

Obs4MIPs Data Specifications

ODS2.5

(last update: March 2, 2024)

NEAR FINAL DRAFT

Contributors to ODS2.5 and/or earlier versions: Peter Gleckler (PCMDI), Karl Taylor (PCMDI), Paul Durack (PCMDI), Denis Nadeau (LANL), Jim Biard (NOAA/NCEI), Robert Ferraro (JPL), Stephan Finkensieper (DWD), Birgit Hassler (DLR), Andrew Manaster (RSS), Carl Mears (RSS), Simon Pinnock (ESA), Scott Stevens (NOAA/NCEI), Matthias Tuma (WCRP), Briony Turner (CMIP-IPO), Allison Waterfall (CEDA) and Kate Willet (MOHC)

[Latest obs4MIPs tables and controlled vocabulary](#)

Previous version [ODS2.1 : September 22, 2017](#)

[CONTENTS](#)

[Introduction](#)

[What is new in ODS2.5?](#)

[Global attributes](#)

[Data Reference Syntax \(DRS\) components:](#)

[File name template:](#)

[Directory structure template:](#)

[Sample file header](#)

[Appendix 1: Coordinate bounds](#)

[Appendix 2: Guidance for defining and registering source information \(product description\)](#)

[Appendix 3: Document version information](#)

Introduction

The purpose of [obs4MIPs](#) is to facilitate comparison of gridded observational data with model output from WCRP intercomparison projects, notably the [Coupled Model Intercomparison Project, CMIP](#). To accomplish this, the organization and description of CMIP and obs4MIPs data are closely coordinated, as described in this document. The details of this technical alignment, including the data structure and metadata requirements, facilitate coordinated delivery (via the [Earth System Grid Federation](#), ESGF) and use of the data for model evaluation, research and development.

[The original set of obs4MIPs contributions adhered to guidelines](#) (circa 2012, hereafter ODS1.0) aligned with the CMIP5 data specifications. This early phase of obs4MIPs was challenging because the software infrastructure relied upon for CMIP ([discussed below as “CMOR”](#)) did not readily accommodate observations. This deficiency in CMOR was rectified in CMIP6, as the obs4MIPs team worked closely with the [Working Group on Coupled Models \(WGCM\) Infrastructure Panel \(WIP\)](#) aligning ODS2.1 with the [CMIP6 data specifications](#). At that time obs4MIPs coordination and governance was provided by the WCRP’s Data Advisory Council’s (WDAC) “Observations for Model Evaluation Task Team”, but since 2022 this role has been fulfilled via continued coordination with the WIP and an obs4MIPs Steering Panel with the technical support of the [CMIP International Project Office \(CMIP IPO\)](#).

To avoid naming conflicts in the identification of data sources and institutions, it is necessary for data providers to register the name of their institution and other information about their data sets *prior to generating* obs4MIPs data sets. This is done by submitting an “issue” on the [obs4MIPs-cmor-tables Github repository](#). (See [Appendix 2](#) for the information requested.)

Obs4MIPs metadata specifications for the following categories are described in this document:

1. [Global attributes](#): Explicitly defined metadata included in netCDF files that describe the contents of the dataset and provenance (e.g., <source_id>).
2. [Data Reference Syntax \(DRS\)](#): Some attributes (e.g.,<institution_id>, <source_id>, <variable_id>) comprise the data reference syntax (DRS) for obs4MIPs, which closely parallels the DRS of CMIP6. The DRS is used, for example, in file names, directory structures, and in facets of some search tools such as ESGF.
3. [Directory structure](#): This can be thought of in the traditional sense and includes selected DRS information used as subdirectories. As in CMIP, it is explicitly defined in obs4MIPs to facilitate organization of a federated database and searching for data with ESGF.
4. [Filenames](#): Like the directory structure, filenames are constructed based on a template that relies mainly on CV entries.

This document also includes a discussion of the inclusion (or not) of grid cell bounds. This information makes it possible for users of the data to weight the influence of each cell based on its surface area (or volume), e.g., for calculating a domain mean. A user can estimate

the grid cell bounds, but in general those closest to the original data are better positioned to construct or estimate them. We recommend that data providers include grid bounds ([Appendix 1](#)).

