ecsat

Fermi Avenue
Harwell Campus
Didcot, Oxfordshire OX11 0FD
United Kingdom
T +44 (0)1235 444200
www.esa.int

STANDARDS

CCI Data Standards

Prepared by **ESA Climate Office**

Reference **CCI-PRGM-EOPS-TN-13-0009**
Issue/Revision **2.2**
Date of Issue **19/05/2020**
Status **For Information Only**

European Space Agency
Agence spatiale européenne

CHANGE LOG

Reason for change	Issue Nr.	Revision Number	Date
First issue of Data Standards Requirements (Victoria Bennett & Sarah James)	1	0	20/05/2013
Revision incorporating comments from DSWG (Victoria Bennett & Sarah James)	1	1	24/05/2013
Changing definitions in section 2.1 of source global attribute, platform attribute, and Envisat term in Section 2.1, Appendix A and section 2.7 Product String examples §	1	2	09/03/2015
Rearranged content, added more examples, additional requirements and clarifications ahead of, and following, CCI Colocation March 2018 and to apply during CCI+. Led by Ed Pechorro (ESA), Alison Waterfall (CCI ODP), Victoria Bennett (CCI ODP) and supported by Kevin Marsh (ECMWF) and Cedric Bergeron (ECMWF).	2	0	17/09/2018
Version v2.1 produced based on feedback for v2.0. Led by Alison Waterfall (CCI ODP)	2	1	25/03/2019
Version 2.2. released. Led by Alison Waterfall (CCI ODP), supported by Ed Pechorro (ESA Climate Office).	2	2	19/05/2020

Table of contents:

1 BACKGROUND

The ESA Climate Change Initiative (CCI) is a programme which is generating a set of validated, error-characterised, Essential Climate Variables (ECVs) from existing satellite observations. The CCI Programme was launched in 2009 and initially consisted of thirteen Projects, each addressing a particular ECV, and CMUG, the Climate Modelling User Group. CCI+ began in 2018, building on this earlier work and introducing a further set of ECV's.

This document is based on the "Guidelines for Data Producers – CCI Phase 1"[1], and the subsequent "Data Standards Requirements for Data Producers"[2]. The motivation and content for these documents came from the first CCI collocation meeting in September 2011, subsequent discussions at the CCI collocation meeting, the CCI integration meetings and within the Data Standards Working Group. The start of the CCI+ programme provided the opportunity to revise the data standards outlined in this document in the light of the experience of the earlier phases of the CCI programme, with version 2.1 released in summer 2019. This version provides a subsequent update to the document.

This document provides the minimum requirements to CCI data producers to ensure consistency between output products from the CCI projects. The requirements in this document are intended to apply to all CCI projects in CCI+.

In case of new developments (e.g. due to evolving user requirements, international standards and formats), these requirements may still evolve during the course of the programme. Any updates will only be made after discussion with, and agreement in the CCI Data Engineering Working Group.

The overall aim of these standards is to make the CCI datasets as widely usable as possible, in user communities outside the existing ones. Using common formats and a number of agreed and consistent metadata conventions will facilitate this. Further help and advice on data formatting and support for checking compliance with these standards is available from the CCI Open Data Portal, and CCI Projects are encouraged to liaise with the Portal when developing their data formats.

CCI Sustainability

Version 2.0 of the CCI Data Standards was the first edition of the standards developed with the support of the European Centre for Medium-Range Weather Forecasts (ECMWF). A number of CCI datasets have served as input to, and evolved under, ECMWF implementation of the Copernicus Climate Change Service (C3S). The cooperation between the ESA Climate Office and ECMWF/C3S on the CCI Data Standards, from v2.0 onwards, optmises the sustainability of CCI datasets beyond the CCI programme, and facilitates their use and further development by ECMWF/C3S.

In addition, the CCI Data Standards supports a complementary activity between the ESA Climate Office and ECMWF/C3S on climate data interoperability, following CCI Toolbox Iteration Review Meeting 9 (IRM9) on 20 June 2018 participated by ECMWF/C3S : the development of a proof-of-concept data link between the (ESA) CCI Toolbox API and the (C3S) CDS API.

