-01 to 2015-12	1.2 (2016-12-13; updated)	Margreet van Marle m.j.e.van.marle@vu.nl
CO2 and CH4 Emissions	Available	1750-01 to 2014-12	2017-05-18, 2017-08-30 (Aircraft; updated)	Steven Smith ssmith@pnnl.gov
Land-use	Available	850 to 2015	2.1h (2017-01-26)	George Hurtt gchurtt@umd.edu
GHG concentrations	Available	0-01 to 2015-12	1.2.0 (2016-07-01)	Malte Meinshausen malte.meinshausen@unimelb.edu.au
Ozone concentrations	Available	1850-01 to 2014-12	1.0 (2016-07-11)	Michaela Hegglin m.i.hegglin@reading.ac.uk
Nitrogen deposition	Available	1850-01 to 2014-12	2.0 (2016-12-07; updated)	Michaela Hegglin m.i.hegglin@reading.ac.uk
Simple plume aerosol	Available	1850 to 2100	1.0 (2017-02-01)	Bjorn Stevens bjorn.stevens@mpimet.mpg.de
Solar	Available	1850-01 to 2299-12	3.2 (2017-01-03; updated)	Katja Matthes kmatthes@geomar.de
Stratospheric aerosol	Available	1850-01 to 2014-12	3.0 (2017-10-04; updated)	Beiping Luo beiping.luo@env.ethz.ch
AMIP SST and SIC	Available	1870-01 to 2016-06	1.1.2 (2017-04-19; updated)	PCMDI pcmdi-cmip@llnl.gov

Download links, input4MIPs website: <https://esgf-node.llnl.gov/search/input4mips>

Also see the live google doc at <https://goo.gl/r8up31>

DECK/historical forcing data status

Satellite MIP	Status	Host(s); Version	Committed to input4MIPs	Contact
CFMIP	Ready	http://doi.org/10.5194/gmd-2016-70 ; 1.0	?	Mark Webb mark.webb@metoffice.gov.uk
DAMIP	Ready	1.0 (2017-08-14)	-	David Plummer david.plummer@canada.ca
DCPP	Ready	1.1 (2017-01-23)	-	Christophe Cassou christophe.cassou@cerfacs.fr
FAFMIP	Ready	http://www.met.reading.ac.uk/~jonathan/FAFMIP/ ; (2015-08-21)	Yes	Jonathan Gregory j.m.gregory@reading.ac.uk
HighResMIP	Ready	2.2.0.0-r1 (2017-05-05)	-	Malcolm Roberts malcolm.roberts@metoffice.gov.uk
LS3MIP	Unknown	-	?	Sonia Seneviratne sonia.seneviratne@ethz.ch
OMIP	Ready	http://data1.gfdl.noaa.gov/nomads/forms/core/COREv2.html http://amaterasu.ees.hokudai.ac.jp/~tsujino/JRA55-do-v1.2/ CORE (Ready); JRA55-do v1.2 (Ready)	Yes	Gokhan Danabasoglu gokhan@ucar.edu
PMIP	Unknown	https://pmip4.lsce.ipsl.fr/doku.php ; ?	Yes	Masa Kageyama Masa.Kageyama@lsce.ipsl.fr
RFMIP	Ready	0.4 (2017-01-18)	-	Robert Pincus Robert.Pincus@colorado.edu
ScenarioMIP	Ready / In Prep.	Land-use – 2.1f (2017-10-05); emissions (in prep.)	-	Detlef van Vuuren Detlef.vanVuuren@pbl.nl
VolMIP	Ready	3.0 (2017-10-04); EVA module (Ready – GMD below) https://doi.org/10.5194/gmd-9-4049-2016	-/Yes	Davide Zanchettin davidoff@unive.it

Forcing datasets for CMIP6: Input4MIPs issues & future

- Rush to meet “deadlines” resulted in:
 - Errors (QC insufficient)
 - Insufficient review and feedback by other experts
 - Retraction/replacement of original release of data
- Resulting CMIP6 piControl and historical simulations now span 3 data collection versions
 - 6.0.0 (20th December 2016) - JAMSTEC simulations relied on this
 - 6.1.1 (22nd May 2017)
 - 6.2.0 (11th September 2017)
- Different approach needed for CMIP7
 - A procedure for better vetting/testing of forcing datasets is needed
 - Contact Paul Durack (durack1@lnl.gov) if you wish to contribute to a white paper suggestion options