What is new in ODS2.5?

The standards used to describe CMIP data are now quite mature, with the general structure not changing significantly from one phase of CMIP to the next. ODS2.5 thus only includes minor changes to ODS2.1 while striving to retain consistency with existing (CMIP6) protocols. Minor updates to ODS2.5 may be required as the data protocols of CMIP7 are established. As with CMIP, making changes to ODS only as necessary helps ensure continuity and backwards compatibility as both projects advance. The primary refinements since ODS2.1 are summarized here and further described in this document.

Changes to obs4MIPs Global Attributes, DRS, filenames and directory structure

- New global attributes related to provenance: documenting who/where/when data was retrieved ((`<data_set_contributor>`, `<source_data_url>`, `<source_data_retrieval_date>`, `<source_data_notes>`) and how the data was prepared (`<processing_code_location>`).
- Replace use of `<grid_label>` in directory structure with `<nominal_resolution>` (without spaces); retain `<grid_label>` in file name.
- The filename template in ODS2.5 is the same as in ODS2.1
- Expanded possible entries of `<product>`, e.g., including “exploratory_product”
- `<variant_label>` is now used to identify the contributing institution (see below and Table1), and if applicable, the identification of individual members of an observational product provided as an ensemble.
- `<further_info_url>` has been deprecated.

Changes to the data preparation protocol

- Demo's and "recipe" documentation now available along with codes used to prepare existing products
- The institution preparing the obs4MIPs-compliant data is identified in the <variant_label> and can be the institution where the data is curated (same as "institution_id") or a 3rd party institution that is registered with obs4MIPs.
- For a dataset to be "obs4MIPs-compliant" it must:
 - Have the <source_id> registered on the obs4MIPs GH repo along with the other required registered content.
 - Provide codes used to make data obs4MIPs-compliant available on the obs4MIPs GH metadata repo

Three stages of preparing obs4MIPs data are now identified as follows:

An obs4MIPs-compliant dataset is a netCDF file(s) that has been prepared according to all specifications described in this document and on the GH repository mentioned above. This includes submitting the required registered content to the repository and inclusion of the associated processing codes on the GH repository. Anything short of this may be *obs4MIPs-like*, but not obs4MIPs-compliant.

An obs4MIPs product is an obs4MIPs-compliant dataset that has been properly published to the ESGF obs4MIPs project.

A reviewed obs4MIPs product is an obs4MIPs-product that has been assessed by the obs4MIPs Steering Panel and has been assigned indicators as described in Fig 2a of Waliser et al. (2020).

Global attributes

The global attributes are constructed to facilitate organization of the obs4MIPs datasets, and in particular for providing a useful set of options (or facets) for data exploration via the [ESGF metagrid](#). The purpose of the DRS is more behind the scenes than the search functions - but it is critical in defining how data from the CMIP6 and obs4MIPs databases is organized (notably the directory structure). These obs4MIPs data structures are curated by the authors of this document in coordination with an obs4MIPs Steering Panel and the

WIP.

Satisfying these (or the CMIP6) data requirements is facilitated by using the [Climate Model Output Rewriter](#) (CMOR3), currently required for producing obs4MIPs data because it ensures that the necessary metadata for ESGF distributed data searching is included. [Following a recipe](#), a user populates an input table with the entries of the required global attributes, and data is read in (typically via a python script) and output via CMOR3 which automatically produces the DRS and filenames in the preparation of obs4MIPs-compliant data. Once the [CMOR utility PrePare](#) has been generalized to apply for obs4MIPs, the requirement of using CMOR for obs4MIPs may change to “strongly recommended”.

Table 1 contains the list of obs4MIPs global attributes, indicating which are required and which are optional. The values for many of the global attributes must be drawn from special obs4MIPs “controlled vocabularies” (CVs). A CV, in simplest form, is a list of the permitted values that can be assigned to a given global attribute. The lists of permitted values can be found in the [reference CVs for obs4MIPs](#) maintained on a github repository.