[1] CCI-PRGM-EOPS-TN-11-0003, issue 4, revision 1, May 2013
[2] CCI-PRGM-EOPS-TN-13-0009, issue 1, revison 2, March 2015

2 PRODUCT GUIDELINES

2.1 Standard products: minimum requirements

All CCI projects shall produce data according to the following minimum requirements for data products:

R-1	CCI projects shall produce data according to the CCI Data Standards Requirements: a. Produced in netCDF-4 format[3] b. Conform with the CF (Climate and Forecasting) convention[4] (currently version 1.8) c. CF standard names used for the main variables d. Include the global attributes listed in section 2.5.1 below

In earlier versions of the CCI Data standards, there was a requirement that data should be produced in netCDF-4 classic format. The recent version of the CF (Climate and Forecasting) convention, now allows some non-classic features, and this restriction has therefore been relaxed to allow the use of new NetCDF 4 data types.

However, datasets shall still only use features which comply with the current version of the CF convention, citing the correct CF and CCI data standards version numbers in the relevant global attributes field (see section 2.5.1 below).

The use of NetCDF 4 groups should only be used after discussion with the CCI data standards team.

For netCDF-4 compression can be invoked.

If netCDF-4 cannot be used (e.g. user tools not yet mature enough), data may be produced in netCDF-3 format.

It is recommended that chunking of the datasets be considered when writing to NetCDF, particularly in case of compression[5]. The appropriate chunking to use will however depend on the intended uses of the product. If specified, it is recommended that the chunk size is an integer fraction of the respective dimension (i.e. that there are no smaller chunks at a border).

2.2 Products for existing user communities

It is understood that existing user communities require data in specific formats that are not in all cases compatible with the requirements in this document (e.g. GeoTiff, shape files).

Projects producing such files in other formats are required to follow these requirements as far as possible in terms of filenaming and metadata for their non-standard data, and to produce a standardised, netCDF format dataset, in addition to their domain specific datasets.

[3] http://www.unidata.ucar.edu/software/netcdf/
[4] http://cfconventions.org/latest.html
[5] https://www.unidata.ucar.edu/blogs/developer/entry/chunking_data_why_it_matters

R-2	Projects who have commitments to produce data in other formats, shall do this **in addition** to the standardized products, and shall ensure these products **comply as much as possible** to the CCI Data Standards (e.g. filenames, metadata)

2.3 Conformance with international standards

The choice of CF and ACDD[6] conventions for CCI data products ensures it will be possible to extract the metadata from the files to a number of standard metadata formats, e.g. ISO19115, FGDC, etc, as required by data discovery systems.

Within the CF-convention there are still areas where there are ongoing discussions, for example, in the representation of swath data under CF (see, for example, ongoing discussions in CF-satellite community http://www.unidata.ucar.edu/mailing_lists/archives/cf-satellite/). CCI data producers who are producing such data products are strongly encouraged to engage with this community to progress the development of CF conventions for satellite swath data, either directly or via the Data Engineering Working Group, which can provide coordination for discussions with the CF-community. Discussions on the cf-conventions generally and requests for new standard names for geophysical variables used in CCI, can be made here: https://github.com/cf-convention/discuss .

R-3	CCI Data Producers shall engage with the CF community to help develop the standards they require for satellite data

The INSPIRE (Infrastructure for Spatial Information in the European Community) directive requires INSPIRE compliant discovery metadata to be generated for geospatial datasets that are within scope of the directive, which covers a broad range of environmental themes. All data products that are held in the CCI Open Data Portal will have INSPIRE compliant metadata records created by the Data Portal team[7] (see, for example:
https://csw-cci.ceda.ac.uk/geonetwork/srv/eng/catalog.search;jsessionid=3F06A64C519329DA3DAB303F8AB0D066#/metadata/3628cb2fdba443588155e15dee8e5352).

