CMIP6 Infrastructure Status and Future

Karl E. Taylor and V. Balaji

With contributions from
The WGCM Infrastructure Panel

Presented at the
CMIP6 Model Analysis Workshop

Barcelona, Spain
26 March 2019

Purpose of CMIP Infrastructure

- Facilitate access to model output produced by MIPs (e.g., CMIP and the CMIP6 endorsed MIPs; obs4MIPs; CORDEX)
- Impose standards that make datasets self-describing
- Define “controlled vocabularies” to enable automated management of data and user-friendly search capabilities
- Provide information and services for
  - Documenting models and experiments
  - Assigning persistent i.d.’s for data citation purposes
  - Reporting errors in data
  - Helping contributors and users of CMIP data to do their work
  - ….

CMIP infrastructure support

- PCMDI
  - DOE has provided 30-years of MIP support
- ESGF
  - Originated by U.S. DOE
  - More recent major contributions from numerous others
- IS-ENES
  - European contribution to ESGF & CMIP infrastructure
- Numerous other projects and institutions, including CEDA, DKRZ, es-doc, IPSL, NASA, NOAA

Centre for Environmental Data Analysis
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

DKRZ

es-doc
Earth System Documentation

Institut Pierre Simon Laplace

WCRP CMIP6
World Climate Research Programme

Purposes of the WGCM Infrastructure Panel*

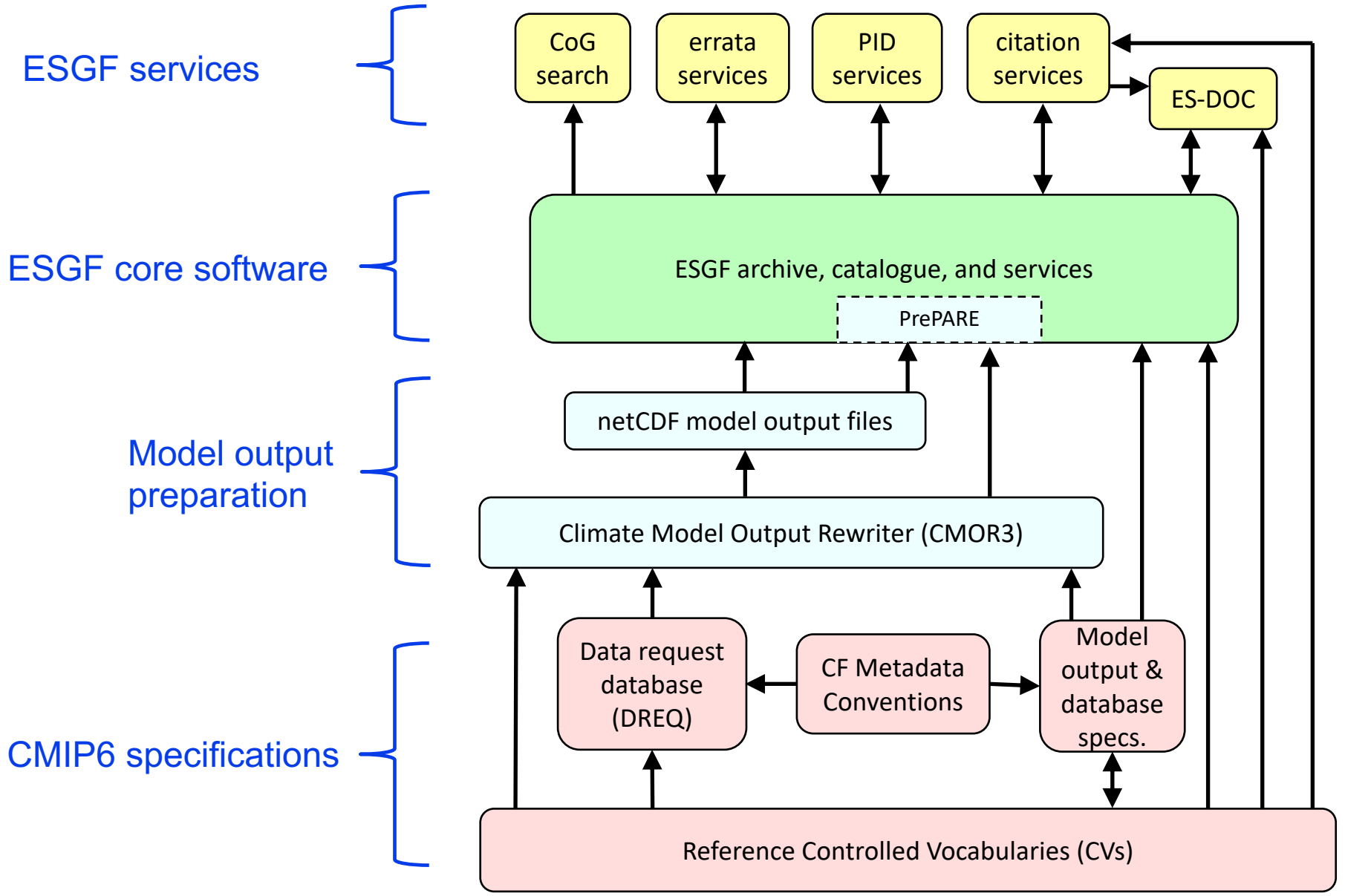
- To set requirements ensuring the infrastructure will serve its purpose
- To write documents defining specifications for the infrastructure and data it hosts.
- To help coordinate development work done under independently-funded projects so that the infrastructure elements work well together
- To communicate and coordinate with data managers at modeling groups via a “CMIP Data Node Operations Team” (CDNOT; Sébastien Denvil, chair)

* Y. Bai, V. Balaji, L. Cinquini, S. Denvil, P. Durack, F. Guglielmo, E. Guilyardi, M. Juckes, S. Kharin, M. Lautenschlager, B. Lawrence, M. Mizielinski, K. Taylor