Table 1: obs4MIPs global attribute description

with comparison to the original obs4MIPs conventions (ODS1.0, circa 2012, and ODS2.1 circa 2017)

Table color key:

Name or form has been changed relative to CMIP5 and/or ODS1.0

New attribute for obs4MIPs added with ODS2.1

New attribute for obs4MIPs added with ODS2.5

Controlled Vocabulary (CV)

Registered Content (RC)

obs4MIPs global attributes see note 1	description	examples	Introduced	form see note 2	when required?	further information and rationale
activity_id see note 3	activity identifier	only value permitted is “obs4MIPs”	originally as project_id in ODS1.0	CV	always	Renamed more generically, since not all activities are projects; also multiple activities may now be listed separated by single spaces
comment	see note 9	see note 9	ODS1.0	free form	never	No change from original obs4MIPs or CMIP5; CF-convention standard

contact	see note 9	see note 9	ODS1.0	free form	always	Still required with ODS-2.5
conventions	convention version	" CF-1.1 ODS-2.5"	ODS1.0	CV	always	Updated version from ODS-1.0 with a list of conventions separated by single spaces now allowed
creation_date	date file was created	see note 5?	ODS1.0	structured form	always	No change from original obs4MIPs specs (automatic with CMOR3)
data_specs_version	version identifier of obs4MIPs CV's and CMOR tables	2.5	ODS2.1	CV	always	This version number is associated with the version or github "tag" for the obs4MIPs CMOR tables and the CV's. See PCMDI/obs4MIPs-cmor-tables which originate from the CMIP6 tables
external_variables	external cell measures	"areacella", "areacello", "volcella", "volcello", as defined in CMIP6	ODS2.1	CV	rarely needed, but include when 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 obs4MIPs data archive. Use of this attribute is expected to be infrequent in contrast to the recommended inclusion of grid bounds (Appendix 1)
frequency	sampling frequency	"mon"	ODS1.0	CV	always	No change from original obs4MIPs. The current options are given in obs4MIPs_frequency.json

grid	grid	see note 7	ODS2.1	free form	always	Briefly describes output grid characteristics
grid_label	grid identifier	“gn”, “gr1” see note 8	ODS2.1	CV (note 11)	always	Used in file name to distinguish among files when the variable is reported on more than one grid. See obs4MIPs_grid_label.json
history	see note 9	see note 9	ODS1.0	free form	never	No change; CF-convention standard
institution	institution name	“NOAA's National Centers for Environmental Information, Asheville, NC 28801, USA”	ODS1.0	CV with registered content	always	Can be used to identify institute responsible for dataset including a curator role if data is no longer managed by original individual or entity. These entries must be registered in: obs4MIPs_institution_id.json
institution_id Will have ESGF alias with institute_id to be backwards compatible with ODS v1.0	institution identifier	“NCEI”	ODS1.0	CV with registered content	always	Was “institute_id”; name changed to align with CMIP; this string is constructed only using the character set: a-z, A-Z, 0-9, and “-”. These entries must be registered on the Github repo
license	license restrictions	see note 13	ODS2.1	some required text	always	Ensures that anyone using the files has access to the terms of use
nominal_resolution	approximate horizontal resolution	“50 km”, “100 km”, “250 km”, “1x1 degree”. (See Appendix	ODS2.1	CV	always	Added in CMIP6 to provide an indication of approximate output grid resolution. See obs4MIPs_nominal_resolution.json

		2 of CMIP6 specifications)				
dataset_contributor	Identifies if obs4MIPs product was prepared by data curator institution or a “3rd party”	Options: “curator” or “3rd party”	ODS2.5	structured form	always	Identifies whether or not the data curators prepared an obs4MIPs compliant version of if it was done by a “3rd party”
processing_code_location	Direct link to code used to process data via GH hashtag	When preparing data via GH repo a function can be called that identifies the precise code version which can be saved as a global attribute. See provenance info in demo	ODS2.5		always	
source_data_notes	Can be used to provide additional information on the accessibility of source data.		ODS2.5		never	
source_data_retrieval_date			ODS2.5	See Note 5	never	
source_data_url	url location where source data was retrieved (normally a recognized curator of the data)		ODS2.5		always	
product	product type	“observations”, “reanalysis”, “exploratory_	ODS1.0	CV	always	As in original obs4MIPs and CMIP5