R-4	The CCI projects shall ensure that INSPIRE compliant metadata records are created for each dataset held in the CCI Open Data Portal, by providing the required information to the CCI Open Data Portal Team.

2.4 CCI Ontology

To ensure consistency where the same global attributes apply to several ECVs, and to avoid different terms being used for the same thing, relevant terms have been gathered in an ontology, defining the CCI entities and the relationship between those entities, including :

[6]Attribute Convention for Data Discovery (http://wiki.esipfed.org/index.php/Attribute_Convention_for_Data_Discovery)

[7] In exceptional cases products are held in external sites with agreement from ESA

- project
- platform
- sensor
- institution

The CCI Ontology is available online on the ESA CCI website at cci.esa.int/working-groups.

R-5	CCI Data Producers shall use terms from the CCI vocabulary tables in the netCDF global attributes, or if terms are missing they shall request that they are added to the tables.

2.5 File contents

2.5.1 Global Attributes

The following section outlines which global attributes shall be included in the files. Example content for these attributes are given in Appendix A. For consistency, CCI output netCDF files shall include these global attributes to provide usage information about the data:

a. title (succinct description of the dataset)
b. institution (where the data was produced, use names from CCI common vocabulary)
c. source (original data source(s), e.g. MERIS RR L1B version 4.02) Multiple source datasets and ancillary datasets used, with their DOI if available, as a free-text, comma-separated list)
d. history (processing history of dataset)
e. references (references to algorithm, ATBD, technical note describing dataset)
f. tracking_id (a UUID (Universal Unique Identifier) value)[8]
g. Conventions (the CF Version e.g CF-1.8)
h. product_version (the product version of this data file (see section 2.8))
i. format_version (the CCI data format used e.g. “CCI Data Standards v2.x”)

Discovery metadata allows information about the data to be harvested into catalogues and data federations. For dataset discovery, the netCDF attributes recommended in http://wiki.esipfed.org/index.php/Attribute_Convention_for_Data_Discovery shall be included in addition to the attributes listed above:

j. summary (a paragraph describing the dataset)
k. keywords (a comma separated list of key words and phrases)
l. id

[8] The unique tracking ID allows files to be referenced, and linked up to processing description, input data, documentation, etc. The software to generate the UUID can be downloaded from : http://www.ossp.org/pkg/lib/uuid/ and CCI shall use version 4 (random number based) for consistency with CMIP5

m. naming authority (the combination of the naming authority and the id should be a globally unique identifier for the dataset)

n. keywords_vocabulary (if you are following a guideline for the words/phrases in your "keywords" attribute, put the name of that guideline here)

o. cdm_data_type (the THREDDS data type appropriate for this dataset)

p. comment (miscellaneous information about the data)

q. date_created (the date on which the data was created)

r. creator_name

s. creator_url

t. creator_email

u. project (the scientific project that produced the data: "Climate Change Initiative – European Space Agency")

v. geospatial_lat_min (decimal degrees north, range -90 to +90)

w. geospatial_lat_max (decimal degrees north, range -90 to +90)

x. geospatial_lon_min (decimal degrees east, range -180 to +180)

y. geospatial_lon_max (decimal degrees east, range -180 to +180)

z. geospatial_vertical_min (assumed to be in metres above ground unless geospatial_vertical_units attribute defined otherwise)

aa. geospatial_vertical_max (assumed to be in metres above ground unless geospatial_vertical_units attribute defined otherwise)

bb. time_coverage_start (format yyyymmddThhmmssZ)

cc. time_coverage_end (format yyyymmddThhmmssZ)

dd. time_coverage_duration (should be an ISO8601 duration string)

ee. time_coverage_resolution (should be an ISO8601 duration string). For L2 data on the original satellite sampling it is acceptable to use 'satellite_orbit_frequency'

ff. standard_name_vocabulary (the name of the controlled vocabulary from which variable standard names are taken e.g. 'CF Standard Name Table v71'.)

