

# CMIP6 Global Attributes, DRS, Filenames, Directory Structure, and CV's

10 September 2018 (v6.2.7)

Document short URL: <https://goo.gl/v1drZl>

Karl E. Taylor, Martin Juckes, V. Balaji, Luca Cinquini, Sébastien Denvil, Paul J. Durack, Mark Elkington, Eric Guilyardi, Slava Kharin, Michael Lautenschlager, Bryan Lawrence, Denis Nadeau, and Martina Stockhause

## Executive Summary

In files containing CMIP model-simulation output, global attributes are used to describe the source of the data, the imposed experiment conditions, the contents of the file, licensing restrictions, and other information useful to those analyzing the data. Here we define the global attributes that should appear in CMIP6 files (some are required, others optional), along with the so-called CMIP6 “data reference syntax” (DRS). The subset of global attributes that defines the DRS is used in constructing the directory structure and file names found in the CMIP6 archive, and also to construct URL's leading to further information about the simulations and in populating search facets.

## Introduction

As in earlier phases of CMIP, a well-defined set of global attributes will be recorded in each CMIP6 model output file, providing information necessary for interpreting the data. Table 1 contains the list of CMIP6 global attributes and indicates which ones are required and which are optional. The values for many of the global attributes must be drawn from special CMIP6 “controlled vocabularies” (CVs). A CV, in simplest form, is a list of the permitted values that can be assigned to a given global attribute. Some of these lists of permitted values appear in this document, but they should not be relied on to be 100% correct. Rather, consult the reference CVs for CMIP6, which are available at [https://github.com/WCRP-CMIP/CMIP6\\_CVs/](https://github.com/WCRP-CMIP/CMIP6_CVs/)

A subset of the global attributes, which comprise the data reference syntax (DRS) for CMIP6, are described following the Table 1 notes, and templates for the CMIP6 filenames and directory structures are also defined. Table 3 provides a summary of the quality assurance checks that should be performed on CMIP6 files.

[Appendix 1](#) describes the rationale underlying the global attributes used to label CMIP6 experiments, and provides additional details. [Appendix 2](#) defines the algorithm used to define the “nominal\_resolution” attribute. [Appendix 3](#) contains a revision history of this document.

Documents of related interest may be found at: <https://www.earthsystemcog.org/projects/wip/>

**Table 1:** CMIP6 global attribute description and comparison with CMIP5 (see [CMIP5 output metadata requirements.pdf](#) and [cmip5 data reference syntax.pdf](#)).

Table key: **name or form has been changed (relative to CMIP5)** **new attribute for CMIP6**

CMIP6 global attribute see note 1	description	examples	corresponding attribute in CMIP5	form see note 2	when required?	further information and rationale
<b>activity_id</b>	activity identifier(s) (part of DRS)	“CMIP”, “PMIP”, “LS3MIP LUMIP” (see note 3 and <a href="#">reference CV</a> )	project_id	CV	always	renamed more generically, since not all activities are projects; also multiple activities may now be listed separated by single spaces. Used in faceted searches.
branch_method	branching procedure	“standard”, “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.
<b>branch_time_in_parent</b>	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 (preferred) or set to 0.0D0.
comment	see note 6	see note 6	comment	free form	never	no change from CMIP5; CF-convention standard
contact	see note 6	see note 6	contact	free form	never	required in CMIP5, but not now because information should be available via further_info_url

Conventions	convention version	"CF-1.7 CMIP-6.2" or "CF-1.7 CMIP-6.2 UGRID-1.0"	Conventions	CV	always	updated version from CMIP5 with a list of conventions separated by single spaces now allowed. The "examples" shows the only options. Note that CMIP-6.2 reflects the version of the present document (without indicating the minor changes), and would have to be updated if the present document were incremented from 6.2.x to 6.3.0.
creation_date	date file was created	see note 7	creation_date	structured form	always	no change from CMIP5
data_specs_version	version identifier	01.00.00, 01.00.01, ... 01.00.xx (see <a href="#">data request</a> for latest release)	-	CV	always	records the version (or "release") number of the <a href="#">data request</a> , which is relied on for certain global and variable attributes.
experiment	short experiment description	"pre-industrial control", "abrupt quadrupling of CO2" (see <a href="#">reference CV</a> )	experiment	CV	always	no change from CMIP5
experiment_id	root experiment identifier (part of DRS)	"historical", "abrupt4xCO2" (see <a href="#">Appendix 1</a> and <a href="#">reference CV</a> )	experiment_id	CV	always	similar to CMIP5. Used in faceted searches.
external_variables	external cell measures	"areacella", "areacello"	-	CV	whenever appropriate	list of cell measure variables (separated by single spaces) that are referenced but not included in the file. These variables will be stored independently in the CMIP data archive.
forcing_index	index for variant of forcing	1, 2, 82, 323	-	integer >0 (see note 8)	always	distinguishes variants that differ in forcing
frequency	sampling frequency	"mon", "day", "6hr" (see <a href="#">reference CV</a> )	frequency	CV	always	no change from CMIP5; value must be consistent with the data

	(part of DRS)					request. Used in faceted searches.
further_info_url	location of documentation	see note 9	-	CV	always	points to definitive documentation that can be updated (even after data files have been written)
grid	grid	see note 10	-	free form	always	briefly describes output grid characteristics
grid_label	grid identifier (part of DRS)	“gn”, “gr”, “gr1”, “gr2”, “grz”, “gm” (see <a href="#">reference CV</a> )	-	CV (see note 11)	always	used in file name to distinguish among files when the variable is reported on more than one grid. Used in faceted searches.
history	see note 6	see note 6	history	free form	never	no change; CF-convention standard
initialization_index	Index for variant of initialization method	1	initialization_method	integer >0 (see note 8)	always	name changed to increase consistency; 0 is forbidden in CMIP6
institution	institution name	“Meteorological Research Institute” (see <a href="#">reference CV</a> )	institution	registered content	always	no change from CMIP5
institution_id	institution identifier (part of DRS)	“IPSL” (see <a href="#">reference CV</a> )	institute_id	registered content	always	Was “institute_id”; name changed to parallel other global attributes; this string is constructed only using the character set: a-z, A-Z, 0-9, and ‘-’ (i.e., hyphen). Used in faceted searches.
license	license restrictions	see note 12	-	some required text	always	ensures that anyone using the files has access to the terms of use
mip_era	activity’s associated CMIP cycle (part of DRS)	“CMIP5”, “CMIP6” (see note 3 and <a href="#">reference CV</a> )	-	CV	always	added to fully define activity and to enable one to determine what cycle of CMIP dictates experiment and data specifications. This means that mip_era = “CMIP6” for all CMIP6 output no matter how