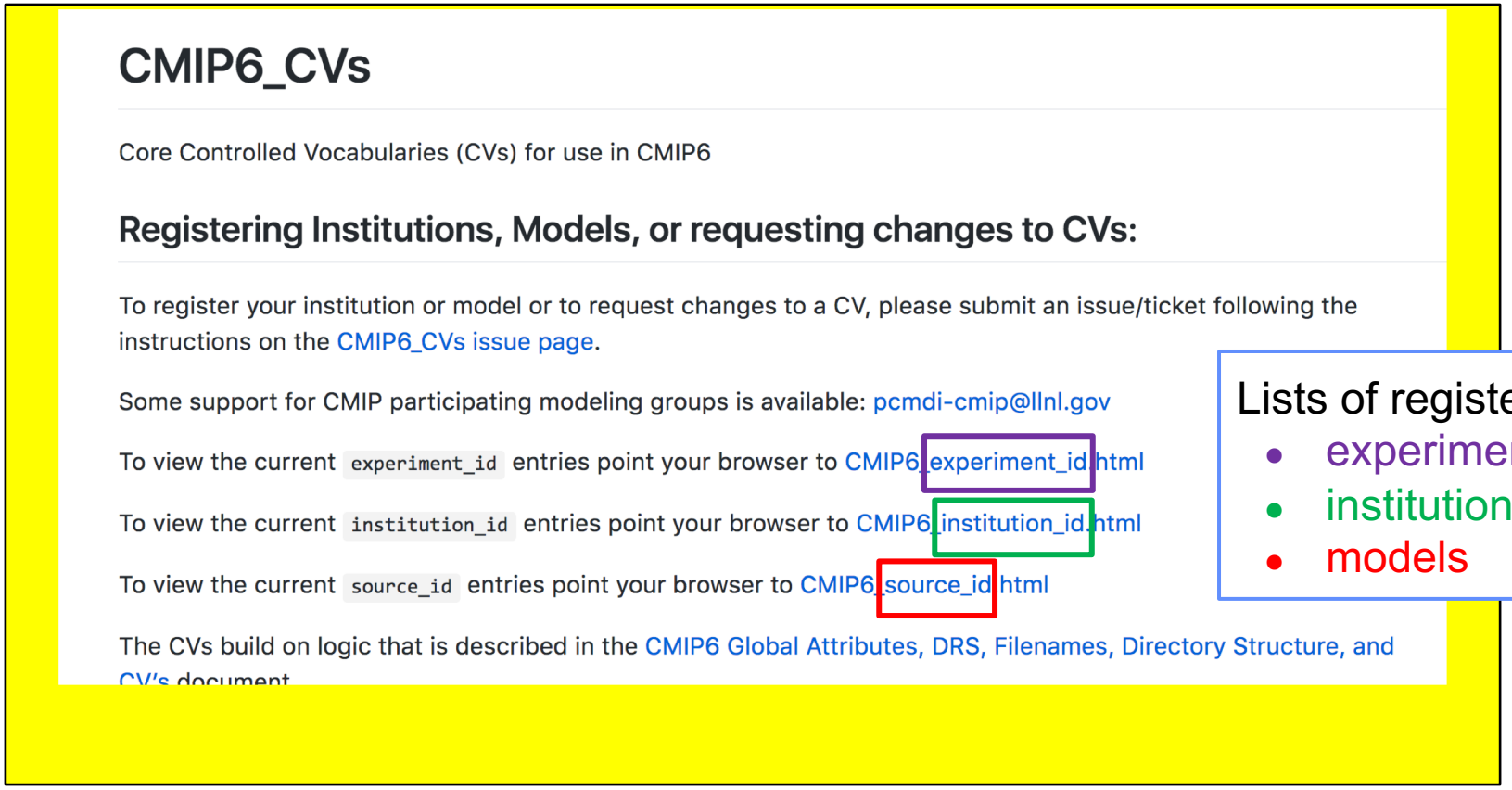
Objectives of this talk

- To provide an overview of the components of the CMIP infrastructure and their status
- To highlight tools and information that may help researchers obtain the data
- Look to the future … cloud-based opportunities

Infrastructure components and dependencies



Formally registered experiments, institutions, and models are recorded by "*controlled vocabularies*" (CV's)



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 following the instructions on the CMIP6_CVs issue page.

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

To view the current `institution_id` entries point your browser to CMIP6_institution_id.html

To view the current `source_id` entries point your browser to CMIP6_source_id.html

The CVs build on logic that is described in the CMIP6 Global Attributes, DRS, Filenames, Directory Structure, and CV's document

Lists of registered:

- experiments
- institutions
- models

All CV's are hosted at:
https://github.com/WCRP-CMIP/CMIP6_CVs

WCRP-CMIP CMIP6_CVs version: 6.2.15.1

Table of Registered Models

Show 5 entries

Search:

source_id	institution id	release year	activity participation	cohort	label	label extended	atmos	natNomRes atmos	ocean	natNomRes ocean	lan
ACCESS-CM2	CSIRO-ARCCSS-BoM	2018	CMIP FAFMIP OMIP RFMIP ScenarioMIP	Registered	ACCESS-CM2	Australian Community Climate and Earth System Simulator Climate Model Version 2	MetUM-HadGEM3-GA7.1 (N96; 192 x 144 longitude/latitude; 85 levels; top level 85 km)	250 km	ACCESS-OM2 (GFDL-MOM5, tripolar primarily 1deg; 360 x 300 longitude/latitude; 50 levels; top grid cell 0-10 m)	100 km	none
ACCESS-ESM1-5	CSIRO	2018	C4MIP CDRMIP CMIP OMIP RFMIP ScenarioMIP	Registered	ACCESS-ESM1.5	Australian Community Climate and Earth System Simulator Earth System Model Version 1.5	HadGAM2 (r1.1, N96; 192 x 145 longitude/latitude; 38 levels; top level 39255 m)	250 km	ACCESS-OM2 (MOM5, tripolar primarily 1deg; 360 x 300 longitude/latitude; 50 levels; top grid cell 0-10 m)	100 km	none
ARTS-2-3	UHH	2015	RFMIP	Registered	ARTS 2.3	ARTS 2.3 (Current development version of the Atmospheric Radiative Transfer Simulator)	none	none	none	none	none
AWI-CM-1-1-HR	AWI	2018	CMIP CORDEX HighResMIP OMIP SIMIP VIACSAB	Registered	AWI-CM 1.1 HR	AWI-CM 1.1 HR	ECHAM6.3.04p1 (T127L95 native atmosphere T127 gaussian grid; 384 x 192 longitude/latitude; 95 levels; top level 80 km)	100 km	FESOM 1.4 (unstructured grid in the horizontal with 1306775 wet nodes; 46 levels; top grid cell 0-5 m)	25 km	none
AWI-CM-1-1-LR	AWI	2018	CMIP CORDEX HighResMIP OMIP SIMIP ScenarioMIP VIACSAB	Registered	AWI-CM 1.1 LR	AWI-CM 1.1 LR	ECHAM6.3.04p1 (T63L47 native atmosphere T63 gaussian grid; 192 x 96 longitude/latitude; 47 levels; top level 80 km)	250 km	FESOM 1.4 (unstructured grid in the horizontal with 126859 wet nodes; 46 levels; top grid cell 0-5 m)	50 km	none

MIPs

Model name

Showing 1 to 5 of 102 entries

Previous 1 2 3 4 5 … 21 Next

Major infrastructure components are operational

- Model output & data base specifications and global metadata requirements
  - Provides the foundation for much of the infrastructure