This talk is structured following the CMIP6 Guide

See <https://pcmdi.llnl.gov/CMIP6/Guide>

- ▶▶▶ Part 1: For participating climate modelers
 - Requirements and expectations
 - Experiment design
 - Forcing data sets
 - **Model output fields (Data Request)**
 - Model output requirements
 - Software for preparing/checking output
 - Archiving/publishing output
 - Documentation process
 - CMIP6 organization and governance

- ▶▶▶ Part 2: For ESGF data node managers and operators (being drafted)
 - CDNOT
 - Tracking and replication

- ▶▶▶ Part 3: For Users of CMIP6 model output (being drafted)
 - Registering published work based on CMIP6
 - obs4MIPs

CMIP data request tools and documentation (Martin Juckes)

- Information available 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](#)

When problems are found, raise an issue!
"CMIP6_DataRequest_VariableDefinitions"

See [Version 01.beta.38 Release Notes](#) for more details

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 defined:
 - e.g., `standard_name`, `units`, `cell_methods`
 - CMOR harvests information and makes it easier to conform to the CMIP requirements
- Status
 - First release last December
 - ~Monthly updates have been made to correct errors and respond to MIP refinements

CMIP data request status and issues

- Variable definitions

- Ongoing revisions respond to MIP requests and modeling group clarification of the definition of variables
- List of missing standard names still significant but shrinking

- Outstanding issue & lessons:

- Scientific review needed to check that key variables are not missed from endorsed-MIP experiments
 - Who should do this?
- The complexity of choices (tiers, priorities, science objectives, MIPs, time slices) is confusing to many, especially the fact that different MIPs have different priorities
 - Can we find a simpler approach?
- MIPs with no technical expertise / familiarity with the data standards and experimental design framework require considerable hand-holding
 - Should we impose new criteria for MIP endorsement?

Many feel that **insufficient resources have been available for preparing the data request**

This talk is structured following the CMIP6 Guide

See <https://pcmdi.llnl.gov/CMIP6/Guide>

►►► Part 1: For participating climate modelers

- Requirements and expectations
- Experiment design
- Forcing data sets
- Model output fields (Data Request)
- **Model output requirements**
- Software for preparing/checking output
- Archiving/publishing output
- Documentation process
- CMIP6 organization and governance

►►► Part 2: For ESGF data node managers and operators (being drafted)

- CDNOT
- Tracking and replication

►►► Part 3: For Users of CMIP6 model output (being drafted)

- Registering published work based on CMIP6
- obs4MIPs

Standard metadata are recorded in all CMIP6 files

- Identify, for example:
 - » Variable
 - » Experiment
 - » Model
 - » Institution
 - » Sponsoring MIP
 - » Grid information
- Controlled vocabularies (CV's) ensure that metadata can be interpreted by infrastructure software.
 - » Reference CVs hosted at: https://github.com/WCRP-CMIP/CMIP6_CVs
- See "CMIP6_global_attributes_filenames_CVS" document
 - » <http://goo.gl/v1drZI>
 - linked from https://www.earthsystemcog.org/projects/wip/position_papers

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.

CMIP6 global attribute see note 1	description	examples	corresponding attribute in CMIP5	form see note 2	when required?	further information and rationale
activity_id	activity identifier(s)	"CMIP", "PMIP", "LS3MIP LUMIP" see note 3	project_id	CV	always	renamed more generically, since not all activities are projects; also multiple activities may now be listed separated by single spaces.
branch_method	branching procedure	"standard", "none provided", "no parent" see note 4	-	free form	whenever parent exists	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 "no parent"
branch_time_in_child	branch time with respect to child's time axis	365.0D0, 0.0D0 see note 5	-	double precision float	whenever parent exists	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.
branch_time_in_parent	branch time with respect to parent time axis	3650.0D0 see note 5	branch_time	double precision float	whenever parent exists	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.