		product”, “in_situ”				
realm	realm(s) where variable is defined	“atmos”, “ocean”, “land”, “sealce” “atmosChem”	ODS1.0	CV	always	As in original obs4MIPs. See obs4MIPs_realm.json
references	see note 9	see note 9	ODS1.0	free form	never	No change; CF-convention standard
region	Pre-defined (CF conventions) approximate regions	“north_america”, “global”, “global_land” (multiple entries ok)	ODS2.1	CV (list object)	always	See obs4MIPs_region.json
source (modified form)	full observational dataset name/version	see Appendix 2	ODS2.1	CV (Generated from registered content)	always	More comprehensive than original obs4MIPs and CMIP5. See Tables/obs4MIPs_CV.json and search for source_id
source_id (modified form)	Unique identifier for dataset	“GPCP-2-4-1” See Appendix 2	ODS1.0	CV (Generated from registered content)	always	Edited version of first part of “source” (with forbidden characters like spaces and periods replaced with hyphens); used in constructing the file name. See obs4MIPs_source_id.json for examples
source_label	label used to identify source (independent of source version)	“GPCP” See Appendix 2	ODS2.1	CV with registered content	always	This should be the same as source_id, but without a version number. It will likely be used in faceted searches to get a truncated list of sources (without all the different versions listed)

source_type	Observational class	'satellite_retrieval', 'satellite_blended', "gridded_insitu", "reanalysis"	ODS2.1	CV see Appendix 2	always	See obs4MIPs_source_type.json . Additional entries may be added over time
source_version_number	Numeric version identifier	v1.0, 1.2 ver2.3.1 See Appendix 2	ODS2.1	CV with registered content	always	See obs4MIPs_source_id.json and search for source_version_number
title	see note 4	see note 4	ODS1.0	free form	never	no change; CF-convention standard
tracking_id	unique file identifier (automatically generated by CMOR)	see note 14	ODS2.1	structured form with some CV	always	Form modified to facilitate use by ESGF
variable_id	variable identifier	"tas", "pr", "ua"	ODS1.0	CV	always	Added to direct users and software to the primary variable of interest in the file. The complete list of CMIP6 variable_ids is here
variant_info	description of "3rd party" identifier and if applicable run variant	"Best Estimate", "Sphere of influence = 20km" See note 15	ODS2.1 refined in ODS2.5	free form	as appropriate	Provides a brief description of variant differences. Not needed in most cases, when there is only a "Best Estimate" (BE)
variant_label	"variant" label	"PCMDI" Ensemble examples: "RSS-BE" "RSS-r1"	ODS2.1 refined in ODS2.5	structured form	always	Attribute serving two purposes: 1) a registered institution_id representing where the obs4MIPs product was prepared (either the institute of the data curator or a 3rd party), AND, if applicable, a

		See note 16				<p>concatenated (“-”) 2nd entry to identify a member of an observational ensemble.</p> <p>If there is a default version of an ensemble, it is identified as a ‘best estimate’ or “BE”. The ensemble entry is relevant only when there are multiple estimates of the same source_id (e.g., constructed with alternate processing choices), in which case ensemble member identification could follow CMIP6, e.g., with “BE”, “r1”, “r2”, “r3”, ...</p>
--	--	-----------------------------	--	--	--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Table Notes:

