A comparison of data and global metadata standards for Obs4MIPs and CCI

Document completed by Debbie Clifford (U Reading; CLIPC) with contributions from Owen Embury (U Reading; SST CCI) and Mike Grant (PML; OC CCI) as part of CLIPC WP5, task 1.

It is a goal of the CLIPC project that datasets from ESA's Climate Change Initiative (CCI) should be accessible via the Earth System Grid Federation (ESGF). ESGF was originally aimed at sharing climate model output, but observational data from the Obs4MIPS project is now being shared via ESG. The Obs4MIPS project has specified a suite of standard metadata fields to describe their observational datasets in terms of the CMIP5 requirements. This report describes how well the CCI metadata standard matches up with the Obs4MIPS metadata requirements.

For SST, as well as being compliant with the CCI data standards, the products are also specified based on the Group for High Resolution SST (GHRSST) Data Specification version 2.0.

Note that this document was completed in early 2015 based on version 1.0 of ocean colour (OC) and sea surface temperature (SST) data. Later versions, particularly in phase 2 of the CCI programme, are likely to have minor changes in some fields.

Obs4MIPS dataset requirements

- General requirements
  - Output file must be written in netCDF3 – CCI-SST and CCI-OC are written in netCDF4 "classic format" i.e. using the chunking/compression from netCDF4, but still using the netCDF3 data model. NetCDF4-classic files can be read by linking to the new library, with no source code modifications. Without the transparent compression, the collection of these files would be unmanageably large or, if externally compressed, difficult to use.
  - Each output file must contain a time series of only one physical variable – CCI-SST and CCI-OC files contain single time steps and multiple SST fields
- Coordinate variables
  - Coordinate variables must be written as double precision floating point – CCI-SST and CCI-OC use single precision
  - Time variable must be stored in units of days since a fixed date – CCI-SST uses seconds since a fixed date. CCI-OC is compliant.
- Physical variables
  - Must be stored in a regular gridded array, with the example given of air temperature as [time, height, latitude, longitude]. CCI-SST currently provide L3 and L4 data as a regular (latitude-longitude) gridded array and would like to add depth, but this would not be compliant with GHRSST standards. L2 data are on the satellite swath i.e. regular in image space [time, ni, nj]. CCI-OC is compliant (gridded lat-long L3).

- o Value of missing data must be 1E20 – for CCI-SST the missing data value depends on the data type (i.e. whether the data are stored as floats, integers, etc). CCI-OC uses the netcdf default float fill value (9.96921 E 36) for physical variables and zero (0) for empty number-of-observations values (int type).
- o Metadata required for the physical variable are:

Obs4MIPS	CCI-SST	CCI-OC
Units	Used as "units" (i.e. lowercase field name)	Lowercase "units"
long name	Used	Used
standard name	Used	Used where available
cell methods	Not used	Not used
cell measures	Not used	Not used
missing value	Not used	Not used
fill value	Used (_FillValue)	Used (_FillValue)
associated files	Not used	Not used
coordinates	Used for L2 files where dimensions are [time, ni, nj]. Coordinates attribute is not applicable to data stored as [time, lat, lon] as the dimensions are the coordinates	Used where appropriate (sinusoidal projection files) but not applicable to L3 lat-long gridded product, as with CCI-SST.
flag_values	Used	Not used (no flags)
flag_meanings	Used	Not used (no flags)
grid_mapping	Not used	Used.

- Naming convention
  - o Obs4MIPS filenames must not contain underscores within the strings represented within angle brackets. CCI-SST allow underscored but not hyphens. CCI-OC has no restrictions beyond those in the CCI filename convention.
  - o Obs4MIPS satellite datasets:
    <variable>_<instrument>_<processing_level>_<processing_version>_<start_date>-<end_date>.ncCCI-SST:

<datetime>-ESACCI-<level>-<producttype>-<instrument>-<extrainfo>-v2.0-fv<version>.nc

  CCI-OC:
  Example: ESACCI-OC-L3S-OC_PRODUCTS-MERGED-1M_MONTHLY_4km_GEO_PML_OC4v6_QAA-200103-fv1.0.nc
  Convention: ESACCI-OC-<level>-<producttype>-MERGED-<extrainfo>-<datestamp>-fv<VERSION>.nc