Further information about data requirements:

- Reference "controlled vocabularies" (CV's) for CMIP6
 - ▶ https://github.com/WCRP-CMIP/CMIP6_CVs
- Specifications for file names, directory structures, and CMIP6 Data Reference Syntax (DRS)
 - ▶ <http://goo.gl/v1drZI>
- Specification of output file content, structure, and metadata
 - ▶ not yet available,
 - ▶ with notable exceptions will follow CMIP5 requirements.
 - ▶ Use of CMOR3 will ensure compliance

This talk is structured following the CMIP6 Guide

See <https://pcmdi.llnl.gov/CMIP6/Guide>

► Part 1: For participating climate modelers

- Requirements and expectations
- Experiment design
- Forcing data sets
- Model output fields (Data Request)
- Model output requirements
- **Software for preparing/checking output**
- Archiving/publishing output
- Documentation process
- CMIP6 organization and governance

► Part 2: For ESGF data node managers and operators (being drafted)

- CDNOT
- Tracking and replication

► Part 3: For Users of CMIP6 model output (being drafted)

- Reporting suspected errors/Errata
- Registering published work based on CMIP6
- obs4MIPs

CMOR3 / PrePARE facilitate (and check) conformance of files to CMIP6 requirements

- **CMOR3: (Climate Model Output Rewriter 3)**
 - Code for writing model output following CMIP6 specs
 - Code available at <https://github.com/PCMDI/cmor>
 - Documentation available at <http://cmor.llnl.gov/>
- **PrePARE (Pre-Publication Attribute Reviewer)**
 - Code for checking some metadata for compliance with CMIP6 specs
 - Call the CF-checker to check compliance with the CF-conventions
 - See https://cmor.llnl.gov/mydoc_cmip6_validator/
- **Status**
 - Major development completed about a year ago
 - Several enhancements made this past year
 - CMOR being used by CMIP6 modeling centers, input4MIPs, obs4MIPs
 - PrePARE being used by modeling groups not using CMOR3 and ESGF to screen out non-conforming datasets

This talk is structured following the CMIP6 Guide

See <https://pcmdi.llnl.gov/CMIP6/Guide>

► Part 1: For participating climate modelers

- Requirements and expectations
- Experiment design
- Forcing data sets
- Model output fields (Data Request)
- Model output requirements
- Software for preparing/checking output
- **Archiving/publishing output**
- Documentation process
- CMIP6 organization and governance

► Part 2: For ESGF data node managers and operators (being drafted)

- CDNOT
- Tracking and replication

► Part 3: For Users of CMIP6 model output (being drafted)

- Registering published work based on CMIP6
- obs4MIPs

The Earth System Grid Federation (ESGF) status

- Includes funded partners worldwide.
 - ▶▶▶ DOE, IS-ENES, NASA, NCI, NOAA
 - ▶▶▶ International Executive Committee
- Development is organized around 18 task teams
- Annual face-to-face meeting summarizes progress in an annual report
 - ▶▶▶ https://esgf.llnl.gov/media/pdf/2017-ESGF_F2F_Conference_Report.pdf

ESGF task teams

Task Team	Focus
CoG User Interface	Improved ESGF search and data cart management and interface
Compute	Developing the capability to enable data analytics within ESGF
Dashboard	Statistics related to ESGF usage
Data Transfer	ESGF data transfer and enhancement of the web-based download
Documentation	Document the ESGF software stack
Identity Entitlement Access	ESGF X.509 certificate-based authentication and improved interface
Installation	Installation of the components of the ESGF software stack
International Climate Network	Increase data transfer rates between the ESGF climate data centers
Metadata and Search	ESGF search engine based on Solr5; discoverable search metadata
Node Manager	Management of ESGF nodes and node communications
Provenance Capture	ESGF provenance capture for reproducibility and repeatability
Publication	Capability to publish data sets for CMIP and other projects to ESGF
Quality Control	Integration of external information into the ESGF portal
Replication	Replication tool for moving data from one ESGF center to another
Software Security	Security scans to identify vulnerabilities in the ESFF software
Tracking / Feedback Notification	User and node notification of changed data in the ESGF ecosystem
User Support	User frequently asked questions regarding ESGF and housed data
Versioning	Managing multiple versions of ESGF published data sets