gg. license (describe the restrictions to data access and distribution)

The following CCI project specific attributes shall be included in the files:

hh. platform (satellite name, e.g. Envisat; use names from CCI common vocabulary list. Separated by commas if more than one, and use angled brackets for a platform series, eg. 'Envisat, NOAA-<12,14,16,17,18>, Metop-A')

ii. sensor (sensor name, e.g. AATSR, use names from CCI common vocabulary. Separated by commas if more than one)

jj. spatial_resolution (a free-text string describing the approximate resolution of the product. For example, "1.1km at nadir". This is intended to provide a useful indication to the user, so if more than one resolution is relevant e.g. the grid resolution and the data resolution, then both can be included.)

kk. key_variables (a comma separated list of the key primary variables in the file i.e. those that have been scientifically validated, and are appropriate for display in the CCI Open Data Portal and CCI Toolbox. These should be identified using the variable id's in the file)

For gridded (level3/level4) data on a regular lat/lon grid, the following attributes shall be included in the files (not compulsory for level2 data, or data on other grids, e.g. polar stereographic) :

ll. geospatial_lat_units

mm. geospatial_lon_units

nn. geospatial_lon_resolution

oo. geospatial_lat_resolution

Projects are free to include any number of additional attributes which are specific to their domain or user community, or for compatability with other external initiatives e.g. Copernicus C3S.

2.5.2 Variables

Variables should be stored in the NetCDF file in conformance with the CF-convention (http://cfconventions.org/) and using the standard name attribute where appropriate. Where an existing standard name exists this should be included as an attribute on the variable. Where a standard name does not exist, or has not yet been accepted by CF, the standard name attribute should not be included for that particular variable. Data producers are strongly encouraged to submit new standard names for inclusion in CF as appropriate. Names can be proposed and discussed directly here: https://github.com/cf-convention/discuss. The CCI Open Data portal team can also provide advice on the process.

In addition the following should be considered:

a. The key primary variables in the file (i.e. those that are scientifically validated, and for instance, are appropriate for visualisation through the CCI Portal and Toolbox), shall be identified by the use of the 'key_variables' global attribute. This should be in the form of a comma separated list.

b. Ancillary and coordinate variables should also be identified as specified in the CF conventions[4]:

   - Ancillary variables such as uncertainty or quality flags should be identified by the 'ancillary_variables' attribute of the related primary variable. Ancillary variables that are required to correctly interpret a key primary variable must be identified in this way. (Section 3.4 of CF-1.7 conventions document)
   - Coordinate variables are identified as those with the same name as the corresponding dimensions, whilst related auxiliary coordinate variables can be identified using the 'coordinates' attribute of a variable (Section 4 and 5 of CF-1.7 conventions document)

c. For each variable, the attributes 'valid_range' (or alternatively 'valid_min','valid_max') and 'actual_range' shall be provided, if appropriate.

   - The 'valid_range' should define the expected valid range of the variable. Under the CF-convention, values outside this range are treated as missing data.
   - The 'actual_range' provides the actual range of the data within the file, within the limits of the valid range.

d. Where quality flags are included, it is recommended that a value of '0' (the number zero) or no bits set is taken to indicate good quality data.

e. For variables with a defined colour table (e.g. for land cover classes), this colour mapping information shall be included in the NetCDF file.

In addition, to allow the datasets to be best displayed through the CCI Open Data Portal and CCI Toolbox, display information are requested to be provided for the variables e.g. a recommended colour bar and display min/max. Such information could be included in the comment attribute of the respective variables, or alternatively could be provided outside of the data file, directly to the CCI Open Data Portal.

R-6	The key primary variables in the file and their related ancillary variables (e.g. uncertainty) shall be identified, and the range of their expected values shall be indicated.

