The PMIP4-CMIP6 Database: using standards to successfully share and use climate model data

Jean-Yves Peterschmitt[1] , Sébastien Denvil[2] , Guillaume Levavasseur[2] , Mark Greenslade[2] , Atef Ben Nasser[2]

1) LSCE-IPSL, L'Orme des Merisiers - Bat 701, 91191 Gif-sur-Yvette, France, Jean-Yves.Peterschmitt@lsce.ipsl.fr 2) IPSL, UPMC, Paris, France

Paleoclimate Modelling PMIP4 Intercomparison Project

The PMIP4-CMIP6 experiments…

…contribute to the CMIP6 question:
How does the Earth System respond to forcing?

Other PMIP4-CMIP6 experiments
- past1000
- lig127k
- midPliocene-eoi400

PMIP4-CMIP6 entry card
- CMIP6 entry card
- At least one of:
  - midHolocene
  - lgm

CMIP6 entry card
- *DECK*
- historical

***DECK* (CMIP entry card)**
- amip
- piControl
- 1pctCO2
- abrupt4xCO2

Other PMIP4 experiments (*non-CMIP6*)
- lgm sensitivity
- deglaciation
- quaternary interglacials
- Isotope modelling
- …

DECK = *Diagnosis, Evaluation, and Characterization of Klima Experiments*

PMIP4-CMIP6 data in The ESGF distributed DB

5+ *paleo* experiments X 16+ modelling groups
= Terabytes of model data

The PMIP4 participants !

Information as of April 13th 2016

	Institute	Country	0k *piControl*	LM *past1000* (1000 years)	6k *midHolocene*	21k *lgm*	Last Interglacial	Mid Pliocene warm period	Atm i_lon x j_lat x lev	Ocn i_lon x j_lat x lev	Model id
1	AWI	Germany	Yes	No	Yes	Yes	Yes	192x96 x L47	MPI-ESM-1.2 AWI-CM		
2	BCCR	Norway	Yes	Yes	Yes	Yes	Yes	1.9°x2.5° x L32	360x180 x L53	NorESM2-LM	
3	CAU-GEOMAR	Germany	Yes	No	Yes	No	Yes	No	96x48 x L19	182x149 x L31	KCM2
4	IAP	China	Yes	Yes	Yes	Yes	Yes	Yes	180x90 x L26	360x180 x L50	FGOALS3
5	INM	Russia	Yes	No	Yes	Yes	Yes	Yes	180x120 x L21	360x318 x L40	INMCM48
6	IPSL	France	Yes	Yes	Yes	Yes	Yes	Yes	144x142 x L79	144x142 x L79	IPSL-CM6-LR
7	MPI-M	Germany	Yes	Yes	Yes	Yes	No	No	192x96 x L47	256x220 x L40	MPI-ESM-1-2
8	MRI	Japan	Yes	Yes	Yes	Yes	Yes	Yes	320x160 x L80	360x364 x L61	MRI-ESM2
9	NASA GISS	USA	Yes	Yes	Yes	Yes	Yes	Yes	144x90 x L40 Cubed90 x L96	360x180 x L32 360x180 x L56	GISS-E2-1-R GISS-E3-R
10	NCAR	USA	Yes	Yes	Yes	Yes	Yes	Yes	288x192 x L32	320x384 x L60	CESM2
11	NUIST	China	Yes	Yes	Yes	Yes	Yes	Yes	192x96 x L47	362x292 x L46	NUIST-CSM
12	Stockholm University	Sweden	Yes	Yes	Yes	Yes	Yes	Yes	320x160 x L62	362x292 x L75	EC-EARTH3-LR EC-EARTH3-CC-LR
13	UK Met Office	UK	Yes	Yes	Yes	Yes	Yes	Yes	N96 = 192x145	ORCA1.0	UKESM1-hi and/or HadGEM3
14	University of Tokyo	Japan	Yes	Yes	Yes	Yes	Yes	Yes	2.8°x2.8° x L40 ?	0.5-1°x1° x L43	MIROC-ESM for CMIP6
15	University of Tasmania	Australia	Yes	Yes	Yes	No	Yes	No	64x56 x L18	128x112 x L31	CSIRO-Mk3L-1-3
16	VUAmsterdam	The Netherlands	Yes	No	Yes	Yes	Yes	No	64x32 x L3	122x65 x L20	iLOVECLIM1.2

- **How can this work ???**
  - **Use STANDARDS !**
- **What do YOU need ?**
  - **YOU tell us !**

What is the CMIP6 Database ?

- The *CMIP6 DB* is the climate model output DB that will be used for the IPCC 6th Assessment Report
- The *distributed* CMIP6 DB is a giant collection of model output data files
  - The expected size is… 10-50 Petabytes !
  - The files are located on the data servers (*data nodes* in/near the computing centers) of the ESGF (Earth System Grid Federation)
- We expect the first PMIP4-CMIP6 data to be available in early 2018
  - You can already use **PMIP3 data in the CMIP5 DB** !