						old the model is that produced it. Used in faceted searches.
nominal_resolution	approximate horizontal resolution	"50 km", "100 km", "250 km", "1x1 degree". (See <a href="#">Appendix 2</a> )	-	CV	always	Added in CMIP6 to provide an indication of approximate output grid resolution. See <a href="#">Appendix 2</a> . Used in faceted searches.
parent_activity_id	parent activity identifier	"CMIP", ScenarioMIP (see <a href="#">reference CV</a> )	-	CV	whenever parent exists	to help identify parent run (when parent comes from a different MIP); when no parent, omit or set to "no parent"
parent_experiment_id	parent experiment identifier	"piControl" (see <a href="#">reference CV</a> )	parent_experiment_id	CV	whenever parent exists	If the parent run is immaterial or if an experiment is initialized in some other way (e.g., from observations), then omit or set to "no parent"
parent_mip_era	parent's associated MIP cycle	"CMIP5", "CMIP6" (see <a href="#">reference CV</a> )	-	CV	whenever parent exists	with parent_activity_id, this fully defines parent activity; when no parent, omit or set to "no parent"
parent_source_id	parent model identifier	"CanCM4" (see <a href="#">reference CV</a> )	-	registered content	whenever parent exists	helps identify parent run; when no parent, omit or set to "no parent"; usually will be the same as source_id
parent_time_units	time units used in parent	"days since 1850-1-1", "days since 1000-1-1 (noleap)" (see note 5)	-	structured form	whenever parent exists	without this, a user would have to read the parent file to interpret branch_time_in_parent; the "calendar" modifier is required only if the calendars used by the parent and child differ. When no parent, omit or set to "no parent".
<a href="#">parent_variant_label</a> (modified form)	parent variant label	"r1i1p1f1", "r1i2p223f3", "no parent"	parent_experiment_rip	see Appendix 1	whenever parent exists	used to distinguish among variants of the parent; when no parent, omit or set to "no parent"
<a href="#">physics_index</a>	index for model physics variant	3	physics_version	integer >0 (see note 8)	always	name changed to increase consistency; 0 is forbidden

product	product type (part of DRS)	“model-output”	product	-	always	As in CMIP5, but “model-output” is now the only option. Used in faceted searches.
<b>realization_index</b>	realization number	5	realization	integer >0 (see note 8)	always	name changed to increase consistency; 0 is forbidden
<b>realm</b>	realm(s) where variable is defined (part of DRS)	“atmos”, “ocean”, “atmosChem atmos” (see <a href="#">reference CV</a> )	modeling_realm	CV	always	name changed to make appropriate for observational data; multiple realms may be listed (separated by single spaces) when appropriate. Value(s) must be consistent with the data request. Used in faceted searches.
references	see note 6	see note 6	references	free form	never	no change; CF-convention standard
<b>source</b> (modified form)	full model name/version	see note 13 and <a href="#">reference CV</a>	source	registered content	always	somewhat different form from CMIP5. The first part (before the colon) should be an abbreviated identifier that is the basis of source_id
<b>source_id</b> (modified form)	model identifier (part of DRS)	“GFDL-CM2-1” (see note 13 and <a href="#">reference CV</a> )	model_id	registered content (limited in length to no more than 16 characters)	always	edited version of first part of “source” (with forbidden characters like spaces and periods replaced with hyphens); used in constructing the file name. Used in faceted searches.
source_type	model configuration	“AGCM”, “OGCM”, “AOGCM”, “ISM”, “AOGCM ISM” (see <a href="#">reference CV</a> )	-	CV see note 14	always	added partly because obs4MIPs defines this (e.g., “in-situ”); This should describe the model most directly responsible for the output (e.g., for dynamical downscaling output, it would describe the regional model, not the global model responsible for driving the regional model). Sometimes it is

						appropriate to list two (or more) model types here. Used in faceted searches.
sub_experiment	description of sub-experiment	see <a href="#">Appendix 1</a> and <a href="#">reference CV</a>	-	CV	always	needed for CMIP6 hindcast and forecast experiments. For other experiments, this should be set to "none".
sub_experiment_id	sub-experiment identifier (part of DRS "member_id")	"s1960", "s1965", "none" (see <a href="#">reference CV</a> )	-	CV	always	needed for CMIP6 hindcast and forecast experiments to indicate "start year". For other experiments, this should be set to "none". Used in faceted searches.
<b>table_id</b>	table identifier (part of DRS)	"Amon", "Oday" (see <a href="#">reference CV</a> )	table_id	CV	always	the word "Table" is no longer included in this attribute (e.g., "Amon", not "Table Amon") and there should be no additional information (such as table date and check sum) included. Used in faceted searches.
title	see note 6	see note 6	title	free form	never	no change; CF-convention standard
<b>tracking_id</b> (modified form)	unique file identifier	see note 15	tracking_id	structured form with some CV	always	form modified to facilitate it use by ESGF.
variable_id	variable identifier (part of DRS)	"tas", "pr", "ua" (see <a href="#">data request</a> )	-	CV	always	added to direct users and software to the primary variable of interest in the file. Used in faceted searches.
variant_info	description of run variant	"forcing: black carbon aerosol only"	-	free form	never, but recommended see note 16	provides brief descriptions of variant differences
variant_label	"variant" label (part of DRS "member_id")	"r1i1p1f1", "f1i2p223f3"	-	see Appendix 1	always	used in faceted searches

-	see note 17	-	forcing	-	not used in CMIP6	no longer needed because this information is now recorded in variant_info
---	-------------	---	---------	---	-------------------	---

#### Table Notes:

1. If CMOR is used to write output files, an additional global attribute will be included: `cmor_version`.
2. “CV” means content must be taken from a “controlled vocabulary” defined by the WIP. “registered content” is special controlled vocabulary defined by each modeling group and approved by the WIP. The reference CV’s are hosted [here](#).
3. The `project_id` used in CMIP5 is being replaced in CMIP6 with two global attributes: 1) a [mip\\_era](#) (a label indicating the CMIP cycle responsible for the experiment and data request), which for CMIP6 must invariably be set to “CMIP6”, and 2) an [activity\\_id](#) identifying the responsible “MIP”. In a few cases it may be appropriate to include multiple activities in the `activity_id` (separated by single spaces). An example of this is “LUMIP AerChemMIP” for one of the CMIP6 land-use change experiments.
4. `branch_method` examples: “perturbations to atmospheric fields were applied at the branch time, followed by a 50 year spin-up period under control conditions”; “fixed historical forcing from year 1850 was applied for a 20 year spin-up period, after which the historically-evolving forcing was imposed.” For a run that had not been spawned from a parent, this attribute should be omitted or set to “no parent”. If external documentation describing the branch method can be retrieved using the url recorded in the `further_info_url` global attribute, then `branch_method` may be set to “see further\_info\_url”, but it is better to provide a brief description as in the examples above. For runs simply restarted from some point in the parent (without special procedures applied), `branch_method` should be set to “standard”.
5. The three attributes related to branching time (`branch_time_in_child`, `branch_time_in_parent`, `parent_time_units`) must be included if branch time is needed to correctly interpret the data (e.g., in simulations spawned from a pre-industrial control run that has not reached equilibrium). Suppose a model with a “noleap” calendar carries out a control run with the time coordinate reported in units of “days since 1000-1-1”. Now suppose a historical simulation branches from that control run at day 3650 (i.e., on the date: 1010-1-1), and suppose the units of the time coordinate for the historical run are “days since 1850-01-01”. Then in the historical run, the following global attributes should appear:

```
branch_time_in_child = 0.0D0; # double precision float
branch_time_in_parent = 3650.0D0; # double precision float
parent_time_units = “days since 1000-1-1”; # character string
```

If the calendar used in the parent is different from the child, then the parent’s calendar should be recorded in “`parent_time_units`”. The CV for the calendar is the same as that for CF calendar attribute. For example in the above case, if the child’s calendar is “gregorian”, but the parent’s is “noleap”, then `parent_time_units=“days since 1000-1-1 (noleap)”`.

