obs4MIPs Global Attributes Requirements

As the purpose of obs4MIPs is to make available observational data that can be readily compared with output from major model intercomparison projects (e.g., CMIP5), the data requirements are similar to those specified for the CMIP5 models. The CMIP5 documents most relevant to obs4MIPs are:

1. CMIP5 Model Output Format and Metadata Requirements
2. CMOR - Climate Model Output Rewriter Software
3. standard_output.xls

and may be found at http://pcmdi-cmip.llnl.gov/cmip5/output_req.html .

Document 1 describes the data format and metadata requirements for CMIP5. The specifications appearing in that document also apply to obs4MIPs data, but modified as noted in the section below. Document 2 describes the CMOR software package that can be used to meet the obs4MIPs requirements. It is strongly recommended, but not required, that CMOR be used to produce obs4MIPs datasets. Document 3 describes the requested output from CMIP5 model runs, and obs4MIPs focuses on providing comparable observational datasets. The CMIP5 variable names and other information appearing in document 3 should also be used by obs4MIPs. When preparing an obs4MIPs dataset, it is the data provider's responsibility to note any deviations from the corresponding CMIP5 dataset description in document 3. In particular it is possible that the cell_methods and/or cell_measures attributes that apply to model output might not carry over to observations. In such cases, these attributes should be appropriately modified in the netCDF files.

obs4MIPs Modifications to CMIP5 Global Attribute Requirements:

The output requirements for obs4MIPs are the same as for CMIP5 except that certain global attributes required for CMIP5 are inappropriate for obs4MIPs and should be omitted. Similarly certain additional global attributes are needed to better describe obs4MIPs datasets and are required. The user is free to define additional global attributes thought to be useful. Below we summarize the global attributes used for CMIP5 and obs4MIPs. **Green** font is used to indicate global attributes not required by CMIP5 specifications. **Red** font indicates attributes that are required by CMIP5 but not for observations.

There are references to "CMOR" and "CMOR tables", which should be helpful to users who rely on the CMOR software package to write data in conformance with the obs4MIPs requirements.

Note that one of the data requirements is that only a single (primary) variable be written per file so that the global attributes should apply to that one variable.

obs4MIPs required global attributes:

cmor_version: (required only if data has been written using CMOR; otherwise omit) the version of CMOR that wrote the data. Note that this attribute will be automatically written by CMOR.

contact: name and contact information (e.g., email, address, phone number) of person who should be contacted for more information about the data. (If using CMOR, pass this to cmor_dataset.)

Conventions = 'CF-1.6' (If using CMOR, supply this via a CMOR input table.)

creation_date: a character string representation of the date when the file was created in the format: "*YYYY-MM-DD*-T*HH*:*MM*:*SS*Z" with replacement of all but= "T" and "Z" by the obvious date or time indicator (e.g., "2010-03-23-T05:56:23Z"). (If using CMOR, this attribute will be automatically written.)

data_structure: a character string indicating the internal organization of the data with currently allowed values of "grid", "station", "trajectory", or "swath". The "structure" here generally describes the horizontal structure and in all cases data may also be functions, for example, of a vertical coordinate and/or time. (If using CMOR pass this in a call to cmor_set_cur_dataset_attribute.)

frequency: a character string indicating the interval between individual timesamples in the dataset. The following are the only options used in CMIP5: "yr", "mon", "day", "6hr", "3hr", "subhr" (sampling frequency less than an hour), "monClim" (climatological monthly mean) or "fx" (fixed, i.e., timeindependent). (If using CMOR, supply this via a CMOR input table.)

institute_id: a short acronym describing "institution" (e.g., 'GFDL'). In some cases it might be appropriate to simply use the acronym of the funding agency (e.g., 'USDOE'). (If using CMOR, pass this to cmor_dataset.)

institution: the institution that generated the data [e.g., 'GFDL (Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA']. This identifier should be consistent, but more self-explanatory than the institute_id. In some cases it might be appropriate to simply use the name of the funding agency [e.g., 'USDOE (U.S. Department of Energy)']. (If using CMOR, pass this to cmor_dataset.)

mip_specs: a space-separated list indicating which model intercomparison project(s)' output specifications have been followed. For example, a dataset that is meant to mimic CMIP5 model output would be assigned the value "CMIP5". (If using CMOR pass this in a call to cmor_set_cur_dataset_attribute.)

product = "observations" or "reanalysis", which indicates what type of data you are writing. "Reanalysis" is reserved for datasets produced by an atmospheric or oceanic forecast model constrained by observations. (If using CMOR, supply this via a CMOR input table.)

project_id = "obs4MIPs" (If using CMOR, supply this via a CMOR input table.)