1. Using CMOR, an additional global attribute will automatically be included: cmor_version, but this is not an obs4MIPs required global attribute
2. “CV” means content must be taken from a “controlled vocabulary” defined in coordination with the WIP. “registered content” (RC) is special controlled vocabulary defined by data contributors and monitored by the obs4MIPs Steering Panel. Data contributors can submit RC at (<https://github.com/PCMDI/obs4MIPs-cmor-tables>) and contact obs4MIPs-admin@lnl.gov
3. For backwards compatibility with ODS1,0 project_id will be an alias of activity_id on ESGF.
4. Since some software uses the ‘title’ for default plotting or describing the contents of a file, it can be used to provide a description that is similar or equivalent to the source. A common entry may be the same as the source but without the release_year.
5. creation_date form: YYYY-MM-DDTHH:MM:SSZ (e.g., “2010-03-23T05:56:23Z”)
6. This note was for further_info_url which has been deprecated in ODS2.5
7. 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 = "data regridded to a CMIP6 standard 1x1 degree latlon grid from the native T63 grid using an area-average preserving method"

grid = "data regridded via bilinear interpolation to a 3x3 deg latlon grid from the native atmosphere T63 gaussian grid (64x128 latlon)"

grid = "data regridded to a CMIP6 standard 1x1 degree latlon grid from the native T63 grid using an area-average preserving method"

grid = "data regridded via bilinear interpolation to a 3x3 deg latlon grid from the native atmosphere T63 gaussian grid (64x128 latlon)"

8. Data providers may choose to report their output on one or more grids, or with special care provide an alternate projection. To distinguish between output reported on different grids, a "grid_label" attribute is defined. The original grid should be labeled with "gn". Additional grids should be labeled using the form "gr[i]" where i is a positive integer. If the data is subsequently regridded to a second and third grid, "gr2" and "gr3" could be used to distinguish these grids from "gr1".

9. A description and examples of the contact, comment, history and references global attributes may be found in the [CMIP6 output metadata requirements](#)

13. The wording of the "license" attribute is up to the dataset creator, but it is recommended that you reference one of the "Creative Commons" licenses. Here is some possible text: "Data in this file produced by <Your Centre Name> is licensed under a Creative Commons Attribution- 4.0 International (**CC BY 4.0**) License (<https://creativecommons.org/licenses/>). Use of the data must be acknowledged following guidelines found at <a URL maintained by you>. Further information about this data, including some limitations, can be found via <some URL maintained by you>."

14. tracking_id should be of the form hdl:21.14102/<uuid> (e.g., "hdl:21.14102/02d9e6d5-9467-382e-8f9b-9300a64ac3cd"). The tracking_id should be unique for each file published in ESGF. CMOR automatically generates a tracking_id. You can review the python built-in UUID library documentation [here](#).

15. Except when variant_label="BE", it is recommended that variant_info include information identifying major distinguishing features of a variant, but care should be taken to record correct information.

16. Via [an obs4MIPs institution_id](#), the variant_label identifies where the obs4MIPs-compliant data was prepared. If an obs4MIPs product is being prepared by the original curators of the source data, the institution_id for the dataset will be the same as the institution_id identified in the source_label. If the obs4MIPs product is prepared by a "3rd party" they will typically be different. Also, if the same source_id is being used to represent multiple versions of an observational "ensemble", the institution_id can be combined (via "-") with labels representing subsequent versions. If there is a default or official member, "BE" can be used to identify this best-estimate, with labels representing subsequent versions (derived from different processing choices), with the CMIP nomenclature recommended: 'BE', 'r1', 'r2', 'r3', ..., 'rN'. Examples: "PCMDI", "RSS-BE", "RSS-r1"

Data Reference Syntax (DRS) components:

The DRS is used, for example, in file names, directory structures and in facets of some search tools. The following components are needed for obs4MIPs:

activity_id	(original obs4MIPs "activity")
realm	(original obs4MIPs "activity")
frequency	(original obs4MIPs: "frequency")
product	(original obs4MIPs: "product")
institution_id	(original obs4MIPs: "institute")
source_id	(original obs4MIPs: "model")
source_label	(new in obs4MIPs ODS2.1)
variable_id	(original obs4MIPs: "variable name")
region	(new in obs4MIPs ODS2.1)
grid_label	(new in obs4MIPs ODS2.1)
variant_label	(original obs4MIPs: "ensemble member")
version	(original obs4MIPs: "version number")

File name template:

The obs4MIPs file name must be constructed consistent with the following template.

obs4MIPs file name template =

<variable_id>_<frequency>_<source_id>_<variant_label>_<grid_label>[_<time_range>].nc

e.g. siconc_mon_OSI-SAF-450-a-3-0_PCMDI-BE_gr1_185001-202301.nc

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

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. The format for this segment is the same as in CMIP6 (see Table 2 of the CMIP6 specs document: <http://goo.gl/v1drZl>).