As another example, suppose a ScenarioMIP simulation “branches” from a historical run at the date 2015-1-1. Suppose a noleap calendar is used in each and both have units of “days since 1850-1-1”. (Note that ScenarioMIP simulations should always adopt the same units as the historical parent.) In this case the following global attributes should appear in the ScenarioMIP output file:

```
branch_time_in_child = 60225.0D0
branch_time_in_parent = 60225.0D0
parent_time_units = “days since 1850-1-1”.
```

As a third example, consider a decadal prediction experiment initialized from observations (e.g., from an “analysis”). In this case there is no parent so none of these attributes should be included. (Note that the initialization time of the decadal prediction experiments will be recorded as a scalar coordinate dimension, pointed to by the “`coordinates`” attribute, which is attached to each variable.)



6. A description and examples of this global attribute may be found in the document: CMIP5\_output\_metadata\_requirements ([http://cmip-pcmdi.llnl.gov/cmip5/docs/CMIP5\\_output\\_metadata\\_requirements.pdf](http://cmip-pcmdi.llnl.gov/cmip5/docs/CMIP5_output_metadata_requirements.pdf)).
7. creation\_date form: YYYY-MM-DDTHH:MM:SSZ (e.g., "2010-03-23T05:56:23Z")
8. For a given experiment, the realization\_index, initialization\_index, physics\_index, and forcing\_index are used to uniquely identify each simulation of an ensemble of runs contributed by a single model. These indices are defined as follows:
  - realization\_index = an integer ( $\geq 1$ ) distinguishing among members of an ensemble of simulations that differ only in their initial conditions (e.g., initialized from different points in a control run). Note that if two different simulations were started from the same initial conditions, the same realization number should be used for both simulations. For example if a historical run with "natural forcing" only and another historical run that includes anthropogenic forcing were both spawned at the same point in a control run, both should be assigned the same realization. Also, each so-called RCP (future scenario) simulation should normally be assigned the same realization integer as the historical run from which it was initiated. This will allow users to easily splice together the appropriate historical and future runs.
  - initialization\_index = an integer ( $\geq 1$ ), which should be assigned a value of 1 except to distinguish simulations performed under the same conditions but with different initialization *procedures*. In CMIP6 this index should invariably be assigned the value "1" except for some hindcast and forecast experiments called for by the DCPP activity. The initialization\_index can be used either to distinguish between different algorithms used to impose initial conditions on a forecast or to distinguish between different observational datasets used to initialize a forecast.
  - physics\_index = an integer ( $\geq 1$ ) identifying the physics version used by the model. In the usual case of a single physics version of a model, this argument should normally be assigned the value 1, but it is essential that a consistent assignment of physics\_index be used across all simulations performed by a particular model. Use of "physics\_index" is reserved for closely-related model versions (e.g., as in a "perturbed physics" ensemble) or for the same model run with slightly different parameterizations (e.g., of cloud physics). Model versions that are substantially different from one another should be given a different source\_id" (rather than simply assigning a different value of the physics\_index).
  - forcing\_index = an integer ( $\geq 1$ ) used to distinguish runs conforming to the protocol of a single CMIP6 experiment, but with different variants of forcing applied. One can, for example, distinguish between two historical simulations, one forced with the CMIP6-recommended forcing data sets and another forced by a different dataset, which might yield information about how forcing uncertainty affects the simulation.

Each data provider can assign whatever positive integers they like for the realization\_index, initialization\_index, physics\_index, and forcing\_index. For each source/experiment pair, however, consistency (in these indices) should be maintained across each parent/child pair whenever sensible (so that, for example, both the ScenarioMIP child and its "historical" parent simulation would be assigned the same set of index values for realization, initialization, physics, and forcing); the integer 1 should normally be chosen for each of these in the case of a single variant or for the primary variant (if there is one). This is only a suggestion, however; there should be no expectation on the part of users that every model will have a value of 1 assigned to any of the r, i, p, f indices, and even if a 1 is assigned it does not imply that it is the primary variant. Note also that a child spawned by a *control* run will not necessarily have the same "ripf" value as the control, since, for example, multiple realizations of an experiment will branch from the same control.

Note that none of the "ripf" indices can be omitted.

Example of a variant\_label: if realization\_index=2, initialization\_index=1, physics\_index=3, and forcing\_index=233, then variant\_label = "r2i1p3f233".

9. further\_info\_url has the form [https://furtherinfo.es-doc.org/<mip\\_era>.<institution\\_id>.<source\\_id>.<experiment\\_id>.<sub\\_experiment\\_id>.<variant\\_label>](https://furtherinfo.es-doc.org/<mip_era>.<institution_id>.<source_id>.<experiment_id>.<sub_experiment_id>.<variant_label>) (e.g.,

- "https://furtherinfo.es-doc.org/CMIP6.CAS\_FGOALS-g3.historical.none.r3i1p1f1"). The further\_info\_url page will be maintained by the es-docs project and will simply be a rendering by the Viewer tool of information provided by modeling groups and recorded in so-called CIM documents.
10. The "grid" global attribute can be used to describe the horizontal grid and regridding procedure. There is no standard form used to record this information, but it is suggested that when appropriate the following be indicated: brief description of native grid and resolution, and if data have been regridded, regridding procedure and description of target grid. Here are some examples:
    - grid = "native atmosphere T63 gaussian grid (128x64 lonxlat)"
    - grid = "data regridded to a CMIP6 standard 1x1 degree lonxlat grid from the native T63 grid using an area-average preserving method."
    - grid = "data regridded via bilinear interpolation to a 3x3 deg lonxlat grid from the native atmosphere T63 gaussian grid (128x64 lonxlat)"
    - grid = "native ocean tri-polar grid with 43200 ocean cells"
  11. Modeling groups may choose to report their output on the model's native grid and/or regrid it to one or more target grids. To distinguish between output reported on different grids, a "[grid label](#)" attribute is defined.

The rules for assigning grid labels should make it easy for users to select (using the ESGF search tools) CMIP output that is on a grid considered by each modeling group to best represent its model -- the so-called "primary" grid. If output is reported on the native grid, this is always deemed the "primary" grid. If output is *not* reported on the native grid, then modeling groups should regrid the data to some primary grid of its choosing. For the "primary" grid the following labels apply:

```
grid_label = "gn" (output is reported on the native grid, usually but not invariably at grid cell centers)
grid_label = "gr" (output is not reported on the native grid, but instead is regridded by the modeling group to a "primary grid" of its choosing)
grid_label = "gm" (global mean output is reported, so data are not gridded)
```

As noted below sometimes a "z" or "a" or "g" is appended to the labels to indicate "zonal means" or grids limited to Antarctica or Greenland.

If besides the "primary" grid, output is regridded to an additional grid, then for this output:

```
grid_label = "gr[i]" (a "secondary" grid), where <i> should be replaced by a positive integer less than 10, which distinguishes this output from other regridded output.
```

Note that:

- If model output is reported on a native grid, then if regridded output is also reported, it must not be labeled "gr", but instead should be of the form gr[i] (e.g., gr1, gr2, ...).
- The grid label provides no information about the grid other than to indicate whether or not the data have been regridded (from the native grid) and whether or not the grid is considered to be a "primary" grid by the data provider.
- Output for different variables may be reported on different grids, so "gn" and "gr" may not uniquely define a grid even within a single model.
- If a variable is reported on more than one grid, one of these grids must be labeled "gn" or "gr" (as appropriate), and others must be labeled as secondary grids ("gr1", "gr2", etc.).
- The grid labels ("gr1", "gr2", etc.) must be defined consistently for all variables reported from a single model on the same secondary grid. Thus, if "gr3" indicates output on a 2x2 degree grid for one variable, then the same label should be used for other variables output on this grid (unless for another variable output is considered to be primary, in which case the output would be labeled "gn" or "gr", as appropriate).
- Output on the same grid but from different models will usually not be labeled with the same grid\_label, since data providers independently assign the labels.

