

# CMIP6 Data Request Compilation Guidance Note

Date: 16 January 2015

Contact for questions: Martin Jukes (email: martin.jukes@stfc.ac.uk)

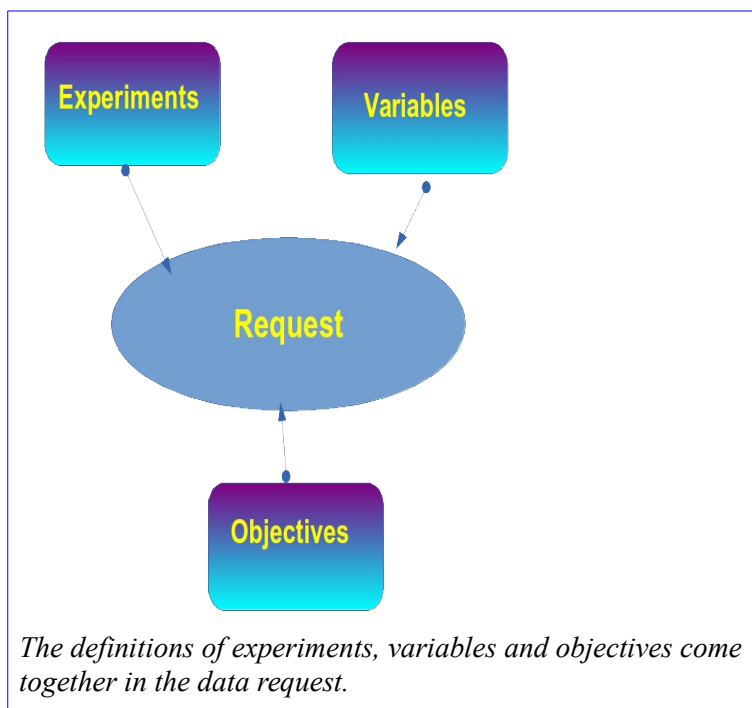
## Table of Contents

1 Introduction.....	1
2 Next steps and web resources.....	3
3 Details by section.....	4
3.1 Objectives.....	4
3.2 Experiments.....	4
3.3 Request Scoping.....	4
Data requested from experiments in other MIPs.....	5
3.4 New Variables – physical identity.....	5
3.5 New Variable Group.....	6
4 Reviewing and shortening CMIP5 MIP tables.....	7
5 Annex 1: Variable groups in CMIP5.....	7

## 1 Introduction

This document provides guidance on how the *CMIP6 Data Request Compilation Template* should be filled and how information on the CMIP5 variables that MIPs want to use again in CMIP6 will be collected.

This information will be used to compile a 1<sup>st</sup> draft of the *CMIP6 Data Request*. Once the final experiment list is submitted by the MIPs together with their final application end of March, a second phase will start where additional information (see details below) from the MIPs will be requested. A summary of the timeline towards the final CMIP6 Data Request is given in Box 1.



### Box 1: Timeline

- Online registration and CMIP5 variables which will be used again [see instructions in Section 2 and 4, respectively, of the *CMIP6 Data Request Compilation Guidance Note*] returned to WIP co-chairs and CMIP Panel Chair, c/o Martin Juckes (*MIP co-chairs*, 31 January 2015)
- Full information except the final experiment list (spreadsheet 2) for *CMIP6 Data Request Compilation Template* returned to WIP co-chairs and CMIP Panel Chair, c/o Martin Juckes (*MIP co-chairs*, 28 February 2015)
- Synthesized responses (i.e., 1<sup>st</sup> Draft of *CMIP6 Data Request*) sent to MIP co-chairs (copied to modelling groups) with request for additional information that can be aligned only after the final experiment list is received (*Martin Juckes, WIP co-chairs and CMIP Panel*, 15 April 2015)
- 1<sup>st</sup> Draft of *CMIP6 Data Request* reviewed and returned, with additional information as requested, to WIP co-chairs and CMIP Panel Chair, c/o Martin Juckes (*MIP co-chairs*, 15 June 2015)
- 1<sup>st</sup> Draft of *CMIP6 Data Request* reviewed and returned to WIP co-chairs and CMIP Panel Chair, c/o Martin Juckes (*Model groups*, 15 June 2015)
- Near-final version of *CMIP6 Data Request* published (*Martin Juckes, WIP co-chairs and CMIP Panel*, 31 July 2015)
- Feedback from modelling group on the actual implementation of the *CMIP6 Data Request* returned (*Model groups*, 15 October 2015)
- Final *CMIP6 Data Request* approved at WGCM meeting (*WIP co-chairs and CMIP Panel*, 31 October 2015)

The objective here is to collect information on the variables needed for the evaluation and analysis of the CMIP DECK experiments, the CMIP6 Historical Simulation, and additional experiments from CMIP6-Endorsed MIPs (see further information on these terms at the CMIP Panel website at <http://www.wcrp-climate.org/index.php/wgcm-cmip/about-cmip>).