For comparison, here is the CMIP6 file name template:

<variable_id>_<table_id>_<source_id>_<experiment_id>_<member_id>_<grid_label>[_<time_range>].nc

and the legacy obs4MIPs file name template:

<variable>_<instrument>_<processing_level>_<processing version>_<start_date>-<end_date>.nc

Directory structure template:

The obs4MIPs directory structure must be constructed consistent with the following template.

obs4MIPs directory structure =

```
<activity_id>/
  <institution_id>/
    <source_id>/
      <frequency>/
        <variable_id>/
          <nominal_resolution>/
            <version>
```

Note any spaces in nominal_resolution are removed in directory structure.

For comparison, here is the CMIP6 directory structure:

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

and the legacy obs4MIPs directory structure:

```
obs4MIPs/
  observations/
    <realm>/
      <variable_id>/
        <frequency>/
          <grid>/
            <Institution_id>
              <instrument>/
                <version>/
```

Notes:

<version> here refers to the CMOR-assigned version number which has the form “vYYYYMMDD” (e.g., “v20170921”), indicating a representative date for the version was produced for obs4MIPs. For those not using CMOR, the convention must be followed.

Sample file header

The example below was produced from the following demo: <https://github.com/PCMDI/obs4MIPs-cmor-tables/demo>

KEY: **yellow** means "absolutely essential"

```
ncdump rlut_mon_CERES-EBAF-4-2_RSS_gn_200003-202307.nc
```

```
dimensions:
```

```
time = UNLIMITED ; // (281 currently)
lat = 180 ;
lon = 360 ;
bnds = 2 ;
```

```
variables:
```

```
double time(time) ;
    time:bounds = "time_bnds" ;
    time:units = "days since 2000-03-01 00:00:00" ;
    time:calendar = "gregorian" ;
    time:axis = "T" ;
    time:long_name = "time" ;
    time:standard_name = "time" ;
double time_bnds(time, bnds) ;
double lat(lat) ;
    lat:bounds = "lat_bnds" ;
    lat:units = "degrees_north" ;
    lat:axis = "Y" ;
    lat:long_name = "Latitude" ;
    lat:standard_name = "latitude" ;
double lat_bnds(lat, bnds) ;
double lon(lon) ;
    lon:bounds = "lon_bnds" ;
    lon:units = "degrees_east" ;
    lon:axis = "X" ;
    lon:long_name = "Longitude" ;
    lon:standard_name = "longitude" ;
double lon_bnds(lon, bnds) ;
float rlut(time, lat, lon) ;
    rlut:standard_name = "toa_outgoing_longwave_flux" ;
    rlut:long_name = "TOA Outgoing Longwave Radiation" ;
    rlut:comment = "at the top of the atmosphere (to be compared with satellite measurements)" ;
    rlut:units = "W m-2" ;
    rlut:cell_methods = "area: time: mean" ;
    rlut:cell_measures = "area: areacella" ;
    rlut:history = "2023-12-05T22:32:04Z altered by CMOR: replaced missing value flag (-999) and corresponding data with standard missing value (1e+20)." ;
    rlut:missing_value = 1.e+20f ;
    rlut:_FillValue = 1.e+20f ;
    rlut:valid_min = "0.00000" ;
    rlut:valid_max = "400.000" ;
```

```

// global attributes:
:Conventions = "CF-1.7 ODS-2.1" ;
:activity_id = "obs4MIPs" ;
:contact = "RSS (support@remss.com)" ;
:creation_date = "2023-12-05T22:32:05Z" ;
:data_specs_version = "2.1.0" ;
:external_variables = "areacella" ;
:frequency = "mon" ;
:grid = "1x1 degree latitude x longitude" ;
:grid_label = "gn" ;
:history = "2023-12-05T22:32:05Z ; CMOR rewrote data to be consistent with CMIP6, CF-1.7 ODS-2.1 and CF standards." ;
:institution = "NASA-LaRC (Langley Research Center) Hampton, Va" ;
:institution_id = "NASA-LaRC" ;
:mip_era = "CMIP6" ;
:nominal_resolution = "100 km" ;
:obs4MIPs_GH_Commit_ID = "https://github.com/PCMDI/obs4MIPs-cmor-tables/tree/77c8a3e419552f672ee5bc8293e00ac3295fb4a8/inputs/RSS
/NASA-LaRC" ;
:originData_URL = "https://ceres.larc.nasa.gov/data/" ;
:originData_retrieved = "20230209; Andrew I. Manaster @ RSS" ;
:product = "observations" ;
:realm = "atmos" ;
:references = "doi: 10.1175/JCLI-D-17-0208.1" ;
:region = "global" ;
:source = "CERES-EBAF-4-2 4.2 (2022): CERES EBAF (Energy Balanced and Filled) TOA Fluxes. Monthly Averages" ;
:source_id = "CERES-EBAF-4-2" ;
:source_label = "CERES-EBAF-4-2" ;
:source_type = "satellite_blended" ;
:source_version_number = "4.2" ;
:table_id = "obs4MIPs_Amon" ;
:table_info = "Creation Date:(18 November 2020) MD5:c7a33bd10bfb7342de23bda43175c1e2" ;
:title = "CERES V4.2 prepared for obs4MIPs (ODS-v2.1.0)" ;
:tracking_id = "hdl:21.14102/9987edbc-097c-404b-87bc-0004ad95ea6c" ;
:variable_id = "rlut" ;
:variant_info = "Best Estimate" ;
:variant_label = "RSS" ;
:license = "Data in this file produced by NASA-LaRC are licensed under a Creative Commons Attribution 4.0 International License
(https://creativecommons.org/licenses/by/4.0/). Use of the data must be acknowledged following guidelines found at https://ceres.larc.nasa.gov/.
Further information about this data, including some limitations, can be found via https://ceres.larc.nasa.gov/." ;
:cmor_version = "3.7.3" ;