- A “nominal\_resolution” attribute must be defined as specified in [Appendix 2](#), and a brief description of the grid should be recorded in the “grid” global attribute (see note 10).
  - For zonal mean output, a “z” should be appended to the grid label that would apply before performing the zonal mean (e.g., “gnz”, “grz”, “gr2z”). The “gnz” label would likely only be appropriate when zonally-averaging data on a native cartesian latxlon grid.
  - For “site” data the label “gn” should be used presuming data are obtained from the single native grid cell located nearest each site.
  - For “transport through a straight” (and the like), the grid label should reflect the grid relied on in calculating the transport (presumably the native grid would be best for this purpose, so “gn”).
  - For output reported only over the region centered on Greenland, “g” should be appended to the grid\_label (e.g., “gng”, “grg”, “gr1g”). Similarly for output reported only over the region centered on Antarctica, “a” should be appended to the grid\_label.
12. The “license” attribute should record the following statement (with segments in square brackets optional, and with required, appropriate text entered in place of <\*> ): “CMIP6 model data produced by <Your Centre Name> is licensed under a Creative Commons Attribution-[\*]ShareAlike 4.0 International License (<https://creativecommons.org/licenses/>). Consult <https://pcmdi.llnl.gov/CMIP6/TermsOfUse> for terms of use governing CMIP6 output, including citation requirements and proper acknowledgment. Further information about this data, including some limitations, can be found via the further\_info\_url (recorded as a global attribute in this file)[ and at <some URL maintained by modeling group>]. The data producers and data providers make no warranty, either express or implied, including, but not limited to, warranties of merchantability and fitness for a particular purpose. All liabilities arising from the supply of the information (including any liability arising in negligence) are excluded to the fullest extent permitted by law.”
- \* The [\*] indicates that institutions may choose to use the Non-commercial version of this license by inserting the words “NonCommercial-” at this point, but this will significantly limit the use of the data in downstream climate mitigation and adaptation applications. Please do not simply copy the statement above when writing data; Some text must be entered, some text is optional and the symbols “[\*]” should not appear in the licensing text.
13. The “source” is used to fully identify the model and version. The first portion of the “source” attribute is used in constructing “[source id](#)”. [The “source\_id”, which must not exceed 16 characters in length, is the same as the “modified source\_id” (which appears as the first part of “source”-- see below) but with forbidden characters removed or replaced by a hyphen (“-”). All characters are forbidden in source\_id except a-z, A-Z, 0-9 and the hyphen (“-”).] Additionally, the “source” attribute must include the year (i.e., model vintage) when this model version was first used in a scientific application. It should also include information concerning the component models. The following template should be followed in constructing “source”:
- ```
<modified source_id> (<year>): atmos: <model_name> (<technical_name>, <resolution_and_levels>); ocean: <model_name> (<technical_name>, <native_resolution_and_levels>); seaIce: <model_name> (<technical_name>); land: <model_name> (<technical_name>); aerosol: <model_name> (<technical_name>); atmosChem: <model_name> (<technical_name>); ocnBgchem: <model_name> (<technical_name>); landIce: <model_name> (<technical_name>)
```

For some models, it may not make sense to include all these components, and none of the text following “<year>” is absolutely mandatory. As an example, “source” might contain the string:

```
CCSM2 (2002): atmos: CAM2 (cam2_0_brnchT_itea_2, T42L26); ocean: POP (pop2_0_ver_1.4.3, 3x2L15); seaIce: CSIM4; land: CLM2.0
```

The source and source\_id should not change even when some of the component models are inactive in some of the CMIP experiments. For example, if an AOGCM is named “SomeAOGCM 1.0 (2016): atmosphere: SomeAGCM; ocean: SomeOGCM; sea ice: SomeSeaIce”, then source\_id would be “SomeAOGCM-1-0” and these same names would be used whether or not the model were run in coupled mode or AMIP mode (with prescribed SST and sea ice).

14. The [source type](#) will depend on which experiment is being performed. Options in the CV include:
- AGCM (atmospheric general circulation model, including a land model),

- OGCM (ocean general circulation model, including a sea-ice model),
- AOGCM (atmosphere-ocean global climate model),
- LAND (land model but only if run “offline”),
- ISM (ice-sheet model, which may be run “offline” or coupled to an AOGCM),
- RAD (radiation code but only if run “offline”),
- BGC (for a model component that includes a biogeochemical treatment which at the very least can account for carbon reservoirs and fluxes in the atmosphere, terrestrial biosphere, and ocean; for some model configurations, only part of the BGC component will be active. For example, in a BGC model coupled to an AGCM, the ocean component of the BGC might be inactive, but the source\_type would be “AGCM BGC”. For a BGC model run coupled to an AOGCM, the source\_type should include “AOGCM BGC” both when atmospheric concentrations are calculated and when they are prescribed),
- CHEM (appears with either AOGCM or AGCM in models that calculate, rather than rely on prescribed concentrations of atmospheric oxidants including at least ozone),
- AER (appears with AOGCM or AGCM in models that calculate tropospheric aerosols driven by emission fluxes, rather than relying on prescribed concentrations),
- SLAB (a slab-ocean model).

All types that apply to the model (i.e., are active in a given experiment) should appear in a list with each type separated by a single space.

Sometimes source\_type for CMIP6 simulations will be described by a single label, as in “AOGCM” or “AGCM” or “OGCM”, but when additional interactive components are included, then multiple labels should be found in source\_type (e.g., “AOGCM BGC”, “AGCM CHEM AER”, “AOGCM AER”, “AOGCM ISM”, “AGCM SLAB”, “OGCM BGC”). Single labels will be used for “offline” models such as “ISM” and “RAD”.

For each of the CMIP6 experiments, the list of components that are required and the ones that are allowed but optional can be found in [http://rawgit.com/WCRP-CMIP/CMIP6\\_CVs/master/src/CMIP6\\_experiment\\_id.html](http://rawgit.com/WCRP-CMIP/CMIP6_CVs/master/src/CMIP6_experiment_id.html) (or, alternatively, [here](#)). The following combinations for source\_type can be found there:

```
AGCM  [BGC]  [AER]  [CHEM]  [SLAB]
AOGCM [BGC]  [AER]  [CHEM]  [ISM]
OGCM
LAND
ISM
RAD
```

where the text in brackets sometimes must appear, other times, may appear, and in still other cases should \*not\* appear, depending on the experiment. The ordering of the components is arbitrary in source\_type, and all brackets should be removed (see examples in earlier paragraph above).

15. tracking\_id should be of the form “hdl:21.14100/<uuid>” (e.g., “hdl:21.14100/02d9e6d5-9467-382e-8f9b-9300a64ac3cd”). The tracking\_id should be unique for each file published in ESGF. The <uuid> should be generated using the OSSP utility which supports a number of different DCE 1.1 variant UUID options. For CMIP6, version 4 (random number based) is required. Download the software from <http://www.oss.org/pkg/lib/uuid/>.
16. It is recommended that variant\_info include information identifying major distinguishing features of a variant, but care should be taken to record correct information. Prudence dictates that this attribute include a warning along the following lines: “Information provided by this attribute may in some cases be flawed. Users can find more comprehensive and up-to-date documentation via the further\_info\_url global attribute.”
17. The information stored in the “forcing” attribute in CMIP5 may in CMIP6 appear in the variant\_info attribute.