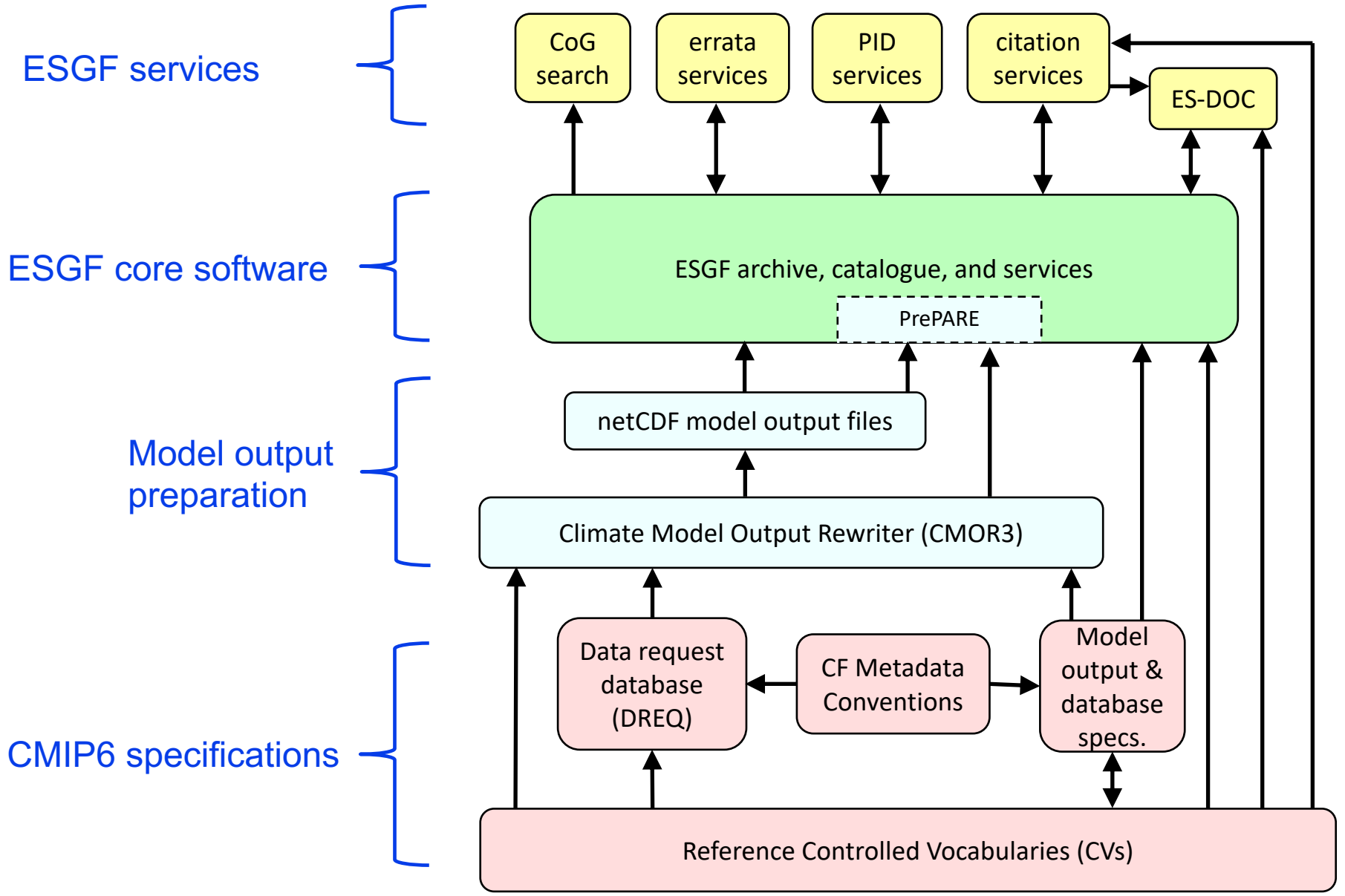
  https://pcmdi.llnl.gov/CMIP6/Guide/modelers.html#5-model-output-requirements
- CMIP data request tools and documentation (M. Juckes' poster)
  - Specifies variables to save and at what frequency for each experiment
  - Specifies variable-specific metadata to include in output files.

  https://www.earthsystemcog.org/projects/wip/CMIP6DataRequest
- CMOR can be used to write model output in conformance with CMIP requirements

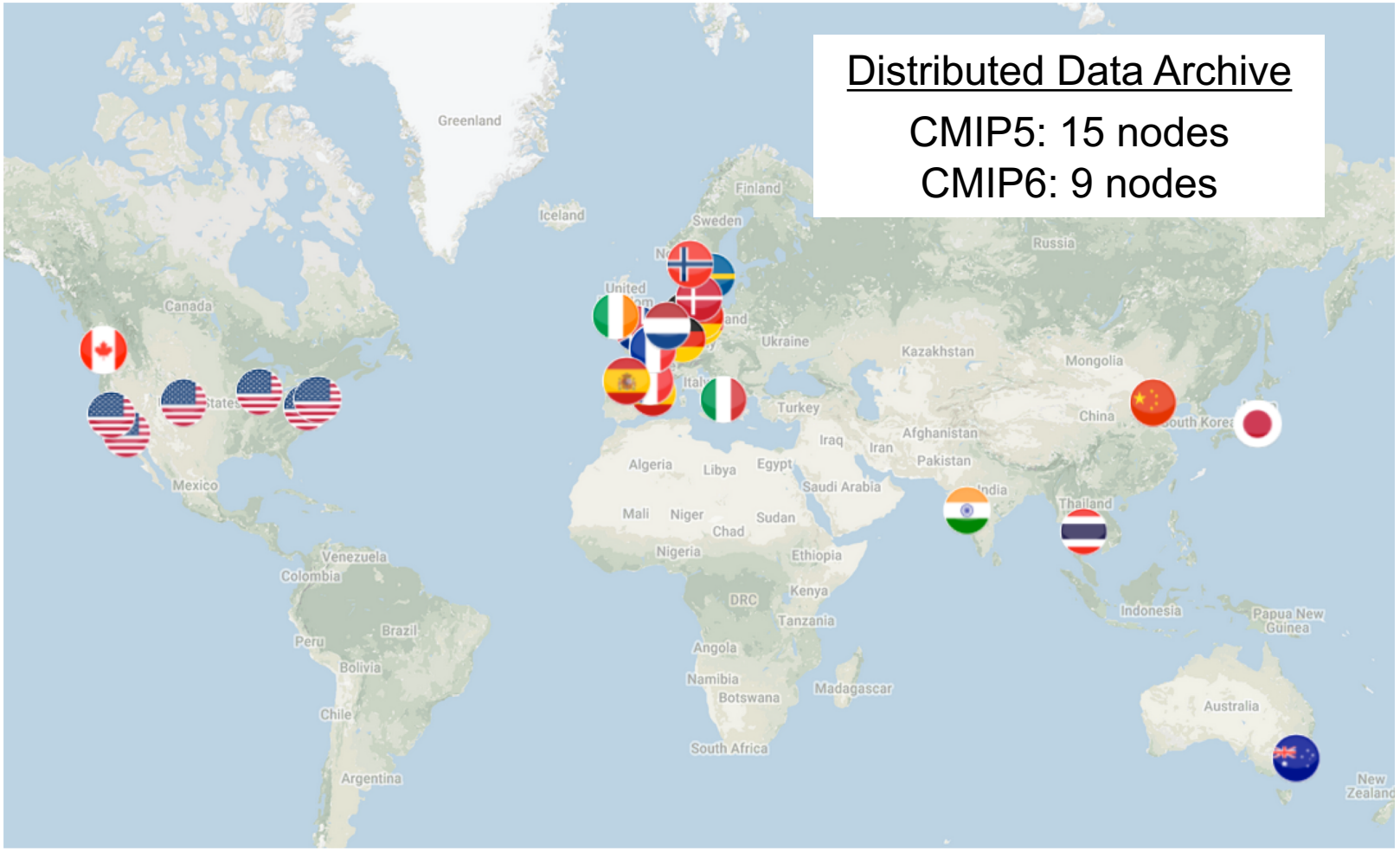
  https://github.com/PCMDI/cmor and http://cmor.llnl.gov/
- Pre-Publication Attribute Reviewer for ESGF (PrePARE)
  - new module checking file metadata before publication