ESGF planning documents for CMIP6

Living Document	Web Link	Description
Governance	https://esgf.llnl.gov/media/pdf/ESGF_Governance_6_4_2015.pdf	ESGF governance structure
Strategic Roadmap	https://esgf.llnl.gov/media/pdf/2015-ESGF-Strategic-Plan.pdf	ESGF mission and international integration strategy for data, database and computational architecture
Implementation Plan	https://esgf.llnl.gov/media/pdf/ESGF-Implementation-Plan-V1.0.pdf	ESGF data management system: deployment, installation, and transition to an operational system.
Software Security Plan	https://esgf.llnl.gov/media/pdf/ESGF-Software-Security-Plan-V1.0.pdf	Systematic approach to releasing a secure ESGF software stack to the community
Policies and Guidelines	https://esgf.llnl.gov/media/pdf/ESGF-Policies-and-Guidelines-V1.0.pdf	Uniform standards for all participants/contributors in the global infrastructure environment
CA Policy	https://esgf.llnl.gov/media/pdf/ESGF-CA-V2.0.pdf	Rules and procedures established for the operation of the ESGF Root CA PKI services
Tier 1 and 2 Node Site Requirements	https://esgf.llnl.gov/media/pdf/ESGF-Tier1and2-NodeSiteRequirements-V5.pdf	Requirements for two classes of distinguished ESGF node types: Tier 1 and 2 node sites
Logo Requirements	https://esgf.llnl.gov/logo_requirements.html	Logo requirements and guidelines for ESGF and CoG

Further information: ESGF-supporting websites

Earth System Grid Federation

esgf.llnl.gov

Living documents (i.e., governance, strategic plan, security plan, etc.)

esgf.llnl.gov → "Documents"

Reports

esgf.llnl.gov/reports.html

Conferences

esgf.llnl.gov/conferences.html

Letters of Support

esgf.llnl.gov/letters-of-support.html

Presentations

<https://esgf.llnl.gov/presentations.html>

Committee Members

esgf.llnl.gov/committee.html

Confluence / JIRA

acme-climate.atlassian.net/wiki/display/ESGF/Earth+System+Grid+Federation;confluence/JIRA

GitHub (software repository)

github.com/esgf; software repository website

Network

icnwg.llnl.gov

Tutorial

www.earthsystemcog.org/projects/cog/tutorials_web; CoG tutorial

Further information: ESGF planning documents

Living Document	Web Link	Description
Governance	https://esgf.llnl.gov/media/pdf/ESGF_Governance_6_4_2015.pdf	ESGF governance structure
Strategic Roadmap	https://esgf.llnl.gov/media/pdf/2015-ESGF-Strategic-Plan.pdf	ESGF mission and international integration strategy for data, database and computational architecture
Implementation Plan	https://esgf.llnl.gov/media/pdf/ESGF-Implementation-Plan-V1.0.pdf	ESGF data management system: deployment, installation, and transition to an operational system.
Software Security Plan	https://esgf.llnl.gov/media/pdf/ESGF-Software-Security-Plan-V1.0.pdf	Systematic approach to releasing a secure ESGF software stack to the community
Policies and Guidelines	https://esgf.llnl.gov/media/pdf/ESGF-Policies-and-Guidelines-V1.0.pdf	Uniform standards for all participants/contributors in the global infrastructure environment
CA Policy	https://esgf.llnl.gov/media/pdf/ESGF-CA-V2.0.pdf	Rules and procedures established for the operation of the ESGF Root CA PKI services
Tier 1 and 2 Node Site Requirements	https://esgf.llnl.gov/media/pdf/ESGF-Tier1and2-NodeSiteRequirements-V5.pdf	Requirements for two classes of distinguished ESGF node types: Tier 1 and 2 node sites
Logo Requirements	https://esgf.llnl.gov/logo_requirements.html	Logo requirements and guidelines for ESGF and CoG

This talk is structured following the CMIP6 Guide

See <https://pcmdi.llnl.gov/CMIP6/Guide>

►►► Part 1: For participating climate modelers

- Requirements and expectations
- Experiment design
- Forcing data sets
- Model output fields (Data Request)
- Model output requirements
- Software for preparing/checking output
- Archiving/publishing output
- **Documentation process**
 - Model and experiment documentation
 - Errata service
- CMIP6 organization and governance

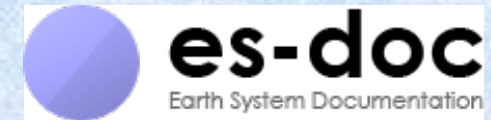
►►► Part 2: For ESGF data node managers and operators (being drafted)

- CDNOT
- Tracking and replication

►►► Part 3: For Users of CMIP6 model output (being drafted)