## Data Reference Syntax (DRS) components:

The DRS identifies experiments, simulations, ensembles of experiments, and atomic datasets. Some of the DRS components are used, for example, to construct file names, directory structures, the further\_info\_url, and in facets of some search tools. The following components are needed for CMIP6 (along with their CMIP5 counterparts):

|                |                            |                                                                                                                            |
|----------------|----------------------------|----------------------------------------------------------------------------------------------------------------------------|
| activity_id*   | (CMIP5: "activity")        | see <a href="#">CMIP6 activity id.json</a>                                                                                 |
| institution_id | (CMIP5: "institute")       | see <a href="#">CMIP6 institution id.json</a>                                                                              |
| source_id      | (CMIP5: "model")           | see <a href="#">CMIP6 source id.json</a>                                                                                   |
| experiment_id  | (CMIP5: "experiment")      | see <a href="#">CMIP6 experiment id.json</a>                                                                               |
| variable_id    | (CMIP5: "variable name")   | see <a href="#">data request</a>                                                                                           |
| table_id       | (CMIP5: "table_id")        | see <a href="#">CMIP6 table id.json</a>                                                                                    |
| variant_label  | (CMIP5: "ensemble member") | construct from realization, initialization, physics, and forcing indices                                                   |
| version        | (CMIP5: "version number")  | indicating approximate date of model output file. (This is the only DRS element that is not stored as a global attribute.) |

We need additional components in CMIP6 to accommodate the more complex structure:

|                   |                                                                     |                                                  |
|-------------------|---------------------------------------------------------------------|--------------------------------------------------|
| sub_experiment_id | (set to "none" for most experiments)                                | see <a href="#">CMIP6 sub_experiment id.json</a> |
| grid_label        | (needed to distinguish the same field stored on more than one grid) | see <a href="#">CMIP6 grid label.json</a>        |
| mip_era           | (needed to distinguish CMIP5 experiments and datasets from CMIP6.)  | set to "CMIP6"                                   |
| member_id         | a compound construction from sub_experiment_id and variant_label    | see below for further information                |

As in CMIP5, we also define additional DRS elements because they can be helpful in providing data discovery services:

|                    |                           |                                                                                  |
|--------------------|---------------------------|----------------------------------------------------------------------------------|
| frequency          | (CMIP5: "frequency")      | see <a href="#">CMIP6 frequency.json</a>                                         |
| realm*             | (CMIP5: "modeling realm") | see <a href="#">CMIP6 realm.json</a>                                             |
| product            | (CMIP5: "product")        | set to "model-output" in CMIP6                                                   |
| nominal_resolution |                           | see <a href="#">CMIP6 nominal_resolution.json</a> and <a href="#">Appendix 2</a> |
| source_type*       |                           | see <a href="#">CMIP6 source_type.json</a>                                       |

The DRS elements marked with an asterisk (\*) are associated with global attributes that may be space-separated lists of values. Only the first item in a list is recognized by the DRS, but in faceted searches all listed items will be recognized.

## File name template:

Before constructing file names and directory structures, it is useful to define a member\_id, which can be used to distinguish among different simulations belonging to a root experiment. The member\_id is constructed from the sub\_experiment\_id and variant\_label using the following algorithm:

```
if sub_experiment_id = "none"
    member_id = <variant_label>
else
    member_id = <sub_experiment_id>-<variant_label>
endif
```

With this segment defined, the file name can be constructed consistent with the following template:

file name = <variable\_id>\_<table\_id>\_<source\_id>\_<experiment\_id >\_<member\_id>\_<grid\_label>[\_<time\_range>].nc

For time-invariant fields, the last segment (time\_range) above is omitted.

Example when there is no sub-experiment: tas\_Amon\_GFDL-CM4\_historical\_r1i1p1f1\_gn\_196001-199912.nc

Example with a sub-experiment: pr\_day\_CNRM-CM6-1\_dcppA-hindcast\_s1960-r2i1p1f1\_gn\_198001-198412.nc

All strings appearing in the file name are constructed using only the following characters: a-z, A-Z, 0-9, and the hyphen ("-"), except the hyphen must not appear in variable\_id. Underscores are prohibited throughout except as shown in the template.

Note that the last segment of the file name indicates the time-range spanned by the data in the file, and is omitted when inappropriate (e.g., if a variable is "fixed" for all time). The format for this segment is as in CMIP5 (see [http://cmip-pcmdi.llnl.gov/cmip5/docs/cmip5\\_data\\_reference\\_syntax.pdf](http://cmip-pcmdi.llnl.gov/cmip5/docs/cmip5_data_reference_syntax.pdf)):

The <time\_range> is a string generated consistent with the following:

```
If frequency = "fx" then
    <time_range>=""
```

else

<time\_range> =  $N_1 - N_2$  where  $N_1$  and  $N_2$  are integers of the form 'yyyy[MM[dd[hh[mm[ss]]]]][<suffix>]' (expressed as a string, where 'yyyy', 'MM', 'dd', 'hh', 'mm' and 'ss' are integer year, month, day, hour, minute, and second, respectively)

endif

where <suffix> is defined as follows:

if the variable identified by variable\_id has a time dimension with a "climatology" attribute then

suffix = "-clim"

else

suffix = ""

endif

and where the precision of the time\_range strings is determined by the "frequency" global attribute as specified in Table 2.

**Table 2:** Precision of time labels used in file names.

| <b>Frequency</b>                                | <b>Precision of time label</b> | <b>Notes</b>                                                                                                                                                                                                                                                                   |
|-------------------------------------------------|--------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| yr, dec, yrPt                                   | “yyyy”                         | Label with the years recorded in the first and last coordinate values.                                                                                                                                                                                                         |
| mon, monC                                       | “yyyyMM”                       | For “mon”, label with the months recorded in the first and last coordinate values; for “monC” label with the first and last months contributing to the climatology.                                                                                                            |
| day                                             | “yyyyMMdd”                     | Label with the days recorded in the first and last coordinate values.                                                                                                                                                                                                          |
| 6hr, 3hr, 1hr,<br>1hrCM, 6hrPt,<br>3hrPt, 1hrPt | “yyyyMMddhhmm”                 | Label 1hrCM files with the beginning of the first hour and the end of the last hour contributing to climatology (rounded to the nearest minute); for other frequencies in this category, label with the first and last time-coordinate values (rounded to the nearest minute). |
| subhrPt                                         | “yyyyMMddhhmmss”               | Label with the first and last time-coordinate values (rounded to the nearest second)                                                                                                                                                                                           |
| fx                                              | Omit time label                | This frequency applies to variables that are independent of time (“fixed”).                                                                                                                                                                                                    |



## Directory structure template:

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

### Note:

- <version> has the form “vYYYYMMDD” (e.g., “v20160314”), indicating a representative date for the version. Note that files contained in a single <version> subdirectory at the end of the directory path should represent all the available time-samples reported from the simulation; a time-series can be split across several files, but all the files must be found in the same subdirectory. This implies that <version> will not generally be the actual date that all files in the subdirectory were written or published.
- If multiple activities are listed in the global attribute, the first one is used in the directory structure.

Example when there is no sub-experiment: CMIP6/CMIP/NOAA-GFDL/GFDL-CM4/1pctCO2/r1i1p1f1/Amon/tas/gn/v20150322

Example with a sub-experiment: CMIP6/DCPP/CNRM-CERFACS/CNRM-CM6-1/dcppA-hindcast/s1960-r2i1p1f3/day/pr/gn/v20160215

The above directory structure is not the same as CMIP5’s directory structure.