WCRP CMIP6
World Climate Research Programme

Infrastructure components and dependencies

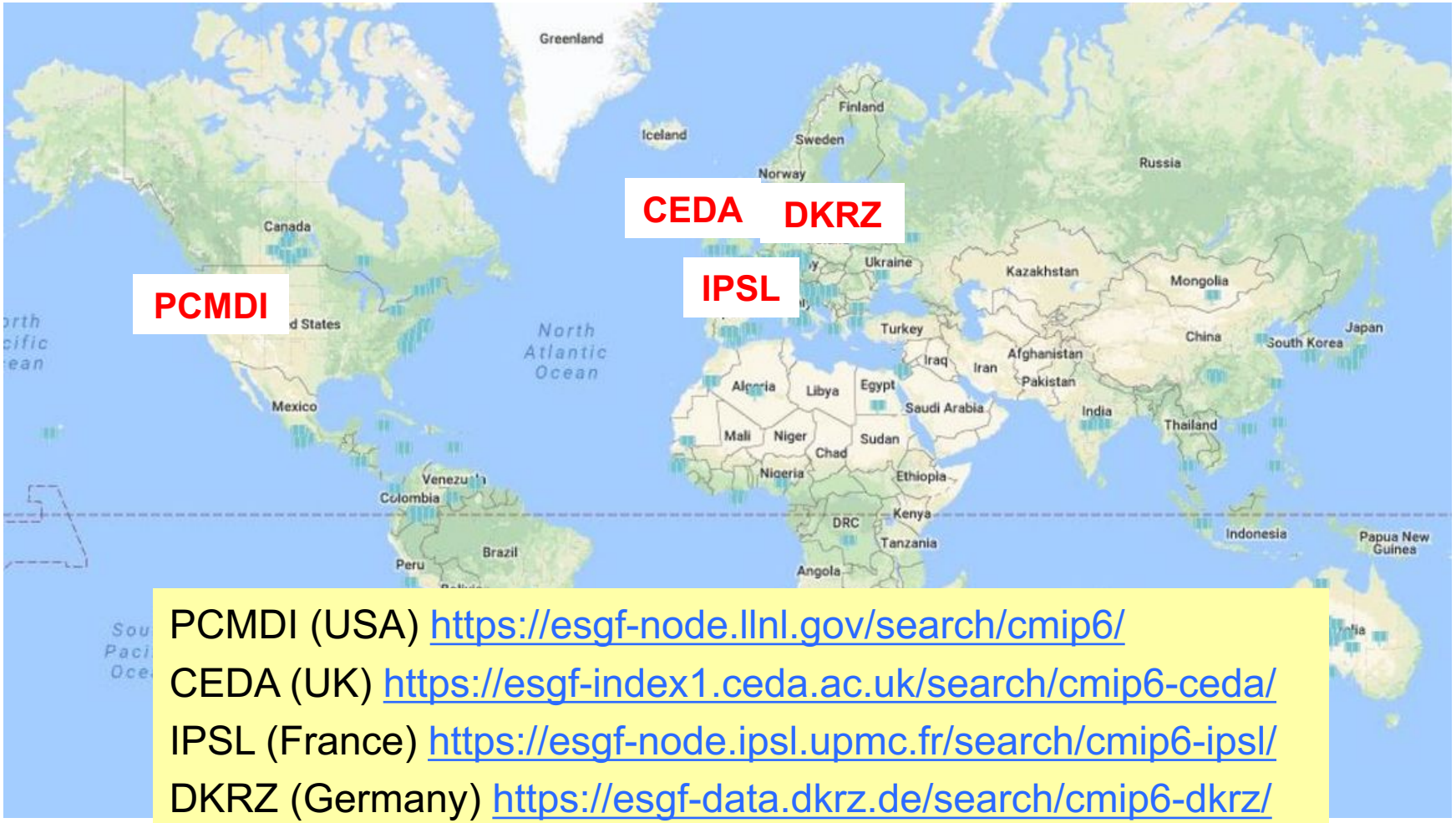


The Earth System Grid Federation (ESGF) develops and manages CMIP and other data



WCRP CMIP6
World Climate Research Programme

All CMIP6 data can be accessed from each of the 4 CMIP6 portals



PCMDI (USA) https://esgf-node.llnl.gov/search/cmip6/

CEDA (UK) https://esgf-index1.ceda.ac.uk/search/cmip6-ceda/

IPSL (France) https://esgf-node.ipsl.upmc.fr/search/cmip6-ipsl/

DKRZ (Germany) https://esgf-data.dkrz.de/search/cmip6-dkrz/

ESGF status for CMIP6

- Summary of data holdings available
  https://pcmdi.llnl.gov/CMIP6/ArchiveStatistics/esgf_data_holdings/
- More detailed info. available from portal search window
  - https://esgf-node.llnl.gov/search/cmip6/
  - 12 institutions; nearly 100 experiments
- Replication is ongoing at PCMDI, DKRZ, and CEDA
- To do (high priority): expand availability of Globus grid ftp to all sites.
- To do (high priority): provide server side computational services

ESGF services: CoG

- Provides users with an interface to browse and download all available data
- All 4 gateways have identical look and feel

Data citation services

Show Citation

1. **CMIP6.CFMIP.IPSL.IPSL-CM6A-LR.abrupt-0p5xCO2.r1i1p1f1.CFmon.albisccp.gr**
   Data Node: vesg.ipsl.upmc.fr
   Version: 20180605
   Total Number of Files (for all variables): 1
   Full Dataset Services: [Show Metadata] [List Files] [THREDDS Catalog] [WGET Script] [LAS] [Show Citation] [PID] [Globus Download] [Further Info]
   Add to Data Cart

Citation page

Metadata for 'CMIP6.CFMIP.IPSL.IPSL-CM6A-LR.abrupt-0p5xCO2'

General Information

General Information

Name CMIP6.CFMIP.IPSL.IPSL-CM6A-LR.abrupt-0p5xCO2

Abstract Coupled Model Intercomparison Project Phase 6 (CMIP6) data sets.
These data includes all datasets published for 'CMIP6.CFMIP.IPSL.IPSL-CM6A-LR.abrupt-0p5xCO2' according to the Data Reference Syntax defined as
'mip_era.activity_id.institution_id.source_id.experiment_id.member_id.table_id.variable_id.grid_label.version'.