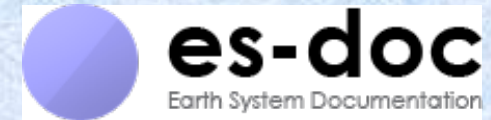
- Registering published work based on CMIP6
- obs4MIPs

ES-DOC for CMIP6 status



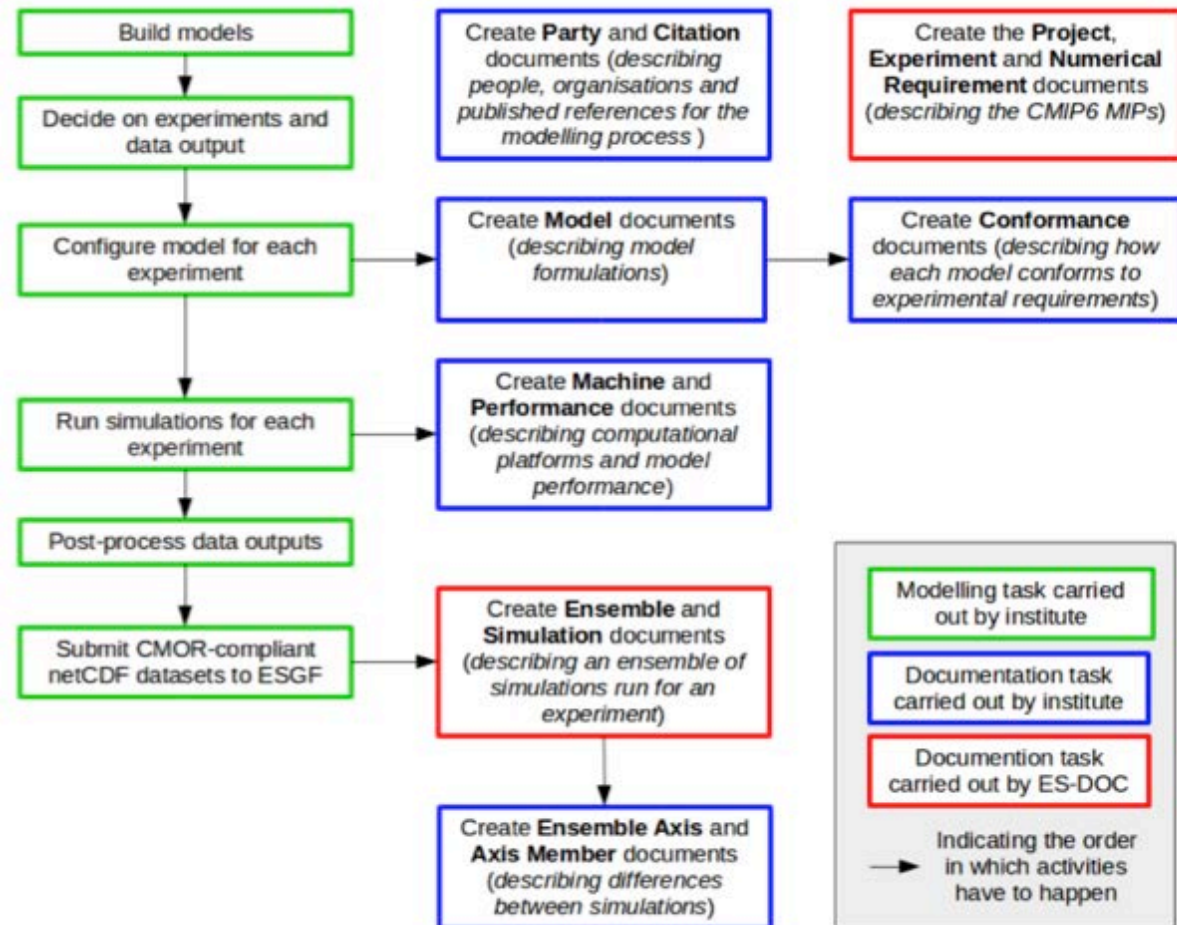
- CMIP6 documentation scope:
 - ➔ WGCM has a responsibility to document data output for users beyond the usual WGCM science community - this is a key issue for many stakeholders
- Designed so that process is easier for modelling groups:
 - ➔ Large fraction is automated
 - ➔ Option to start model description from CMIP5 version
 - ➔ Modular and agile process
 - ➔ Documentation for all steps (+ published WIP white paper)
- Community review :
 - ➔ Science contents of model documentation (realms, short tables) on-going (we need WGCM help to identify more science reviewers)
- Beta testing phase on-going (GFDL, IPSL and CCMA, IITM, MPI soon)
 - ➔ 20 liaison out of 34 groups identified (we need WGCM help to identify the others)
- Time line :
 - ➔ Science contents of realms ready Nov 1st (V1.0)
 - ➔ iPython notebook entry tool with CMIP5 seeding ready Dec 15th
 - ➔ Cdf2cim tool ready for ESGF ingestion
 - ➔ Community support tools (checklist, training, webcasts, ...) ready Dec 15th
 - ➔ Jan 1st for full end-to-end release for model and simulation documentation
 - ➔ Challenges due to reduced funding (IS-ENES gap)