**Table 3:** Quality assurance checks. [Note that “CMIP6 data request” refers to information recorded in the “dreq” xml files.]

| <b>CMIP6 global attribute</b> | <b>CMOR source of global attribute</b>           | <b>Required attribute?</b> | <b>Check that type is</b> | <b>Require check against template or CV</b>                        | <b>Check also in file name?</b> |
|-------------------------------|--------------------------------------------------|----------------------------|---------------------------|--------------------------------------------------------------------|---------------------------------|
| activity_id                   | CMIP6_experiment_id.json                         | yes                        | string                    | CV (consistent with experiment_id)                                 | no                              |
| branch_method                 | user input                                       | conditionally              | string                    | no                                                                 | no                              |
| branch_time_in_child          | user input                                       | conditionally              | double precision          | no                                                                 | no                              |
| branch_time_in_parent         | user input                                       | conditionally              | double precision          | no                                                                 | no                              |
| comment                       | user input                                       | no                         | string                    | no                                                                 | no                              |
| contact                       | user input                                       | no                         | string                    | no                                                                 | no                              |
| Conventions                   | Table 1                                          | yes                        | string                    | CV                                                                 | no                              |
| creation_date                 | CMOR-generated                                   | yes                        | string                    | must be valid date represented as an ISO 8601 date and time string | no                              |
| data_specs_version            | CMIP6 data request                               | yes                        | string                    | CV                                                                 | no                              |
| experiment                    | CMIP6_experiment_id.json                         | yes                        | string                    | CV (consistent with experiment_id)                                 | no                              |
| experiment_id                 | user input (drawn from CMIP6_experiment_id.json) | yes                        | string                    | CV                                                                 | yes                             |
| external_variables            | cell_measure values in CMIP6 data request        | conditionally              | string                    | CV (consistent with variable_id and table_id)                      | no                              |
| forcing_index                 | user input                                       | yes                        | integer >0                | no                                                                 | no                              |
| frequency                     | CMIP6 data request                               | yes                        | string                    | CV (consistent with table_id)                                      | no                              |
| further_info_url              | CMOR-generated, based on other global attributes | yes                        | string                    | CV (consistent with several global attributes)                     | no                              |
| grid                          | user input                                       | yes                        | string                    | no                                                                 | no                              |
| grid_label                    | user input (drawn from CMIP6_grid_label.json)    | yes                        | string                    | CV                                                                 | yes                             |
| history                       | user, supplemented by CMOR                       | no                         | string                    | no                                                                 | no                              |
| initialization_index          | user input                                       | yes                        | integer >0                | no                                                                 | no                              |
| institution                   | CMIP6_institution_id.json                        | yes                        | string                    | CV (consistent with institution_id)                                | no                              |
| institution_id                | user input                                       | yes                        | string                    | CV (consistent with source_id)                                     | no                              |

|                      |                                                                  |               |            |                                                       |     |
|----------------------|------------------------------------------------------------------|---------------|------------|-------------------------------------------------------|-----|
| license              | user input                                                       | yes           | string     | defined structure/format and partially-defined text   | no  |
| mip_era              | Table 1                                                          | yes           | string     | CV (= "CMIP6")                                        | no  |
| nominal_resolution   | user input (drawn from CMIP6_nominal_resolution.json)            | yes           | string     | CV                                                    | no  |
| parent_activity_id   | CMIP6_experiment_id.json                                         | conditionally | string     | CV (consistent with parent_experiment_id)             | no  |
| parent_mip_era       | Table 1                                                          | conditionally | string     | CV (= "CMIP6")                                        | no  |
| parent_experiment_id | CMIP6_experiment_id.json                                         | conditionally | string     | CV (consistent with experiment_id)                    | no  |
| parent_source_id     | user input (drawn from CMIP6_source_id.json)                     | conditionally | string     | CV (usually the same as source_id)                    | no  |
| parent_time_units    | user input                                                       | conditionally | string     | check validity against udunits and the CF calendar CV | no  |
| parent_variant_label | user input                                                       | conditionally | string     | defined structure/format                              | no  |
| physics_index        | user input                                                       | yes           | integer >0 | no                                                    | no  |
| product              | Table 1                                                          | yes           | string     | CV (= "model-output")                                 | no  |
| realization_index    | user input                                                       | yes           | integer >0 | no                                                    | no  |
| realm                | CMIP6 data request                                               | yes           | string     | CV (consistent with table_id)                         | no  |
| references           | user input                                                       | no            | string     | no                                                    | no  |
| source               | Constructed from components recorded in CMIP6_source_id.json     | yes           | string     | CV (consistent with source_id)                        | no  |
| source_id            | user input (drawn from CMIP6_source_id.json)                     | yes           | string     | CV                                                    | yes |
| source_type          | user input (drawn from CMIP6_source_type.json)                   | yes           | string     | CV (consistent with experiment_id)                    | no  |
| sub_experiment       | CMIP6_sub_experiment_id.json                                     | yes           | string     | CV (consistent with sub_experiment_id)                | no  |
| sub_experiment_id    | user input (drawn from list in CMIP6_sub_experiment_id.json)     | yes           | string     | CV (consistent with experiment_id)                    | yes |
| table_id             | user input (drawn from CMIP6_table_id.json)                      | yes           | string     | CV                                                    | yes |
| title                | CMOR-generated (based on, source_id, mip_era, and experiment_id) | no            | string     | no                                                    | no  |
| tracking_id          | CMOR-generated uuid with handle prefix from Table 1              | yes           | string     | defined structure/format                              | no  |

|               |                                                           |     |        |                                                   |     |
|---------------|-----------------------------------------------------------|-----|--------|---------------------------------------------------|-----|
| variable_id   | user input (drawn from CMIP6 data request)                | yes | string | CV (consistent with table_id)                     | yes |
| variant_info  | user input                                                | no  | string | no                                                | no  |
| variant_label | CMOR-generated, based on 4 user-provided indices ("ripf") | yes | string | defined structure/format constructed from indices | yes |

## Appendix 1: Global Attributes for Labeling Experiments

Global attributes that label experiments are needed to construct filenames and directories and can generally be used as search facets. Together, they should have the following characteristics:

- Uniquely label each experiment within CMIP6 and distinguish experiments with specified conditions that differ in any way
- Easily be interpreted and remembered
- Facilitate representations of groups of experiments that are closely related (e.g., same forecast conditions but different start dates, or experiment with an “offline” model driven by output from various models)

Often several simulations will be performed that satisfy the conditions specified for an experiment. For example, simulations of the historical period can branch from various points in a control run, and each of these will satisfy the conditions defining the experiment. Together, such simulations constitute a “conforming ensemble” with members all satisfying the same experiment specifications. There are also occasional cases where the experiment designers (MIP leaders) define a family of related simulations and choose to label these with a common “root” experiment name, followed by a “sub-experiment” name. An example of this is the set of decadal prediction hindcasts that are all run similarly but started from different start dates (with each simulation identified by a different sub-experiment label). Such “defined ensembles” of experiments will be labeled with a “root” experiment name, and a “sub-experiment\_id” will be used to distinguish among members in the ensemble.