The Earth System Model IPSL-CM6A-LR, released in 2017, includes the components:
atmos: LMDZ (NPv6, N96; 144 x 143 longitude/latitude; 79 levels; top level 40000 m), land: ORCHIDEE (v2.0, Water/Carbon/Energy mode), ocean: NEMO-OPA (eORCA1.3, tripolar primarily 1deg; 362 x 332 longitude/latitude; 75 levels; top grid cell 0-2 m), ocnBgchem: NEMO-PISCES, seaIce: NEMO-LIM3.
The model was run by the Institut Pierre Simon Laplace, Paris 75252, France (IPSL) in native nominal resolutions: atmos: 250 km, land: 250 km, ocean: 100 km, ocnBgchem: 100 km, seaIce: 100 km.

Project: These data have been generated as part of the internationally-coordinated Coupled Model Intercomparison Project Phase 6 (CMIP6; see also CMD Special Issue: http://www.geosci-model-dev.net/special_issue590.html). The

Cite this data

Citation **(2018)**. *IPSL IPSL-CM6A-LR model output prepared for CMIP6 CFMIP abrupt-0p5xCO2*. Earth System Grid Federation.
http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CFMIP.IPSL.IPSL-CM6A-LR.abrupt-0p5xCO2

BibTeX RIS

Model and experiment documentation

1. **CMIP6.CFMIP.IPSL.IPSL-CM6A-LR.abrupt-0p5xCO2.r1i1p1f1.CFmon.albisccp.gr**
   Data Node: vesg.ipsl.upmc.fr
   Version: 20180605
   Total Number of Files (for all variables): 1
   Full Dataset Services: [Show Metadata] [List Files] [THREDDS Catalog] [WGET Script] [LAS] [Show Citation] [PID] [Globus Download]
   [Further Info]

   Add to Data Cart

Further Info

further_info_url

es-doc
Earth System Documentation

WCRP World Climate Research Programme

CMIP6 Further Information v0.5.1.0

Support Help

Further Info URL: https://furtherinfo.es-doc.org/cmip6.ipsl.ipsl-cm6a-lr.dcppa-hindcast-niff.s2000.r1

ES-DOC Documentation

MIP Era	CMIP6
Institution	IPSL
Consortia	IPSL
Model	IPSL-CM6A-LR
Experiment	dcppA-hindcast-niff
Ensemble Description	N/A
Machine Performance	N/A

Dataset Documentation

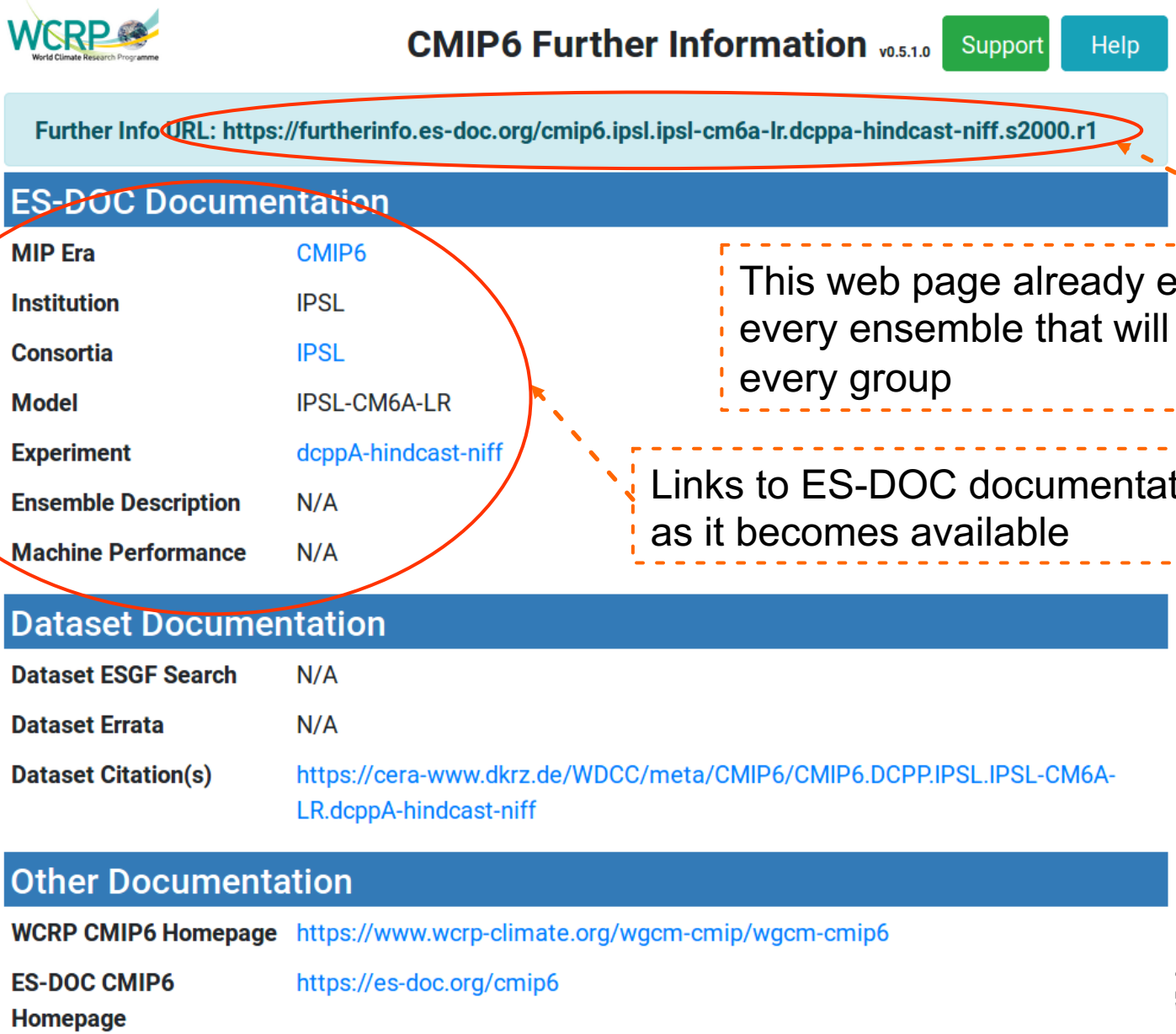
Dataset ESGF Search	N/A
Dataset Errata	N/A
Dataset Citation(s)	https://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.DCPP.IPSL.IPSL-CM6A-LR.dcppA-hindcast-niff