ES-DOC for CMIP6 status



The documentation workflow :

- ▶ About half of the documents automated or ES-DOC generated
- ▶ The others produced by groups when ready
- ▶ Links from "further_info_URL":
 - Institute's general homepage
 - Description of the experiment
 - Scientific description of model
 - Description of the ensemble
 - Institute's own page
 - Dataset errata information
 - Citation information
 - Performance
 - Datasets in ESGF
- ▶ Note: conformance document will capture exact forcing used by groups



New ES-DOC errata service



- Records issues (problems) with published datasets
- Provides service for responding to queries about datasets identified by their "persistent identifiers" (PIDs)
 - Datasets are labeled with "persistent identifiers" (PIDs)
 - User can determine whether a queried version of dataset/file is safe to use or is
 - affected by an unresolved issue.
 - Has been superseded by a newer version
- In development:
 - Exposure of errata service to other services (such as the ESGF CoG front-end and Synda) to ensure real time, automated feedback on data status.
 - Incorporation of the issue declaration process in the conventional publishing workflow.
- November 2017 community release scheduled

This talk is structured following the CMIP6 Guide

See <https://pcmdi.llnl.gov/CMIP6/Guide>

►►► Part 1: For participating climate modelers

- Requirements and expectations
- Experiment design
- Forcing data sets
- Model output fields (Data Request)
- Model output requirements
- Software for preparing/checking output
- Archiving/publishing output
- Documentation process
- CMIP6 organization and governance

►►► Part 2: For ESGF data node managers and operators (being drafted)

- **CDNOT**
- Tracking and replication

►►► Part 3: For Users of CMIP6 model output (being drafted)

- Registering published work based on CMIP6
- obs4MIPs

Each modeling group should appoint a technical contact to join the *CMIP6 Data Node Operations (CDNOT)*

- The CDNOT (Sébastien Denvil, chair) is charged with
 - applying and operationalizing ESGF for CMIP6
 - Imposing technical requirements on ESGF nodes used to serve CMIP6 data
- The CDNOT serves to:
 - Communicate WIP discussion to modeling center representatives
 - Provide input to the WIP of data node/modeling center concerns
- Members represent each site hosting CMIP6 data (i.e., most modeling centers and major data centers)
- Membership overlaps with bodies responsible for requirements (WIP) and software development (ESGF, ESDOC, ...)

Establish (or partner with) an ESGF data node

- Tier 1: Serves multiple models and provides full suite of ESGF services, including
 - ≥ 10 petabytes of spinning disk storage space
 - ≥ 10 gigabits per second connection to a wide-area network provider
 - Run a 10 gbits/s perfSONAR host
 - Deploy at least four 10 gbits/s Data Transfer Nodes (DTNs)
 - Publish data using GridFTP and Globus URLs in addition to wget URLs,
 - Use Synda for data replication between Tier 1 sites.
- Tier 2
 - For centers that typically have fewer physical or staff resources available for ESGF but need to distribute CMIP6 data
 - Document describing minimum requirements under development

This talk is structured following the CMIP6 Guide

See <https://pcmdi.llnl.gov/CMIP6/Guide>

►►► Part 1: For participating climate modelers

- Requirements and expectations
- Experiment design
- Forcing data sets
- Model output fields (Data Request)
- Model output requirements
- Software for preparing/checking output
- Archiving/publishing output
- Documentation process
- CMIP6 organization and governance