The process will be slightly modified relative to CMIP5 in order to deal with the additional complexity of multiple simultaneous MIPs, to cope with the expected increase in number of experiments and variables, to manage the additional details about each variable which are needed, and to develop a final data request which is more amenable to automated processing.

Where new variables are required, they should be specified in the accompanying template. Where variables used in CMIP5, CCMI, SPECS or CORDEX are required, there are two choices: (1) include a reference to an entire table or group of variables (see below) or (2) create a new table with a selection of variables from tables used in CMIP5, CCMI, SPECS or CORDEX (further details below).

CMIP6-Endorsed MIPs are asked to provide contact information in an online form (see section 2) and technical details in an excel workbook with the following sections:

- **Objectives:** a set of scientific and dissemination objectives. Referred to from the variable group sheet.
- **Experiments:** This sheet collects information about experiments from the MIP proposals submitted to the CMIP Panel.
- **Request scoping:** this sheet provides the link between Objectives, Experiments and Variables. Each row in the data entry section will defined the experiments for which a set of variables is needed and the objectives served by this request.
- **New Variables:** If you only need variables used in CMIP5, CORDEX, CCMI, or SPECS this sheet can be left blank. Variables which have not previously been used in in one of these MIPs should be defined here.
- **New Variable Groups:** In order to organise the request, groups of variables should be designated aimed at a common set of objectives, with common realm, common frequency and common duration (number of model years for which data is requested). If you are only using variables used in CMIP5,

CORDEX, CCM1, or SPECS this sheet can be left blank. Note, however, that standard names will be required for all CMIP6 variables and these were not assigned for some variables used in CCM1.

Further details will be required in a second phase (starting mid-April and due mid June), including:

- Full details of dimensions to be used, including details of vertical levels, transects, selected grid points etc.
- Details of the time periods for which data is required;
- Valid ranges of variables;
- Where variables are only needed for models which represent target processes (e.g. interactive carbon cycle or atmospheric chemistry);
- Where variables are only needed for models which include certain diagnostics tools interactively (e.g. COSP simulator has been installed);
- Whether this variable is of interest to downstream users (such as impacts researchers, WG2 users) or whether its principal purpose is for understanding and analysis of the climate system itself.

For some MIPs the process of generating the variable request may be quite simple: e.g. identifying a set of variables defined in the CMIP5 variable request, and requesting that they be output for all years of all the requested experiments. Other MIPs may need a range of new variables and different output for different experiments. In order to provide a simple process where appropriate and a robust process for inclusion of new variables in a structured way (bearing in mind that the task of delivering the data for a significant number of MIPs will be challenging for the modelling centres) we have a dual approach allowing review of CMIP5 request and proposals of extensions.

**Box 2: character restrictions for short names**

“Short Names”, which act as identifiers for experiments, experiment groups, etc, should use the restricted character set a-z, A-Z, 0-9 or “-”.

## 2 Next steps and web resources

**The template should be filled in and returned (to WIP co-chairs and CMIP Panel Chair, c/o [martin.juckes@stfc.ac.uk](mailto:martin.juckes@stfc.ac.uk)) by February 28, 2015.**

**Information on CMIP5 variables to be used (see section 4) should be returned to [martin.juckes@stfc.ac.uk](mailto:martin.juckes@stfc.ac.uk) by January 31, 2015.**

In order to get the process started please complete the following three registration steps by January 31, 2015:

1. In order to facilitate the process of the developing the request, a data request consolidation web site has been set up at <https://www.earthsystemcog.org/projects/wipdatarequest/>. In order to gain access, you will firstly need to obtain an ESGF user account (<https://pcmdi9.llnl.gov/esgf-web-fe/createAccount> or <https://services.ceda.ac.uk/cedasite/register/start/> ). If you have an account, please provide the associated ESGF OpenID it in the contact information form (see next paragraph), otherwise supply when available. The site will only be open to registered users so that it can be used to share documents which have not been finalised – representatives of modelling centres will also be invited to register. When documents are ready for release they will be moved to another site.
2. Fill out the form at <http://goo.gl/forms/U7lv2ING4H> giving contact details of the project scientific lead and a technical lead for the data specification. There are also 4 questions designed to give us an idea of the level of complexity we can expect in the data requested by your MIP (answers should be indicative – please answer “not known” rather than expend effort trying to gather information quickly or delaying submission of contact details);
3. If you would like to take part in a telco in January to discuss the data request, fill out the doodle poll here: <http://doodle.com/daste93aqqq3aziq>

Public documents produced by the WGCM Infrastructure Panel (such as this document and the associated template) will be available here: <https://www.earthsystemcog.org/projects/wip/CMIP6DataRequest> .

Please use the “checkdreq.py” script (see link given in previous paragraph for details) to check the consistency of your data request spreadsheet, and let us know if there are any issues.

You will be sent instructions later on how to set up your own CoG site.

## 3 Details by section

### 3.1 Objectives

The “Objectives” section, which is new compared to CMIP5, is intended to support the consolidation activity and to maintain some transparency so that at the end of this process, when a finalised data request is delivered to modelling centres, the individuals receiving the request will have information not only about what is being requested, but why it is being requested. These objectives should, of course, be consistent with the overall objectives of CMIP6, but in finer detail:

- **Short name of objective:** a label (restricted character set);
- **Long name:** a title;
- **Description:** a paragraph specifying the scientific objective. Be specific as possible: these objectives will be linked to output requests in the “Request Scoping” sheet described below and may be used by modelling centres to prioritise resources. If applicable, specify support for downstream users as a separate objective or objective(s).

### 3.2 Experiments

This sheet contains information extracted from the MIP proposals submitted to WGCM. Please review the entries relevant to your MIP for completeness and correctness: where you make changes please highlight in red.

- **Experiment short name:** a label (with restrictive character set); If you wish to assign a more recognisable name please edit this column, but make sure that you use only the permitted characters (see box 2 above).
- **Experiment group:** a group label: entries in the “Request Scoping” section will refer to groups of experiments, rather than individual experiments, in order to reduce the need for duplication of entries. The experiment groups suggested in the template have been defined in such a way that the tier and the entries in columns Q to Z are the same for all experiments in each group. If the output requirements are not the same for all experiments in these suggested groups, please create additional groups by editing this column.
- **Experiment number;**
- **MIP short name;**
- **Experiment name:** from application;
- **Tier:** tier of the experiment, as defined in MIP proposal;
- **Number of start dates:** this entry is to provide information about start date ensembles (e.g. decadal experiments). It should be set to “1” if there is only one start date;

The remaining columns are intended to be self explanatory – please review these to ensure that information which has been taken from your MIP proposal has been entered accurately. Please highlight any rows that you alter in red. If you insert additional rows for further experiments, please highlight in blue.

### 3.3 Request Scoping

This sheet joins together the other elements of the document into a draft request, providing information about the scale of the data request and the links to the objectives.

Each row identifies a set of variables required from a set of experiments.

The first 3 columns identify a set of variables – the first identifies a variable group and the 2<sup>nd</sup> and 3<sup>rd</sup> may be used to identify a smaller set of variables within that group.

- **Short name of variable group:** specifies a group of variables: it could be one of the CMIP5 groups (see appendix 1), one of the SPECS, CORDEX or CCMi tables, a new group defined using the tem-

plate described in section 3.5 below, or a new short name for a CMIP5 table which has been modified as described in section 4 below.

- **Selection method:** the method used to identify variables from within the group: must be one of:
  - all: all variables will be used;
  - shape: a shape will be specified as a combination of XYZTK (e.g. XYT or XYZT) where X,Y,Z,T,K indicate the presence of longitude, latitude, vertical, time and other dimensions respectively.
  - priority: one or more integers specifying the priority of the variables to be included (e.g. there could be one request row for priority 1 and a separate row for priority 2 and 3);
  - list: select a list of variables from a CMIP5 table. Create a modified copy of the CMIP5 data request as described in section 4 below. Place the name of the relevant CMIP5 table in the “selection option” column.
- **selection option:** one of the options corresponding to the selection method.