Other Documentation

WCRP CMIP6 Homepage	https://www.wcrp-climate.org/wgcm-cmip/wgcm-cmip6
ES-DOC CMIP6 Homepage	https://es-doc.org/cmip6

enes
RE FOR THE EUROPEAN NETWORK
STEM MODELLING

further_info_url

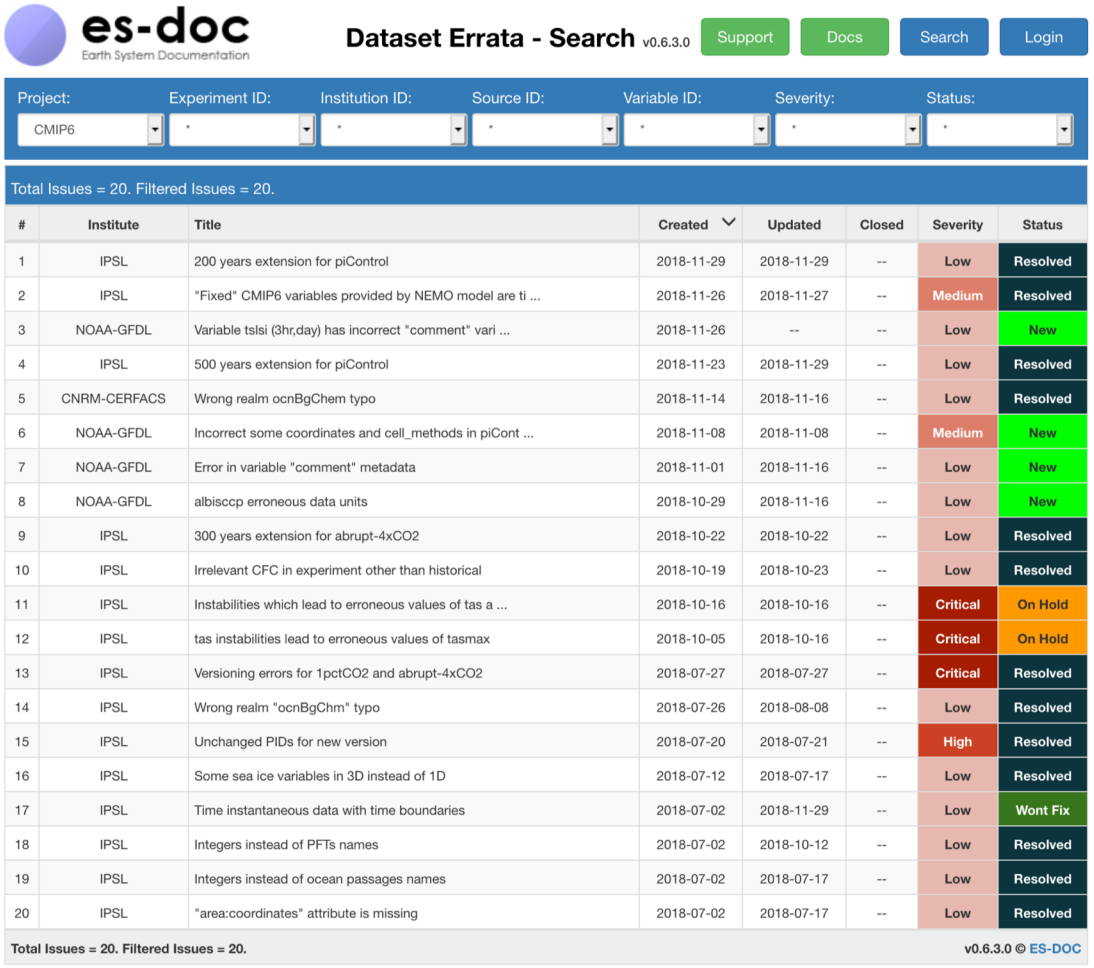


This web page already exists for every ensemble that will be run by every group

Links to ES-DOC documentation, as it becomes available

CMIP6 data errata is provided by ES-DOC

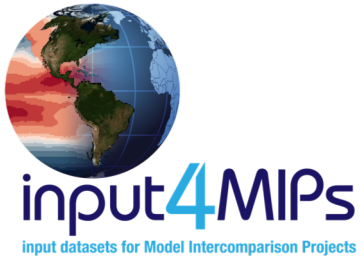
- Entry page https://search.es-doc.org/
- Any simulation/data problems must be reported using the ES-DOC errata service
- This will trigger a response (and resolution) by the contributing modelling group



#	Institute	Title	Created	Updated	Closed	Severity	Status
1	IPSL	200 years extension for piControl	2018-11-29	2018-11-29	--	Low	Resolved
2	IPSL	"Fixed" CMIP6 variables provided by NEMO model are ti ...	2018-11-26	2018-11-27	--	Medium	Resolved
3	NOAA-GFDL	Variable tslsi (3hr,day) has incorrect "comment" vari ...	2018-11-26	--	--	Low	New
4	IPSL	500 years extension for piControl	2018-11-23	2018-11-29	--	Low	Resolved
5	CNRM-CERFACS	Wrong realm ocnBgChem typo	2018-11-14	2018-11-16	--	Low	Resolved
6	NOAA-GFDL	Incorrect some coordinates and cell_methods in piCont ...	2018-11-08	2018-11-08	--	Medium	New
7	NOAA-GFDL	Error in variable "comment" metadata	2018-11-01	2018-11-16	--	Low	New
8	NOAA-GFDL	albisccp erroneous data units	2018-10-29	2018-11-16	--	Low	New
9	IPSL	300 years extension for abrupt-4xCO2	2018-10-22	2018-10-22	--	Low	Resolved
10	IPSL	Irrelevant CFC in experiment other than historical	2018-10-19	2018-10-23	--	Low	Resolved
11	IPSL	Instabilities which lead to erroneous values of tas a ...	2018-10-16	2018-10-16	--	Critical	On Hold
12	IPSL	tas instabilities lead to erroneous values of tasmax	2018-10-05	2018-10-16	--	Critical	On Hold
13	IPSL	Versioning errors for 1pctCO2 and abrupt-4xCO2	2018-07-27	2018-07-27	--	Critical	Resolved
14	IPSL	Wrong realm "ocnBgChm" typo	2018-07-26	2018-08-08	--	Low	Resolved
15	IPSL	Unchanged PIDs for new version	2018-07-20	2018-07-21	--	High	Resolved
16	IPSL	Some sea ice variables in 3D instead of 1D	2018-07-12	2018-07-17	--	Low	Resolved
17	IPSL	Time instantaneous data with time boundaries	2018-07-02	2018-11-29	--	Low	Wont Fix
18	IPSL	Integers instead of PFTs names	2018-07-02	2018-10-12	--	Low	Resolved
19	IPSL	Integers instead of ocean passages names	2018-07-02	2018-07-17	--	Low	Resolved
20	IPSL	"area:coordinates" attribute is missing	2018-07-02	2018-07-17	--	Low	Resolved

https://errata.es-doc.org/static/index.html

ESGF provides access to CMIP6 forcing datasets and some observations



- CMIP6 forcing datasets (P. Durack's Poster)