►►► Part 2: For ESGF data node managers and operators (being drafted)

- CDNOT
- **Tracking, replication, and citation**

►►► Part 3: For Users of CMIP6 model output (being drafted)

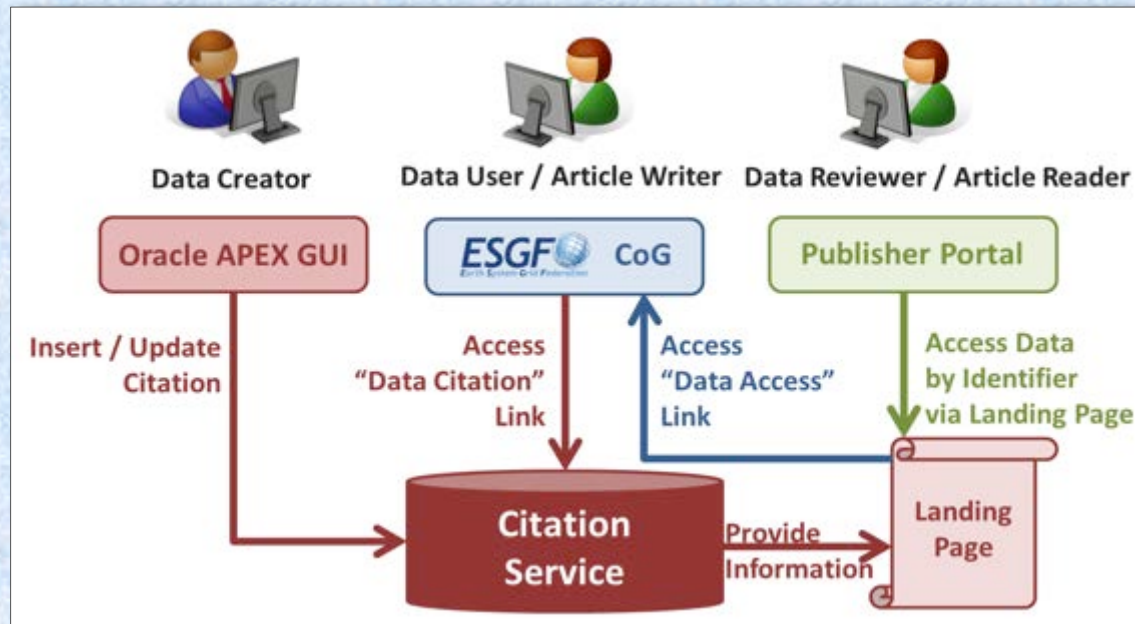
- Registering published work based on CMIP6
- obs4MIPs

Citation and data tracking

- DOI's will be assigned at a fairly high level (model/experiment?)
 - Data granularity: DataCite DOI together with citation reference will be assigned to the collection of data from a single experiment and model
 - A reasonably short list of DOI's plus citation reference can be included in publications.
 - Main requirement: ensure proper citation of data acknowledging contributions by modeling groups
- Persistent IDs (PIDs) will be assigned at fine granularity
 - Data granularity: PIDs are assigned to
 - Each CMIP6 NetCDF file during the ESGF data publication and
 - The collection of files comprising an atomic dataset
 - Web service planned for recording lists of PIDs along with citation info. for CMIP6 publications.
 - ES-DOC errata services will be PID-based
 - Potential use of PIDs in replication workflow.

Data Citation Service: Workflow

Data Citations are provided on Model/MIP and experiment granularities.



- Workflow for Modeling Centers:

- Contact person for citation required
- Provide at least basic data citations for model data: author lists + titles via GUI
- First data delivered to ESGF ◇ DataCite DOI registered
- Update and extend model data citations and add citations for simulation data via API
- At AR6 WG I literature-cutoff date, an "snapshot" of the archive will be recorded and citation information will be frozen.

Data Citation Service: Information and Status

Data citations are provided on Model/MIP and experiment granularities.