The next 4 columns are:

- **grid:** this is a new option: if appropriate the data may be requested on a specified grid. The grid should be one of 200km, 100km, 50km, 25km, 10km. If, as is likely, different MIPs request the same data at different resolutions the WIP will seek to reach a solution which avoids calculation and distribution on multiple grids;
- **Sensitive:** [yes/no] set this to “yes” if there is information in the variables of this group which would be compromised by regridding the data;
- **Comment:** a comment may be entered here, primarily for your own convenience;
- **Objectives served:** a comma separated list of the objectives supported by this request line (using the short names from column A of sheet “Objectives”). Please do not request data unless you have specific and reasonable plans to exploit and analyse the data.
- Columns H to Q collect information about data needed from the DECK experiments. Any data which is needed should be requested here – please do not assume that it will be archived just because it was archived last time. For each experiment there is a column for the number of years and a column for the number of ensemble members. Enter “all” if you want data for all the years in all the simulations requested.
- Columns R onwards collect information about data needed from your MIP. There are two columns for each experiment group. Enter the experiment group names in row 4, and years and ensemble sizes in rows 6 onwards. Add more column pairs if needed.

### **Data requested from experiments in other MIPs**

- After entering the information for your own MIP, please use additional space to indicate any requests for data from experiments proposed by other for each MIP enter the MIP name in row 4 and fill out 3 columns: the number of experiments, number of years and number of ensemble members.

### ***3.4 New Variables – physical identity***

This table presents physical entities which have not previously been defined in the MIP system. This is collected as a single list here to aid consolidation of new variable definitions across range of participating MIPs. We do not, at this stage, want the full details of the dimensions and cell\_methods, but enough details to understand the scope of the request and whether there are overlaps between MIPs.

- **Short name:** label (character restriction);
- **Standard\_name:** a name from the CF Standard Name table or proposed for inclusion in the table;
- **Standard name status:** 1: in the current list; 0: proposed to CF and under discussion or accepted pending list update; -1: to do. MIP coordinators are responsible for ensuring that standard names are registered for the variables they request. The WIP will provide advice if needed;
- **Native grid:** Specify which model grid this variable is expected to be on: ocean, atmosphere, land, ice, or other.
- **Units:** the units in which the variable should be reported;

- **Long name**; a descriptive name, following the pattern of existing MIP long names.
- **Description**; a simple description of the variable.
- **Priority**:
  1. modelling centres must commit to supplying all priority 1 variables associated with at least one science question for Tier 1 experiments of any MIP which they participate in;
  2. expected to be used in multi-model diagnostics; models not supplying these variables may be omitted from some parts of the inter-comparison;
  3. experimental -- used for exploring new capabilities and/or unlikely to be used in multi-model diagnostic.
- **Observation dataset** (if applicable): a registry of observational datasets is being assembled here: [https://docs.google.com/spreadsheets/d/1mJGJ\\_7nfyY8oLC8-XZqL3DtWUM\\_tF6VuOuyOPsJO-c/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1mJGJ_7nfyY8oLC8-XZqL3DtWUM_tF6VuOuyOPsJO-c/edit?usp=sharing). Please enter the short name of an observational dataset from the registry in column G. If the dataset you are interested in is not in the registry, please use this form to create an entry: <http://goo.gl/forms/LBN9yninxS> , providing a URL and citation for each dataset.
- Different shape (e.g. height dependent or not) and sub-grid scale representations (masking, point vs. time mean) should be specified in the variable group sheet(s), allowing each basic variable definition to be used more than once.
- WGCM has specified that each new variable should be linked to specific objective(s): these objectives should be listed in the “Objectives” sheet and associated with variables through the entries in the “Request Scoping” sheet. E.g. if you have a collection of new variables which are required at daily frequency throughout a set of experiments for some science goals and at 3 hourly intervals for specific time slices for downstream users:
  - (1) enter the basic definitions of the variables here;
  - (2) create one or two “Variable Group” sheets, specifying additional details about how the variables should be archived;
  - (3) Specify the different objectives to be served by different data collections in the “Objectives” sheet;
  - (4) Enter two rows in the data request sheet, one for the science objectives and one for the downstream users.

### 3.5 New Variable Group

This sheet should be copied for each new group that is defined.

There are two rows to provide information about the variable group, and then a row for each variable to be defined. Variables may be physical entities defined in previous MIPs, but here used with different frequency, masking, or shape, or new physical entities introduced for the purpose of this MIP.