To accommodate the various CMIP6 experiments, we define three global attributes:

- `experiment_id`: the label identifying the “root” experiment
- `sub_experiment_id`: the label identifying a sub-experiment of a “defined ensemble”; otherwise set to “none”
- `variant_label`: a label constructed from 4 indices stored as global attributes:

`variant_label = r<k>i<l>p<m>f<n>`

where

`k = realization_index`

`l = initialization_index`

`m = physics_index`

`n = forcing_index`

Besides the above identifiers, additional descriptive information concerning the experiment is provided by the following global attributes:

- `experiment`: brief descriptor of experiment (using CV)
- `sub_experiment`: brief descriptor of sub-experiment (using CV); if `sub_experiment_id = “none”`, then `sub_experiment = “none”`.
- `variant_info`: brief descriptor of what is unique about this “ripf” variant.

For the group of experiments included in CMIP6, the following structure will usually be followed:

- Each experiment\_id will comprise one or more segments separated by hyphens.
- The first segment indicates that an experiment should be run with a model other than an AOGCM or a concentration-driven ESM. (This segment is omitted in experiments for AOGCMs and concentration-driven ESMs.) CMIP6 examples of the first segment (shown in parentheses) include:
  - Offline radiation code experiments (“rad”)
  - Uncoupled ice-sheet models forced by AOGCM output (“ism”)
  - Atmosphere (and land surface) models forced by prescribed SSTs and sea ice (e.g., “amip”, “piSST”, “piClim”, “histSST”, “ssp370SST”, “aqua”, “futureSST”, “G6SST1”, “G6SST2”, “G7SST1”, “G7SST2”, “highresSST”, “a4SST”, “a4SSTice”)
  - Offline land-surface model (“land”)
  - Ocean and sea ice model forced by prescribed atmospheric conditions (“omip1” or “omip2”)
  - Earth system model forced by emissions (rather than concentrations) of CO2 (“esm”)
- The next segment is the first indication of experiment conditions
- Any additional segments indicate some relatively small variation on experiment conditions defined by the previous segment.

We now provide two examples of the global attributes relevant to identifying a CMIP6 experiment, and the filenames and directory structures that make use of these global attributes.

Example 1: The common case when there are no sub-experiments:

Global attributes (relevant to experiment definition):

```
experiment_id = "1pctCO2"  
experiment = "1 percent per year increase in CO2 concentration"  
sub_experiment_id = "none"  
sub_experiment = "none"  
realization_index = 1  
initialization_index = 1  
physics_index = 1  
forcing_index = 1  
variant_label = "r1i1p1f1"  
variant_info = "realization 1"
```

file name: tas\_Amon\_CCSM2-1\_1pctCO2\_r1i1p1f1\_gn\_202001-202912.nc

directory structure: CMIP6/CMIP/NCAR/CCSM2-1/1pctCO2/r1i1p1f1/Amon/tas/gn/v20150320/

Example 2: The uncommon case (in CMIP6) when there are sub-experiments defined:

Global attributes (relevant to experiment definition):

experiment\_id = "dcppA-hindcast"

experiment = "year 1-5 hindcast initialized based on observations and using historical forcing"

sub\_experiment\_id = "s1960"

sub\_experiment = "initialized near end of year 1960"

realization\_index = 1

initialization\_index = 2

physics\_index = 1

forcing\_index = 1

variant\_label = "r1i2p1f1"

variant\_info = "initialized using anomaly approach (method 2)"

file name: tas\_Amon\_CCSM2-1\_hindcast\_s1960-r1i2p1f1\_gn\_198001-198412.nc

directory structure: CMIP6/DCPP/NCAR/CCSM2-1/dcppA-hindcast/s1960-r1i2p1f1/Amon/tas/gr/v20150320/

## Appendix 2: Algorithms for Defining the “nominal\_resolution” Attribute

There are various ways grid resolution might be defined, but in CMIP6 this should be done in the same way by all models. If the following procedure seems inappropriate for a model, the modeling group may request an exception from the WGCM Infrastructure Panel (WIP). In general, the nominal resolution characterizes the resolution of *the grid used to report model output fields*, which may differ from the native grid on which the fields are calculated by the model.

### Algorithm for defining the nominal\_resolution global attribute:

1. For each grid cell, calculate the distance (in km) between each pair of cell vertices and select the maximum distance ( $d^{\max}$ ). For lonxlat grid cells, for example,  $d^{\max}$  would be the diagonal distance.
2. Calculate the mean over all cells of  $d^{\max}$ , weighting each by the grid-cell’s area ( $A$ ). This defines the “mean resolution” ( $\overline{d^{\max}}$ ). The formula is:

$$\overline{d^{\max}} = \frac{\sum d_i^{\max} A_i}{\sum A_i}$$

where the sums are over all grid cells except for the following cases:

- For a global ocean grid, only sum over the ocean grid cells.
- For a sea ice model calculated on the ocean grid, include *all* ocean cells, whether or not they contain sea ice.
- For land surface models calculated on the atmospheric grid include all grid cells (not just those over land).
- For a land surface model calculated on its own grid, include all land grid cells .
- For data reported on a sub-domain of the globe (e.g., northern high latitudes only), include only those grid cells in the domain.
- For data reported at individual sites, calculate as if every grid cell contained one site (i.e., include all grid cells).
- For zonal means, global means, sector or basin means, and similar area-means, the data provider may either report the nominal resolution of the native grid or the resolution of the primary grid on which data are reported.

3. Except in the case of a CMIP6 “standard grid” (see item 4 below), define the global attribute “nominal\_resolution” according to:

if  $\overline{d^{\max}} < 0.72$  km, nominal\_resolution = “0.5 km”  
else if  $\overline{d^{\max}} < 1.6$  km, nominal\_resolution = “1 km”  
else if  $\overline{d^{\max}} < 3.6$  km, nominal\_resolution = “2.5 km”  
else if  $\overline{d^{\max}} < 7.2$  km, nominal\_resolution = “5 km”



else if  $\overline{d^{max}} < 16$  km, nominal\_resolution = "10 km"  
 else if  $\overline{d^{max}} < 36$  km, nominal\_resolution = "25 km"  
 else if  $\overline{d^{max}} < 72$  km, nominal\_resolution = "50 km"  
 else if  $\overline{d^{max}} < 160$  km, nominal\_resolution = "100 km"  
 else if  $\overline{d^{max}} < 360$  km, nominal\_resolution = "250 km"  
 else if  $\overline{d^{max}} < 720$  km, nominal\_resolution = "500 km"  
 else if  $\overline{d^{max}} < 1600$  km, nominal\_resolution = "1000 km"  
 else if  $\overline{d^{max}} < 3600$  km, nominal\_resolution = "2500 km"  
 else if  $\overline{d^{max}} < 7200$  km, nominal\_resolution = "5000 km"  
 else nominal\_resolution = "10000 km"

The different nominal\_resolution values are approximately spaced logarithmically and the bounds on each are logarithmically approximately half-way between the values.

For a regular latlon global grid it is possible to calculate the approximate mean resolution analytically:

$$\overline{d^{max}} = \frac{r_{earth} \Delta\phi}{2} \left[ 1 + \frac{\Delta\phi^2 + \Delta\lambda^2}{\Delta\phi \Delta\lambda} \tan^{-1} \left( \frac{\Delta\lambda}{\Delta\phi} \right) \right]$$

where  $r_{earth}$  is the radius of the Earth (in km), and  $\Delta\phi$  and  $\Delta\lambda$  are the latitude and longitude angular dimensions of each cell (measured in radians). If these dimensions are identical, then the mean resolution is:

$$\overline{d^{max}} = \frac{r_{earth} \Delta\phi}{2} \left[ 1 + \frac{\pi}{2} \right] \approx 1.2854 r_{earth} \Delta\phi$$

For a 0.5x0.5 degree grid, the mean resolution, according to this formula is 71.5 km (given 6371 km as the Earth's radius).