realm: a character string that indicates the portion of the earth system for which the variable is particularly relevant. For CMIP5, permitted values are: "atmos", "ocean", "land", "landIce", "seaIce", "aerosol" "atmosChem", or "ocnBgchem" (ocean biogeochemical). Note that sometimes a variable will be equally (or almost equally relevant) to two or more "realms", in which case a space-separated list of realms is acceptable, with the first item being designated the primary "realm". (If using CMOR pass this in a call to cmor_set_cur_dataset_attribute.)

references: a character string containing a list of published or web-based references that describe the data or the methods used to produce it. Typically, the user should provide references

describing the model formulation here. (If using CMOR, pass this to cmor_dataset. Note that this attribute was optional for CMIP5.)

source: a character string fully identifying the observational product and version. The first portion of the string should be a copy of the global attribute “source_id”, and it is often helpful to include the year when the version of the data was released (e.g., “ARMBE-Atm-1.0 2010 Atmospheric Radiation Measurement Program Best Estimates, atmospheric state profiles, Version 1.0”). (If using CMOR, pass this to cmor_dataset.)

source_id: a character string containing an acronym that most users would associate with the data product, which should usually include a version number (e.g., ARMBE-ATM-1.0). (If using CMOR pass this in a call to cmor_set_cur_dataset_attribute.)

table_id: a character string identifying the CMOR table where this variable appears. This string should be of the form “Table <table name>” followed by “(<date of table>”, where the date is the date of the requested output table (e.g., “table_id=Table obsSites-ARMBE (10 June 2010)”). (If using CMOR, supply this via a CMOR input table.)

source_type: a character string indicating the intrinsic nature of the data, with currently allowed values of ‘satellite_retrieval’, ‘satellite_merged’, ‘in-situ’, ‘ground_retrieval’, or ‘reanalysis’. We anticipate this might be used in future MIPs with values like ‘AGCM’, ‘OGCM’, ‘AOGCM’, ‘EMIC’, ‘RCM’, or ‘ESM’. (If using CMOR pass this in a call to cmor_set_cur_dataset_attribute.)

tracking_id: a character string that is almost certainly unique to this file and must be generated using the OSSP utility which supports a number of different DCE 1.1 variant UUID options. For CMIP5 version 4 (random number based) is required. Download the software from http://www.ossp.org/pkg/lib/uuid/. The tracking_id might look something like: 02d9e6d5-9467-382e-8f9b-9300a64ac3cd. (If using CMOR, this attribute will be automatically generated and written.)

obs4MIPs optional global attributes

comment: a character string containing additional information about the data or methods used to produce it. (If using CMOR, pass this to cmor_dataset.)

history: a character string containing an audit trail for modifications to the original data. Each modification is typically preceded by a "timestamp". (If using CMOR, pass this to cmor_dataset.)

title: A description of the data found in the file. (If using CMOR, this attribute will be automatically generated and written.)

location: a character string containing a very short name identifying the location of the measurement site (e.g., “Barrow”). The sites of most interest to CMIP5 can be found listed at http://pcmdicmip.llnl.gov/cmip5/docs/pointlocations.txt . (If using CMOR pass this in a call to cmor_set_cur_dataset_attribute.)

Optional global attributes that should be eliminated as soon as practical from obs4MIPs datasets.

experiment_id = ‘obs’. If using CMOR, the “experiment_id” attribute is required but should be removed from the CMOR-produced file through postprocessing.

model_id = ‘Obs-’<source_id>, where <source_id> is defined above (e.g., “Obs-ARMBE-ATM-1.0”). If using CMOR, remove from the list of required_global_attributes in CMOR table, and don’t pass to cmor_dataset.

modeling_realm = <realm>, where <realm> is defined above as the “realm” attribute. If using CMOR, the “modeling_realm” attribute is always written, but should be removed from the CMOR-produced file through postprocessing.

CMIP5 global attributes that should ***not*** appear as metadata in obs4MIPs datasets

branch_time: If using CMOR, remove from the list of required_global_attributes in CMOR table, and don’t pass to cmor_dataset.

experiment: If using CMOR, the “experiment” attribute is always written but should be removed from the CMOR-produced file through post-processing.

forcing: If using CMOR, remove from the list of required_global_attributes in CMOR table, and don’t pass to cmor_dataset.

initialization_method: If using CMOR, the “initialization_method” attribute is always written but should be removed from the CMOR-produced file through post-processing.

parent_experiment_id: If using CMOR, remove from the list of required_global_attributes in CMOR table, and don’t pass to cmor_dataset.

parent_experiment_rip: If using CMOR, remove from the list of required_global_attributes in CMOR table, and don’t pass to cmor_dataset.

physics_version: If using CMOR, the “physics_version” attribute is always written but should be removed from the CMOR-produced file through postprocessing.

realization: If using CMOR, the “realization” attribute is always written but should be removed from the CMOR-produced file through post-processing.