The CMIP5 data request was organised into tables and groups within tables. At this point, we are asking for definitions of new variable groups, if needed (note that if you only need to use variables previously used in other MIPs, you can enter the information directly in the “Request Scoping” sheet as described in section 5.4 above). When defining groups of new variables here, please remember that the data request on the request scoping sheet will be the same for all variables in the group: where different requests are intended for different variables separate groups should be created.

Create a page for each group, listing variables.

The following items need to be specified for each variable:

1. **Short name of group**: a label (with character description)
2. **Short name**: variable short name
3. **Table**: enter “new” if this is a new physical entity, or name of previous table (e.g. if it occurred in CMIP5\_Amon and is being included here in a daily group); If a variable short name occurs twice in the “New variables” table, enter the row number here as well (e.g. “new 24”) [this should be rare, but may happen, for instance, if a quantity is required on both atmosphere and ocean grids and the name is reused].
4. **Frequency**: a frequency (one of fx, yr, monClim, mon, day, 6hr, 3hr, subhr – if additional frequencies are needed please provide details on an additional sheet);

5. **Description (optional)**: enter text here if the original description needs modification (e.g. to describe masking).
6. **Shape**: the shape entry specifies the spatial dimensions associated with the variables in the group (though without requiring full details at this stage). Some combination of XYZTK -- using K for sites or transects or any collection of points;
7. **Number of levels** (blank, integer or all): specify "all" if needed on all model levels. Specify an integer if the variable is to be requested on a particular number of levels is to be requested (e.g. 8 pressure levels). If left blank, a single level field will be assumed;
8. **Time mean, point or climatology**: specify whether the data should be time means, instantaneous values, or climatologies (this information will be used in comparing requests of different MIPs);
9. **Masking**: state whether a mask is to be applied (e.g. land, sea, ice).

## 4 Reviewing and shortening CMIP5 MIP tables

If you want to comment on variables defined in the CMIP5 data request (e.g. to suggest clarifications of long names or comments) or create a shorter version (by deleting variables) of one of the CMIP5 MIP tables for your request, please follow the steps described below. Note that if you want to request, from a CMIP5 table, all variables of a particular shape or priority there is no need to create a new copy: you can just follow the steps described in section 3.4 to define the variables which you are requesting from the existing tables.

To comment on variables in the CMIP5 data request:

1. take a copy of [http://cmip-pcmdi.llnl.gov/cmip5/docs/standard\\_output.xls](http://cmip-pcmdi.llnl.gov/cmip5/docs/standard_output.xls) and rename <your\_MIP>\_standard\_output.xls
2. place your comments in column Z;

To create a short version of CMIP5 tables:

1. take a copy as above (or use the same file as for comments, if applicable);
2. enter 0/1/2 in column Y: 0: not wanted (you may also deleted variables you do not want); 1: wanted at the same priority as given in the CMIP5 table; 2: wanted at the highest priority.

Finally, upload your modified table to the WIP Data Request pages (<https://www.earthsystemcog.org/projects/wipdatarequest/> ).

## 5 Annex 1: Variable groups in CMIP5

Group name	MIP Table	Selection within table (all elements in one group if blank).
Oclim	Oclim	
Oyr	Oyr	
Amon	Amon	
Omon-3d	Omon	lat*lon*olev
Omon-oth		Other
Lmon	Lmon	
Llmon	Llmon	
Olmon	Olmon	
Aero-oth	aero	lat*lon

Aero-3d		lat*lon*alev
Day-ss	day	Subset specified as for requested fields only*
Day-oth		Others (huss, tasmin, tasmax, tas, pr, psl, sfcWind, tossq, tos, omldamax)
6hrLev	6hrLev	
6hrPlev	6hrPlev	
3hr	3hr	
CfMon-3dstd	cfMon	3D (normal radiative calc).
CfMon-2dmod		2D (diagnosed with modified)
CfMon-3dmod		3D (diagnosed with modified)
CfMon-sim		Inline (simulators)
cfOff	cfOff	
CfDay-2d	cfDay	2D
CfDay-3d		3D
Cf3hr-grid	cf3hr	Model grid
Cf3hr-sim		Inline (simulators)
cfSites	cfSites	

\* In the CMIP5\_day table, variables defined in row 34 onwards were only requested for pre-industrial controls, historical, RCPs and ESM-RCPs, AMIP & the 2030 time-slice run.