- Note that for so-called "standard CMIP6 grids," nominal\_resolution will be assigned a string defined by the WGCM Infrastructure Panel (WIP), rather than applying the above algorithm. Currently the only WIP-assigned standard grid is a 1x1 degree longitude by latitude grid with 360 longitudes and 180 latitudes of equal (angular) width, and with one longitude centered at 0.5 degree east longitude (consistent with the World Ocean Atlas). For this CMIP6 standard grid, nominal\_resolution = "1x1 degree", rather than "100 km" (which would be the result of the above algorithm). Defining the standard grid resolution in this way makes it easy for users to download only data that have been regridded to the standard grid, since "1x1 degree" will be selectable from the "nominal\_resolution" search facet on ESGF.

5. When the above formula for a regular lonxlat global grid is inapplicable, one can rely on a python code to calculate nominal\_resolution for any grid. The following links lead to the code and its documentation:

- Code documentation: [https://pcmdi.github.io/nominal\\_resolution/html/index.html](https://pcmdi.github.io/nominal_resolution/html/index.html).
- The code can be obtained via a conda package:

```
conda install -c pcmdi nominal_resolution
```
- The package repository is hosted on Github at: [https://github.com/pcmdi/nominal\\_resolution](https://github.com/pcmdi/nominal_resolution)
  - The library source (api.py) is in the [lib](#) directory.
  - The test codes reside in the [tests](#) directory.

### Appendix 3: Document version information

The document version number consists of 3 integers separated by “.” The first integer is “6”, indicating the document applies to CMIP6. The second integer will be incremented if changes are made that likely will require modifications in existing software or output files (e.g., an addition of a new global attribute). The third digit will be incremented whenever a new release involves only minor changes (e.g., to a list of values appearing in a CV).

6.0.0 (14th September 2016) - Initial release

6.0.1 (5 October 2016):

- Replaced source\_type option “BCM” with “BGCM” (biogeochemical model).
- Eliminated source\_type options: “RCM (regional climate model)”, “ESD (empirical statistical downscaling model)”, and “EMIC (earth-system model of intermediate complexity)”. This was done
- Modified algorithm for determining grid\_resolution so that the label value will now nearly always be logarithmically the closest one to d-max. This is consistent with the labels being logarithmically approximately evenly spaced.
- Corrected typos in note 13 after Table 1: “sea ice” is now “sea\_ice” (for consistency with other labels in that note).
- Added text at the beginning of this Appendix about incrementing the version number.

6.1.0 (15 November 2016):

- Changed “grid\_resolution” to “nominal\_resolution”, which is less definitive and more appropriate.
- Noted that the length of source\_id is limited to no more than 16 characters.

- Description of grid\_label now includes specifications for global mean data and data stored on regional grids centered over Greenland or Antarctica.
- Corrected model\_id to read source\_id and model\_type to read source\_type in a couple of places.
- After discussions with Pierre Friedlingstein, renamed “BGCM” to read “BGC” and eliminated distinction between ESM and “AOGCM BGC”.
- Expanded description of source\_type attribute to help clarify.

6.2.0 (2 December 2016; shortened URL is [goo.gl/v1drZl](http://goo.gl/v1drZl)):

- Reversed the order of the <source\_id>\_<experiment\_id> segment of the filename (with “source” now first) to be consistent with subsequent examples of filename in this document and more like the ordering of the CMIP5 filename and the directory structure.
- Eliminated one of the optional sentences in the “license” attribute because it seemed unnecessary, given the more general statement made in the sentence that followed.

6.2.1 (21 December 2016):

- Corrected the URL in the license statement to now point to <https://pcmdi.llnl.gov/home/CMIP6/CitationRequirements6-0.html> where the acknowledgement guidelines will be located.

6.2.2 (23 March 2017):

- Corrected the Creative Commons link in the license statement to now point to <https://creativecommons.org/licenses/>. Also changed “should be” to “must be”, and modified spaces, hyphens, and quotes in the “Attribution-ShareAlike” part of the statement.. Also corrected link to terms of use document.
- Change “UGRID-0.9”, which can appear in the conventions attribute to “UGRID-1.0”.
- Clarified that one statement in footnote 5 referred to the decadal prediction runs only.
- Changed CMIP6 specification of “product” global attribute from “output” to “model-output”. This better describes it and clearly differentiates it from “observations”, which is the text assigned to this attribute in input4MIPs and obs4MIPs.
- Corrected a few entries in the CMOR Source column of Table 2 (which has now been renumbered Table 3).
- Added links in several places to the reference CV’s found at [https://github.com/WCRP-CMIP/CMIP6\\_CVs](https://github.com/WCRP-CMIP/CMIP6_CVs).
- Clarified that the data\_specs\_version attribute records the version of the data request relied on in preparing model output.
- Added specific details on defining the time labels that appear in file names, which required addition of a new table and required Table 2 to be renumbered Table 3.
- Adjusted (by less than 3%) the values used to determine nominal resolution in [Appendix 2](#), so that models with latlon resolution of 0.25, 0.5, 2.5 and 5.0 are now classified with nominal resolution of “25 km”, “50 km”, “250 km” and “500” km, respectively (instead of 50, 100, 500, and 1000 km).

#### 6.2.3 (4 April 2017):

- Deleted a double entry in table 2 and added a missing entry.
- Added links to the reference controlled vocabularies in the section defining the DRS.

#### 6.2.4 (14 July 2017):

- Revised note 8 for improved clarity.
- In note 13, corrected the terms used to describe each realm to be consistent with the CMIP6 CV.
- Augmented Table 2 with notes on how time-labels for file names should be constructed. Also, changed frequency “decadal” to “dec” and eliminated 3hrClim since it isn’t needed for CMIP6.
- Corrected the examples of file name to make the order of the elements consistent with the template.
- Reworded a few sentences to improve clarity
- Corrected some instances of “product” to indicate for CMIP6 this should have the value “model-output”.
- Corrected CMIP6 version in Conventions attribute to be “CMIP-6.2”.

#### 6.2.5 (14 September 2017):

- Corrected/added options for frequency appearing in Table 2.
- Expanded notes describing the directory template.
- Corrected a few typos and URL addresses.

#### 6.2.6 (20 December 2017):

- Expanded description of realization\_index, initialization\_index, physics\_index, and forcing\_index in note 8 following Table 1.
- Removed “none provided” as an example of a branch\_method in Table 1.
- Corrected the example of further\_info\_url given in note 9 following Table 1.
- Revised the description of source\_type=“BGC” in note 14 following Table 1.
- Expanded explanation about how the tracking\_id should be generated (in note 15 following Table 1).
- Added missing items to the DRS: member\_id, nominal\_resolution, and source\_type.
- Changed the name of institution and source appearing in the examples of file names and directory structures so that they are now consistent with names currently registered for CMIP6.
- Corrected a few typos.
- Indicated in Table 1 which global attributes are part of the DRS and which will likely be used as search facets in CoG.

#### 6.2.7 (10 September 2018):

- Added to Appendix 2 an item (5) pointing to a python code that can be used to calculate nominal resolution.

- Made corrections so that when a rectangular longitude-latitude grid is described, the longitude dimension always appears first.