The screenshot shows the 'Test Project' web interface. At the top, there is a navigation bar with 'Home', 'About Us', 'Resources', and 'Contact Us'. A 'Technical Support' button is also visible. Below the navigation bar is a search bar with the text 'Enter Text:' and a search button. To the right of the search bar, there are options for 'Display' (set to 50) and 'results per page', along with a link to 'More Search Options'. Below the search bar, there are checkboxes for 'Show All Replicas', 'Show All Versions', and 'Search Local Node Only (Including All Replicas)'. The search constraints are listed as 'x CMIP6 | x ssp126'. The total number of results is 1. Below this, there is a link to 'Please login to add search results to your Data Cart' and a note for 'Expert Users: you may display the search URL and return results as XML or return results as JSON'. The main content area displays a single result: '1. cmip6.ScenarioMIP.PCMDI.PCMDI-test-1-0.ssp126.r1i1p1f1.6hrLev.ua.gn'. Below this, there is a 'Data Citation (Landing Page)' section with the following information: 'Data Node: esgf-dev.dkrz.de', 'Version: 20170303', 'Total Number of Files (for all variables): 1', and 'Full Dataset Services: [Show Metadata] [List Files] [THREDDS Catalog] [WGET Script] [Hide Citation] [PID]'. The 'Data Citation' section includes the 'Identifier URL: http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.ScenarioMIP.PCMDI.PCMDI-test-1-0', 'Creators: Jungclaus, Johann; Giorgetta, Marco; Reick, Christian et al.', 'Title: PCMDI PCMDI-test1.0 model output prepared for CMIP6 ScenarioMIP', 'Publisher: Earth System Grid Federation', and 'Publication Year: 2017'.

- Information and status: (responsible party for citation service: DKRZ)
 - General information: <http://cmip6cite.wdc-climate.de>
 - Graphical User Interface (released: early 2017): <http://cera-www.dkrz.de/citeXA>
 - Tool to provide citation information in user test phase (release planned for early 2018)
 - Data citation accessible in ESGF CoG portal (part of next ESGF publisher release)
 - Citations will be integrated in furtherInfo URL landing page ◊ discussion with ES-DOC
 - Citation service used also for input4MIPs, e.g. <https://doi.org/10.22033/ESGF/input4MIPs.1122>.
 - Registered citation managers (2017-09-29): CMIP6 - 21; input4MIPs - 13

This talk is structured following the CMIP6 Guide

See <https://pcmdi.llnl.gov/CMIP6/Guide>

► Part 1: For participating climate modelers

- Requirements and expectations
- Experiment design
- Forcing data sets
- Model output fields (Data Request)
- Model output requirements
- Software for preparing/checking output
- Archiving/publishing output
- Documentation process
- CMIP6 organization and governance

► Part 2: For ESGF data node managers and operators (being drafted)

- CDNOT
- Tracking and replication

► Part 3: For Users of CMIP6 model output (being drafted)

- **Registering published work based on CMIP6**
- obs4MIPs

Users of CMIP data are obliged to record publications as in CMIP5.

- PCMDI maintains a web-based service for recording publications based on CMIP output
 - ➔ <https://cmip-publications.llnl.gov/>
- This has been recently improved to facilitate entering input
 - ➔ Enter DOI and pre-populate most information
- We collect additional (non-mandatory) input identifying the output used, including which experiment, models, and variables
- We plan to provide functionality for recording a list of “tracking i.d’s” documenting the data relied on by a study.
 - ➔ Permanent record of provenance
 - ➔ Can be used to meet government and scientific mandates to make data available.

This talk is structured following the CMIP6 Guide

See <https://pcmdi.llnl.gov/CMIP6/Guide>

►►► Part 1: For participating climate modelers

- Requirements and expectations
- Experiment design
- Forcing data sets
- Model output fields (Data Request)
- Model output requirements
- Software for preparing/checking output
- Archiving/publishing output
- Documentation process
- CMIP6 organization and governance

►►► Part 2: For ESGF data node managers and operators (being drafted)

- CDNOT
- Tracking and replication

►►► Part 3: For Users of CMIP6 model output (being drafted)

- Registering published work based on CMIP6
- **obs4MIPs**
 - Data requirements based on CMIP
 - Peter Gleckler to describe tomorrow

CMIP6 infrastructure closing remarks

- Progress in developing all components of the infrastructure is satisfactory but taxing the resources of nearly all contributors
- The WIP has helped identify and prevent incompatibilities from developing across the many infrastructure components.
- The establishment of systematic and well-defined procedures, vocabularies, versioning systems, citation methods, and standards for forcing datasets (input4MIPs) provide a foundation for evolving toward future CMIP phases.

We invite questions/input.