Global metadata attribute comparison

Table 1: Global attributes required by both Obs4MIPS and CMIP5

Obs4MIPS metadata attribute	Description	Format or exampleCCI-SST	OC-CCI	
contact	Contact information		creator_name, creator_email, creator_url, publisher_email	creator_name, creator_email, creator_url
Conventions	Which version of Climate and Forecast (CF) conventions used	'CF-1.6'	Same	Same
creation_date	When the file was created	YYYY-MM-DD-T*HH:MM:SSZ*	date_created	date_created
frequency	interval between time samples within dataset	e.g. "yr", "day"	Not applicable – only ever one time per file (either one orbit or one day, depending on processing level) Note that the field time_coverage_dur ation contains the difference between start_time and stop_time. This is not an equivalent to frequency, but may be considered related.	Follows CCI metadata convention: time_coverage_resol ution = "P1M" (or P1D, P1W, etc) time_coverage_durati on = "P1M"
institute_id	Short acronym		Not used	Not used
institution	Long name consistent with institute_id		creator_ processing_ institution	institution
product	type of data	"observations" or "reanalysis"	Not applicable	Not used
project_id		"obs4MIPS"	institution = "ESACCI" (GHRSST	project = "Climate Change Initiative -

			RDAC code specification), project = "ESA Climate Change Initiative SST Project"	European Space Agency" ;
source	character string fully identifying the product and version		Same	Identifies source data. Product and version of the data are in "title", "id" and "product_version"
table_id	string identifying CMOR table where the variable appears		Not applicable	Not applicable
tracking_id	Character string unique to this file	UUID	Same	tracking_id id

Table 2: Global attributes optional for CMIP5 but required by Obs4MIPS

Obs4MIPS metadata attribute	Description	Format/example CCI-SST	equivalentCCI-OC	
cmor_version	Version of CMOR[1] that wrote the data		Not applicable	N/A
Data_structure	internal organization of the data	e.g "grid" , "swath"	cdm_data_type	cdm_data_type
mip_specs			Not applicable	Not applicable
realm	which part of the Earth system is relevant	e.g. "atmos", "land"	Not used	Not used
references	character string of		Same	Same

---

[1]

      CMOR is a software package that can be used to produce obs4MIPS datasets that meet these requirements.

	references describing the data and methods used to produce it			
source_id	acronym/short name users associate with the product		Not used	Not used
source_type	instrinsic nature of the data	e.g. "satellite_merged", "in situ"	Not used	Not used

Table 3: Global attributes optional for both Obs4MIPS and CMIP5

Obs4MIPS metadata attribute	Description	Format/exampleC CI-SST equivalentCCI-OC		
comment	Additional information about the data		Same	Same
history	string containing an audit trail for modifications		Same	Same
title	Description of the data		Same	Same
location	name of measurement site		Not applicable	Not applicable (global product)

Table 4: CCI metadata useful for CLIPC not included in Obs4MIPS

Attribute	Description or example
file_quality_level	A code value: 0 = unknown quality 1 = extremely suspect (frequent problems, e.g. with known satellite problems) 2 = suspect (occasional problems, e.g. after launch) 3 = excellent (no known problems)
spatial_resolution	A string describing the approximate resolution of the product. For example, "1.1km at nadir"

platform	Satellite(s) used to create this data file.
sensor	Sensor(s) used to create this data file.
keywords	Typically GCMD Science Keyword, e.g.: "Oceans > Ocean Temperature > Sea Surface Temperature"
keywords_vocabulary	E.g. "NASA Global Change Master Directory (GCMD) Science Keywords"
processing_level	GHRSST definitions are the options: L2P, L3U, L3C, L3S, L4 and GMPE

Important notes:

- Obs4MIPS requirement that only a single variable be written per file, which is not the case for CCI-SST and CCI-OC files.
- Higher level data (e.g. level 4) gets very complicated in terms of source – multiple platforms/instruments etc. CCI-OC data is level 3, but is a merged product.
- Updates to the CCI metadata standard would be achieved via the data standards working group.

References

Obs4MIPS Global Attributes requirements:
https://www.earthsystemcog.org/site_media/projects/obs4mips/obs4MIPsGlobalAttributesRequirements_v1.1.pdf

Obs4MIPS Dataset requirements:
https://www.earthsystemcog.org/site_media/projects/obs4mips/obs4MIPsDatasetRequirements_v1.2.pdf

CCI Phase 1 (SST) Product Specification Document (PSD), SST_CCI-PSD-UKMO-001:

http://www.esa-sst-cci.org/PUG/documents