Standard file names

- Naming conventions makes it easier to **reliably share and compare data**
- CMIPn DRS (*Data Reference Syntax*) document
  - All the CMIPn files have the same naming convention, metadata information, directory structure
  - File names: `<variable_id>_<table_id>_<experiment_id>_<source_id>_<member_id>_<grid_label>[_<time_range>].nc`
    - `tas_Amon_CCSM2-1_1pctCO2_r1i1p1f1_gn_202001-202912.nc`
  - The value of `experiment_id` has to be taken in a CV (*Controlled Vocabulary*): e.g. *piControl*, *past1000*, *midHolocene*, *lgm*, …
- PMIP4 and the CMIP6 DRS: https://pmip4.lsce.ipsl.fr/doku.php/database:drs

Standard variable names and metadata

- CMIPn variable names follow (and extend) the CF (*Climate and Forecast*) convention: http://cfconventions.org/
- The CF convention defines both variable names and the associated *metadata* (including the *axes* definition)
  - CMIP5 vars: http://cmip-pcmdi.llnl.gov/cmip5/docs/standard_output.xls
  - CMIP6 vars: https://earthsystemcog.org/projects/wip/CMIP6DataRequest
- Example: *tasmax*
  - Standard name = *air_temperature*
  - Long name = *Daily Maximum Near-Surface Air Temperature*
  - Units = *K*
  - Description = *maximum near-surface (usually, 2 meter) air temperature*

Standard file format

- CMIP data files are in *NetCDF* format: https://www.unidata.ucar.edu/software/netcdf/
  - Self-documented, binary format: **can be efficiently used by programs**
- One variable per file (vertical levels and multiple time steps are OK)
- Using the CMOR3 (*Climate Model Output Rewriter*) library makes it easier to comply with the standards: https://cmor.llnl.gov/

Standard documentation

The modelling groups have to comply with the standards mentioned in this poster, which makes the data partially self-documented. The rest of the documentation is handled by the ES-DOC project (*Earth System Documentation*): https://es-doc.org/

- ES-DOC can document: projects, experiments, models and simulations (i.e. experiments of a project run by a model)
- CMIP5 documentation example: https://documentation.es-doc.org/cmip5/
- Note: checking that the documentation is correct is part of the **CMIP quality control** !
- Added value: the model comparator!
  - https://compare.es-doc.org/

Searching and accessing CMIPn data

- CMIP data can be searched on the ESGF *Search nodes*
  - eg IPSL search node: https://esgf-node.ipsl.upmc.fr/
- Point and click Search
  - This allows you to discover the available data and select what you need. You can then download the data by:
    - Clicking on the download links
    - Generating a *wget* script that you can execute to download all the selected files
- Command line Search
  - You can use the synda tool: https://github.com/Prodiguer/synda/
    - synda can also be used to create a multi-user mirror of part of the CMIPn data!