```

Appendix 1: Coordinate bounds

While climate models have grid cells with boundaries that have exact definitions allowing analysts to precisely compute area integrals averages, they are often not well defined for gridded observational products. However, to objectively compare simulations and observations, the area

associated with each gridded value in an observational product must be estimated. We therefore recommend that the data provider include their own estimates of the latitude-longitude grid cell bounds from which the grid cell areas can be calculated (since they are most familiar with the data). This will help ensure uniformity among different researchers analyzing the data.

Appendix 2: Guidance for defining and registering source information (product description)

We need 7 pieces of information from obs4MIPs data providers to enable search functioning and other services on ESGF. (Note that we are referring here to only the “source” information which collectively is used to identify a dataset; data providers will also need to register their institution, and if they are contributing a new variable, not already defined by CMIP6, they will need to register the variable.). The 7 attributes are listed below, each with 3 examples:

1. **source_name** ("brand name") of the dataset. (See below for further guidance.)

"REMSS PRW"

"GPCP"

"NOAA NCEI AVHRR NDVI"

2. **release_year** (year this version of the data was produced and made available; the year that the data was reformatted for obs4MIPs is irrelevant):

"2017"

"2003"

"2013"

3. **source_description** of the dataset which usually is simply an expansion of the acronyms or abbreviations appearing in source_name

"Remote Sensing Systems precipitable water"

"Global Precipitation Climatology Project"

"NOAA Nat Cent ... AVHRR Normalized difference vegetation index"

4. **source_version_number** of the dataset (following whatever convention the data provider prefers)

"V6.6.0"

"2.3"

"V4.0"

5. **institution_id** (acronym(s) used to identify the institution, group, or consortium responsible for producing the data set)

"RSS"

"UofMD"

"NOAA NCEI"

6. the **region** covered by the dataset (see [obs4MIPs_region.json](#) for options; the smallest appropriate region among the options should be used)

"global"

"global_ocean"

"global_land"

7. the **source_type** (see [obs4MIPs_source_type.json](#) for options)

"gridded_insitu"

"satellite_blended"

"satellite_retrieval"

From the information above, the obs4MIPs team constructs the following 3 items:

8. **source_label** = <source_name>, but substituting "-" for certain forbidden characters (including ".", "_", "(", ")", "/", and " ").