  https://esgf-node.llnl.gov/search/input4mips/

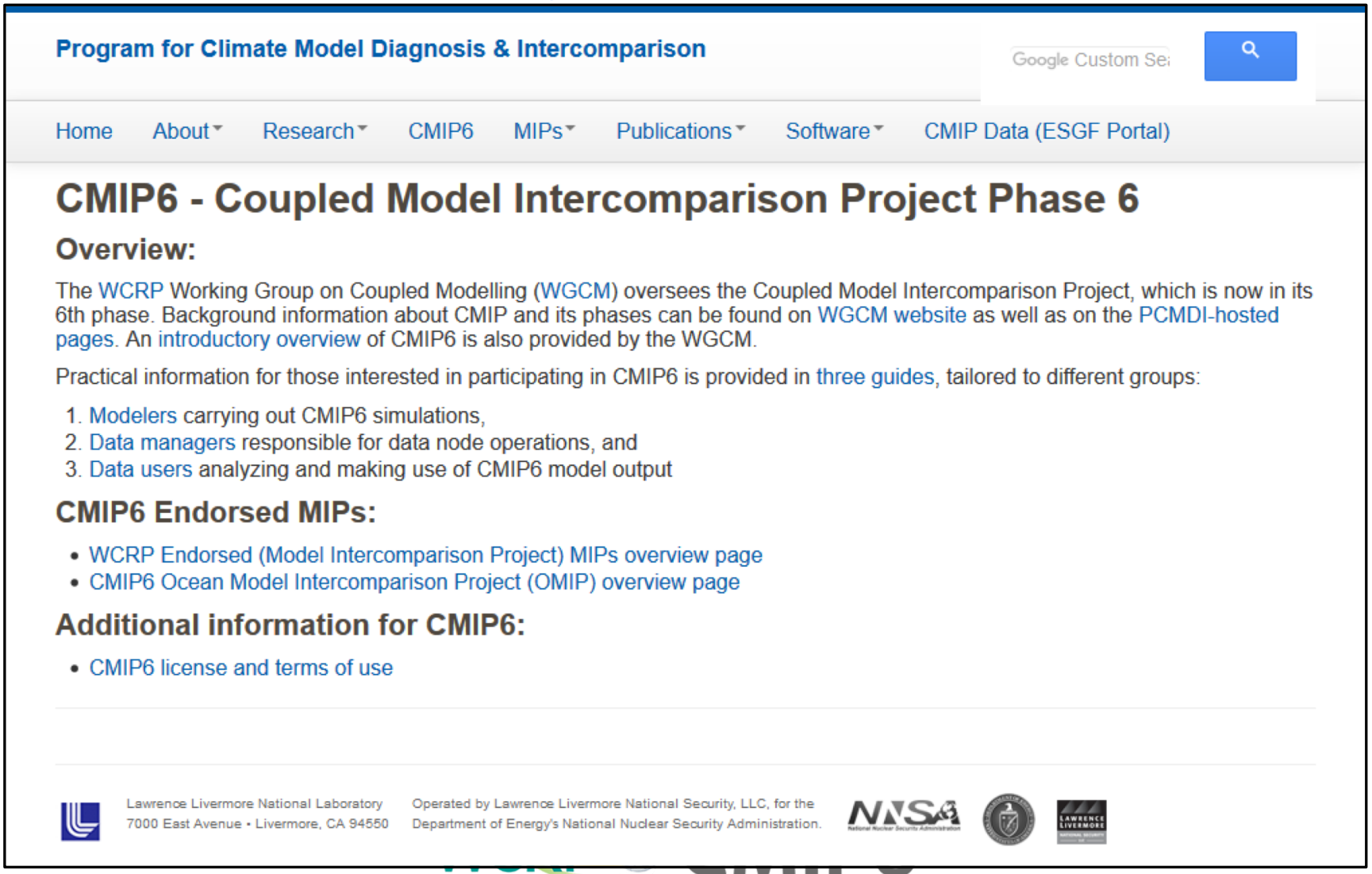
- Observations conforming to CMIP6 data specifications (P. Gleckler and D. Waliser)

  https://esgf-node.llnl.gov/projects/obs4mips/

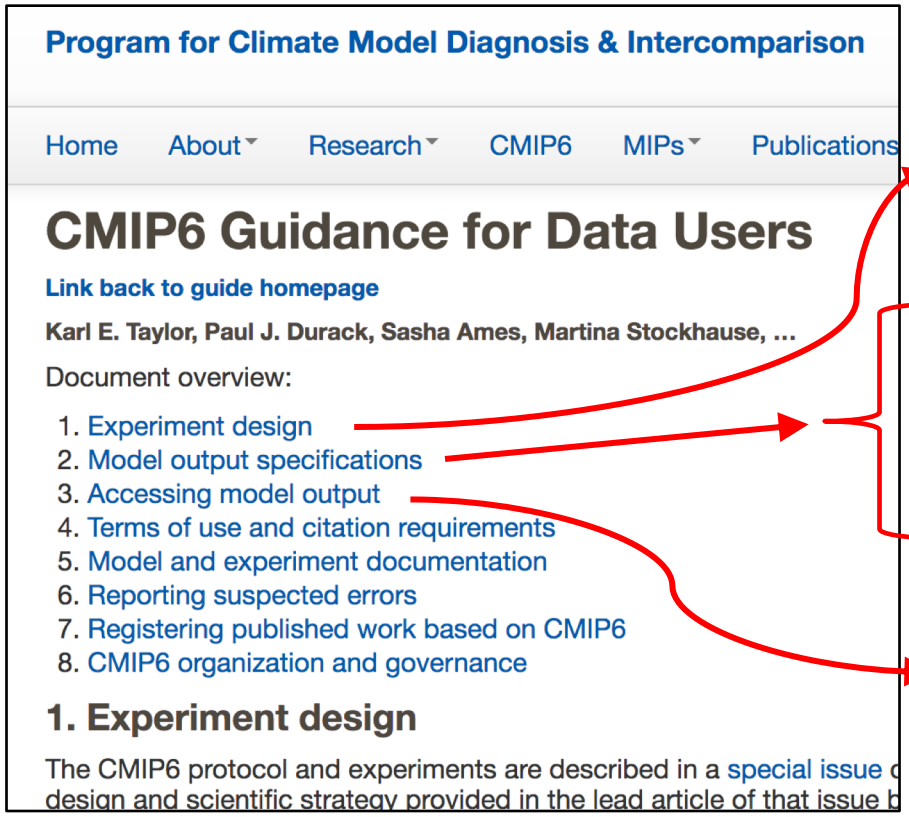
WCRP CMIP6
World Climate Research Programme

CMIP6 practical guides for contributors and users

CMIP6 Guide: https://pcmdi.llnl.gov/CMIP6/



Example: Guidance document for data users



- Points to information on experimental design
- Describes and links to controlled vocabularies
- Documents essential model output specifications
- Describes how to access output

https://pcmdi.llnl.gov/CMIP6/Guide/dataUsers.html

Please record you CMIP publications.

https://cmip-publications.llnl.gov/view/CMIP6/

Publication Hub Submit Edit View Search Login

Show Citations Show BibTex

Cmip5:

Year ▾	Author	Title
2017	Courtney TA, Lebrato M, et al.	Environmental controls on modern scleractinian coral and reef-scale calcification (Citation) (More Info) (Abstract) (BibTex)
2017	Quesada Benjamin, Arneth Almut, et al.	Atmospheric, radiative, and hydrologic effects of future land use and land cover changes: A global and multimodel climate picture (Citation) (More Info) (Abstract) (BibTex)
2017	J. Saynisch., J. Petereit, et al.	Impact of oceanic warming on electromagnetic oceanic tidal signals: A CMIP5 climate model-based sensitivity study (Citation) (More Info) (Abstract) (BibTex)
2017	Brady, Riley X., Alexander, Michael A., et al.	Emergent anthropogenic trends in California Current upwelling (Citation) (More Info) (Abstract) (BibTex)
2017	Esteban Abellán, Shayne McGregor, et al.	Analysis of the southward wind shift of ENSO in CMIP5 models (Citation) (More Info) (Abstract) (BibTex)
2017	Zazulie, N, Rusticucci, M, et al.	Regional climate of the subtropical central Andes using high-resolution CMIP5 models—part I: past performance (1980–2005) (Citation) (More Info) (Abstract) (BibTex)
2016	DeAngelis, Anthony M., Qu, Xin, et al.	Importance of vegetation processes for model spread in the fast precipitation response to CO2 forcing (Citation) (More Info) (Abstract) (BibTex)
2016	Le, Thanh	ENSO response to external forcing in CMIP5 simulations of the last millennium (Citation) (More Info) (Abstract) (BibTex)
2016	Belda, M, Holtanová, E, et al.	Global warming-induced changes in climate zones based on CMIP5 projections (Citation) (More Info) (Abstract) (BibTex)

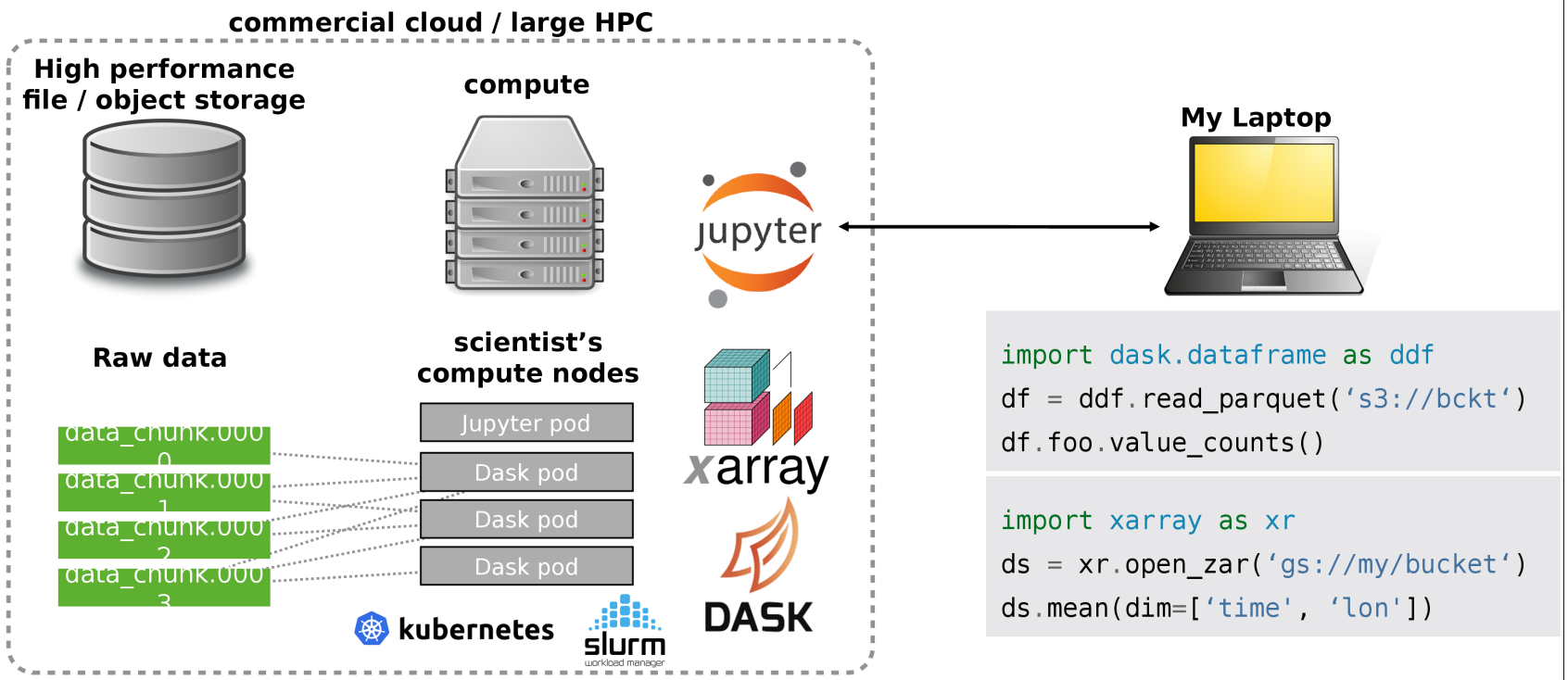
Publication Count: 1161

- Experiment
- Frequency
- Keyword
- Model
- Status
- Type
- Variable
- Year

Community-supported software components

PANGEO

Pangeo Environment



Try **https://tinyurl.com/pangeo-cmip6** right now on the cloud!
Figure courtesy Ryan Abernathey, Columbia.

WCRP CMIP6
World Climate Research Programme

Cloud considerations

- Commercial cloud vendors usually have a “Public Datasets” program for “publicly available high-value cloud-optimized datasets” for users seeking to “democratize access to data by making it available for analysis” (from Amazon Public Datasets webpage).
- For public datasets, typically ingress (upload to cloud) and storage are free, egress (download from cloud) is not. Academic users can apply for free cloud research credits.
- Caveats: programs can be ended at any time, have a time limit (e.g 2 years at AWS): negotiable.
- Many funding agencies evaluating the relative cost of moving to cloud vs purchasing or leasing on-premises.
- Major advantage is democratization: large data volumes available for analysis from anywhere in the world without replication.
- See discussion in *Science*, 8 February 2019: Government data, commercial cloud: Will public access suffer?

WCRP CMIP6
World Climate Research Programme

Computational and storage costs of CMIP6

- The CPMIP Project attempts to measure the computational and storage costs, and energy footprint, of CMIP6. Primavera doing the same.
- IS-ENES3 is taking a leading role in collecting this information, led by our hosts, the BSC team! (contact: Mario.Acosta@bsc.es).
- ES-DOC information request on models includes speed (SYPD), cost (CHSY), energy cost (JPSY), data intensity (GB/CH).
- First look (highly preliminary results!):

	Model Dev (kSY)	CMIP6 (kSY)	C Footprint (tons)
GFDL	53	(planned) 21	2750
IPSL	100	50	650

(1 transatlantic flight = 0.5T carbon per passenger).

Numbers courtesy Marie-Alice Foujols and Casimir Delavergne, IPSL, Alistair Adcroft and Aparna Radhakrishnan, GFDL. No implied endorsement of results by NOAA or IPSL.

A longer term view

- Despite growing international investment in climate modeling infrastructure, it remains fragile: single points of failure need to be addressed because they can lead to enormous disruptions
  - Some individuals are irreplaceable
  - Some software is not well documented
- ESGF has become essential to the climate research community:
  - CMIP, input4MIPs, obs4MIPs, etc.
  - Modeling and analysis groups have invested in it
- Given resource constraints, we should treat ESGF as part of an operational climate research enterprise; it must be reliable and robust
- Underlying data technologies are in flux: ESGF will adopt promising technical evolutions as they mature, and operationalize them for a broad community.

WCRP CMIP6
World Climate Research Programme