Sample NetCDF file

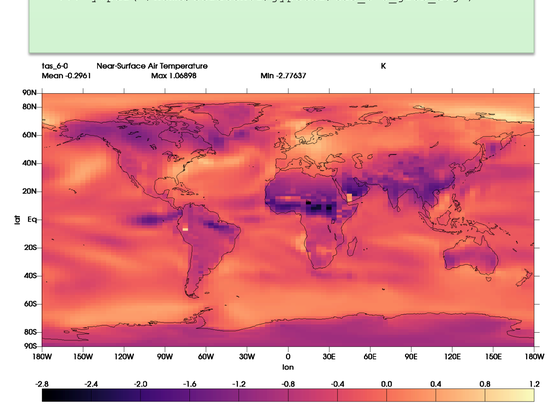
The following shows a sample NetCDF file. Note that this *climatological monthly mean* is a PMIP3 product created from PMIP3-CMIP5 data. PMIP4 is trying to get more climatological monthly mean data officially added to the CMIP6 DB

```
> ls -lh IPSL-CM5A-LR/midHolocene/monClim/Aclim/tas/r1i1p1/tas_Aclim_IPSL-CM5A-
LR_midHolocene_r1i1p1_230101-280012-clim.nc
-rw-r--r-- 1 jypeter clim 878K Mar  6  2014 IPSL-CM5A-
LR/midHolocene/monClim/Aclim/tas/r1i1p1/tas_Aclim_IPSL-CM5A-
LR_midHolocene_r1i1p1_230101-280012-clim.nc

> ncdump -h IPSL-CM5A-LR/midHolocene/monClim/Aclim/tas/r1i1p1/tas_Aclim_IPSL-CM5A-
LR_midHolocene_r1i1p1_230101-280012-clim.nc
netcdf tas_Aclim_IPSL-CM5A-LR_midHolocene_r1i1p1_230101-280012-clim {
dimensions:
        lon = 96 ;
        bnds = 2 ;
        lat = 96 ;
        time = UNLIMITED ; // (12 currently)
variables:
        double lon(lon) ;
                lon:standard_name = "longitude" ;
                lon:long_name = "longitude" ;
                lon:units = "degrees_east" ;
                lon:axis = "X" ;
                lon:bounds = "lon_bnds" ;
        double lon_bnds(lon, bnds) ;
[…]
        float tas(time, lat, lon) ;
                tas:standard_name = "air_temperature" ;
                tas:long_name = "Near-Surface Air Temperature" ;
                tas:units = "K" ;
                tas:_FillValue = 1.e+20f ;
                tas:original_name = "t2m";
[…]
// global attributes:
                :Conventions = "CF-1.4" ;
                :source = "IPSL-CM5A-LR (2010) : atmos : LMDZ4 (LMDZ4_v5, 96x95x39);
ocean : ORCA2 (NEMOV2_3, 2x2L31); seaIce : LIM2 (NEMOV2_3); ocnBgchem : PISCES
(NEMOV2_3); land : ORCHIDEE (orchidee_1_9_4_AR5)" ;
                :institution = "IPSL (Institut Pierre Simon Laplace, Paris, France)" ;
                :institute_id = "IPSL" ;
                :experiment_id = "midHolocene" ;
                :model_id = "IPSL-CM5A-LR" ;
[…]
                :contact = "pmip3db _at_ lsce.ipsl.fr Data manager : Jean-Yves
Peterschmitt" ;
[…]
                :experiment = "mid-Holocene" ;
                :frequency = "monClim" ;
                :creation_date = "2014-03-06T18:35:23Z" ;
                :project_id = "PMIP3" ;
[…]
}
```

Sample data usage

The following shows how to use the *UV-CDAT Python distribution* to use NetCDF data: https://uvcdat.llnl.gov/
(Get in touch with Jean-Yves if you need help about this)

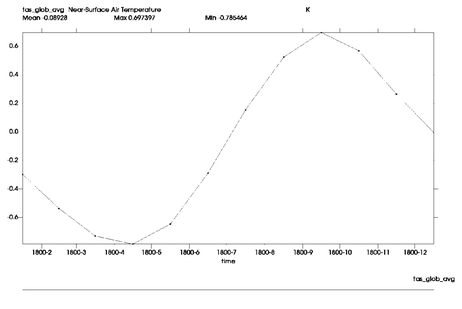
```
> python
>>> import cdms2, vcs
>>> f6k = cdms2.open('IPSL-CM5A-
LR/midHolocene/monClim/Aclim/tas/r1i1p1/tas_Aclim_IPSL-
CM5A-LR_midHolocene_r1i1p1_230101-280012-clim.nc')
>>> tas_6k = f6k('tas')
>>> f6k.close()
>>> f0k = cdms2.open('IPSL-CM5A-
LR/piControl/monClim/Aclim/tas/r1i1p1/tas_Aclim_IPSL-
CM5A-LR_piControl_r1i1p1_180001-279912-clim.nc')
>>> tas_0k = f0k('tas')
>>> f0k.close()
>>> tas_diff = tas_6k - tas_0k
>>> tas_diff.shape
(12, 96, 96)
>>> x = vcs.init()
>>> x.setcolormap('magma')
>>> tas_diff.id = 'tas_6-0'>>>
x.plot(tas_diff[0](longitude=(-180,180)))
<vcs.displayplot.Dp object at 0x2aba151b0168>
>>> x.pdf('/home/scratch01/jypeter/tas_6-0_jan')

>>> import cdutil
>>> tas_diff_avg = cdutil.averager(tas_diff, axis='xy',
weights = ['weighted', 'weighted'])
>>> tas_diff_avg.shape
(12,)
>>> tas_diff_avg.id = 'tas_glob_average'
>>> tas_diff_avg = cdms2.createVariable(tas_diff_avg,
axes=[tas_0k.getTime()])
>>> y.plot(tas_d_a)
>>> y.pdf('/home/scratch01/jypeter/tas_6-0_glob_avg')
```



PMIP references

- PMIP4 wiki: https://pmip4.lsce.ipsl.fr/
- All the PMIP phases: https://pmip.lsce.ipsl.fr/
- Main mailing list, where everybody using or providing PMIP data should be pmip-announce@lists.lsce.ipsl.fr

For questions about this poster, or joining the pmip-announce list, get in touch with Jean-Yves Peterschmitt: pmip4web@lsce.ipsl.fr



Scan this QR code
To download this poster ;-)

PMIP is endorsed by PAGES PAST GLOBAL CHANGES and WCRP World Climate Research Programme