"REMSS-PRW"

"GPCP"

"NOAA-NCEI-AVHRR-NDVI"

9. **source_id** = <source_label>-<source_version_number> but substituting "-" for certain forbidden characters (including ".", "_", "(", ")", "/", and " ").

"REMSS-PRW-6-6-0"

"GPCP-2-3"

"NOAA-NCEI-AVHRR-NDVI-4-0"

10. **source** = <source_name> <source_version_number> (<release_year>): <source_description>

"REMSS PRW v6.6.0 (2017): Remote Sensing Systems precipitable water"

"GPCP 2.3 (2003): Global Precipitation Climatology Project"

"NOAA NCEI AVHRR NDVI v4.0 (2013): NOAA Nat Cent ... AVHRR Normalized difference vegetation index"

In summary, data providers need to *register content* for the following: 1) source_name, 2) release_year, 3) source_description, 4) source_version_number, 5) institution_id, 6) region and 7) source_type. They can do this by submitting an issue on the [obs4MIPs-cmor-tables github repository](#). From this information the source_label, source_id and source are constructed. CMOR will record as global attributes all items except the first 3: source_name, release_year, and source_description (because they are included in "source").

Notes:

- 1) The *source_name* must be unique across all obs4MIPs datasets. It should be as short as possible because it appears in file names, directory structures, search lists, and the like.
- 2) For an existing dataset that is now being prepared for obs4MIPs, the *source_name* might be an established name associated with the dataset and recognizable by the community (e.g., GPCP, GHCN). For new data sets that do not yet have a recognized brand name, *source_name* might (but is not required to) include an acronym with some indication of: who the data provider is, what instrument the observations are based on, and/or the name of the observed variable. The *source_name* should not be too generic because each separate contribution to obs4MIPs (current or future) must have a unique *source_name*. [If, for example, you are currently contributing a sea surface temperature dataset, but expect later to prepare a surface salinity dataset, you should anticipate that when constructing your *source_name*. In this case you wouldn't want the *source_name* to just be the name of your group. Similarly, you wouldn't want your source name to be just the name of the variable because it is likely other groups will also want to contribute the same variable to obs4MIPs.] Consider what name you would like the dataset to be known by in the years to come.
- 3) If an obs4MIPs contribution includes multiple, related datasets (i.e., more than one variable), coming, for example, from a self-consistent analysis procedure, then all the variables would normally share a common *source_name* and *source_version_number*.
- 4) If a single group plans to prepare multiple *unrelated* datasets, then each dataset should be assigned a different *source_label*. By "unrelated" we mean they would normally be independently assigned "release" (or version) numbers. If these multiple, unrelated datasets each deal with a different variable, then one option would be to include the variable name as part of the *source_label*, but this is not a requirement. The provider might have other ways to distinguish between these datasets.

Note: If a minor problem is found in obs4MIPs datasets (e.g., a mistake in one of the global attributes), the mistake can be corrected without issuing a new "release" with a new version. The data can be republished on ESGF, and ESGF's versioning system will hide the deprecated file(s) (and make it easy for users to determine whether they have the most up-to-date files).

Appendix 3: Document version information

0.1 (23 October, 2016) - Document initiated with CMIP6 analog. *Not public*

0.2 (23 October - 22 February 2017) - working on draft document in relation to CMIP6, CMIP5 and original obs4MIPs specs

0.3 (March 01 - April 05 2017) - Added “region” as global attribute, finessing Table 1 footnotes

0.4 (April 25, 2017) - Draft made public on obs4MIPs CoG site enabling broader feedback

[2.0 released \(June 30, 2017\)](#) on obs4MIPs CoG site

[2.1 \(June 25, 2017\)](#) - update made public on obs4MIPs CoG site

2.5 Final DRAFT expected early in 2024

Note: As of November 2024, CoG is no longer available. It has been superseded by metagrid