Example of including variables and corresponding attributes in the file :

```
       float variable1(time, lat, lon) ;
                   variable1:long_name = "a_longer_descriptive_name_of variable" ;
                   variable1:standard_name = "CF_standard_name_here (if appropriate)" ;
                   variable1:units = "unit from UDUNITS" ;
                   variable1:valid_range = 0f,10f ;
                   variable1:actual_range = 1f,9f ;
                   variable1:anciliary_variables = 'variable1_uncertainty variable1_flag' ;
        float variable1_uncertainty(time, lat, lon) ;
              variable1:long_name = " uncertainty associated with variable1" ;
                   variable1:units = "unit from UDUNITS" ;
           byte variable1_flag(time,lat,lon) ;
                   variable1_flag:long_name = "example flag associated with variable1" ;
                   variable1_flag:standard_name = "status_flag" ;
                   variable1_flag:_FillValue = -128b ;
                   variable1_flag:valid_range = 0b,2b ;
                   variable1_flag:flag_values = 0b,1b,2b ;
                   variable1_flag:flag_meanings = "meaning1 meaning2 meaning3" ;

   //global attributes:
                   :key_variables = "variable1" ;

NB. in this example if variable1_flag is a quality flag then 0b should correspond to the best quality
data.
```

2.5.3 Gridded data – coordinate variables

Data products from the CCI which are most likely to be used *outside* the community of EO (Earth Observation) specialists are the global gridded (level3, level4) datasets. To achieve consistency across these ECV products, and enable them to be used and understood by a wide variety of users and tools, including the CCI Open Data Portal and CCI Toolbox, it is important to ensure consistent use of dimensions for variables within the files.

R-7	For consistency across CCI gridded products, variables shall have, as a minimum, the following dimensions: time, latitude, longitude (or alternative horizontal grid)

1. The dimensions time, latitude and longitude of a netCDF variable shall have an associated time, latitude and longitude coordinate variable
    - A coordinate variable is defined in the cf-convention as a one dimensional variable with the same name as its dimensions (e.g. time(time), lat(lat),lon(lon), and it is defined as a numeric data type with values that are ordered monotonically. Missing values are not allowed in coordinate variables.
2. It is recommended that latitude and longitude values from -180 to 180 are used.
3. All products shall have a time dimension, even if there is only a single timestep within an individual file. This is recommended to have be of dimension = 'UNLIMITED'. The corresponding time variable should be an array and not a scalar variable. This allows users to aggregate the data files along the time dimension and facilitates display of the data within the CCI Open Data Portal.
4. Horizontal grid dimensions shall also be included.
    - If coordinates other than latitude and longitude are used for the horizontal grid, then, following the cf-conventions, it is required that the true latitude and longitude coordinates be supplied via the coordinates attribute.
    - The dimensions of any latitude or longitude auxiliary coordinate variable should be the horizontal grid dimensions only (e.g. lat(nx,ny))If it is desired to describe the mapping between the given coordinate variables and the true latitude and longitude coordinates, the attribute grid_mapping should be used to supply this description. For more information and examples see section 5.6 of CF-1.8.
5. For each coordinate variable in a gridded data product, the boundaries of the grid shall be described via the use of boundary variables, as described in section 7.1 of the CF-1.8 conventions document. The boundary variable is linked to the corresponding coordinate variable by the use of the attribute 'bounds', which contains the name of the boundary variable. (see example below)

The GHRSST project (Group for High Resolution Sea Surface Temperature) has developed conventions for non-regular latitude/longitude grids (swath) which have been adopted in the SST community – details and examples are given in the GHRSST Data Specification document[9].

Examples can be found in the CF-conventions document and the CF Unidata NetCDF Users' Guide[10], and a simple example of CDL for CCI gridded lat/lon data is given here:

[9] See https://www.ghrsst.org/documents/q/category/gds-documents/
[10] https://www.unidata.ucar.edu/software/netcdf/documentation/NUG/

```
netcdf CCI_example_grid {
dimensions:
       time = UNLIMITED ;
       lat =  1200 ;
       lon =  1200 ;
        nv = 2 ; //number of vertices (to define the grid boundaries)
variables:
        float lat(lat) ;
                lat:standard_name = "latitude" ;
                lat:units = "degrees_north" ;
                lat:bounds = "lat_bnds" ;
        float lon(lon) ;
                lon:standard_name = "longitude" ;
                lon:units = "degrees_east" ;
                lon:bounds = "lon_bnds" ;
        double time(time) ;
                time:standard_name = "time" ;
                time:units = "seconds since 1970-01-01 00:00:00 0:00" ;
                time:bounds = "time_bnds" ;
        float lat_bnds(lat,nv)
        float lon_bnds(lon,nv)
        double time_bnds(time,nv)
        float ecv_variable(time, lat, lon) ;
                ecv_variable:long_name = "a_longer_descriptive_name_of_the_variable" ;
                ecv_variable:standard_name = "CF_standard_name_here" ;
                    ecv_variable:units = "unit from UDUNITS" ;
                    ecv_variable:valid_range = 0f,10f ;
                    ecv_variable:actual_range = 1f,9f ;
```

2.6 Directory structure

A common CCI directory structure has been defined.

R-8	CCI Data Producers shall use the common directory structure for all output data made available to users

If projects need to use a different structure (e.g. due to conflicting requirements of existing user communities), it shall be possible to map to the CCI common structure using symbolic links. The common directory structure is based on suggestions and existing conventions used by CCI teams and is arranged as follows:

/<archive_root>/<type>/<version>/<time>/

where each <…> may itself be complex, e.g.

<**time**> may be <year>/<month> or <year>/<day-in-year> or <year>/<month>/<day>
<**archive-root**> is /esacci/<cci_project>/
<**cci_project**> is e.g. <sst> or <aerosol>
<**type**> will be different for each ECV, but needs to be defined, and consistent within an ECV
<**version**> uniquely identifies the version (see Section 2.8)

2.7 Filenaming

A common CCI filenaming convention has been developed to simplify access to the project data across the different disciplines involved.

R-9	CCI Data Producers shall use the CCI filenaming convention for all output data made available to users

Overview of CCI Filename Convention

Note that there are two forms that the filename can take:

1. It can begin with the date and time information (which will be digits). This form is compatible with the GHRSST filenaming convention used by Sea Surface Temperature scientists. https://www.ghrsst.org/wp-content/uploads/2016/10/GDS20r5.pdf

   Or

2. It can begin with the string 'ESACCI' in which case the date and time information will be in the field immediately before the version information.

Overview form 1

<Indicative Date>[<Indicative Time>]-ESACCI-<Processing Level>_<CCI Project>-<Data Type>-<Product String>[-<Additional Segregator>][-v<GDS version>]-fv<File version>.nc

Overview form 2

ESACCI-<CCI Project>-<Processing Level>-<Data Type>-<Product String>[-<Additional Segregator>]-<IndicativeDate>[<Indicative Time>]-fv<File version>.nc

The fields in the filename convention are:

<Indicative Date>

The identifying date for this data set.
Format is YYYY[MM[DD]], where YYYY is the four digit year, MM is the two digit month from 01 to 12 and DD is the two digit day of the month from 01 to 31.

The date used should best represent the observation date for the data set. It can be a year, a year and a month or a year and a month and a day.

<Indicative Time>

The identifying time for this data set in UTC.

Format is [HH[MM[SS]]] where HH is the two digit hour from 00 to 23, MM is the two digit minute from 00 to 59 and SS is the two digit second from 00 to 59.

ESACCI

This corresponds to RDAC (Regional Data Assembly Centre) in the GHRSST filenaming convention.

<Processing Level>

Level	<Processing Level> Code	Description	Based on Source
Level 0	L0	Unprocessed instrument and payload data at full resolution. CCI does not make recommendations regarding formats or content for data at this processing level.	GHRSST
Level 1A	L1A	Reconstructed unprocessed instrument data at full resolution, time referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters, computed and appended, but not applied, to L0 data.	GHRSST
Level 1B	L1B	Level 1A data that have been processed to sensor units.	GHRSST
Level 1C	L1C	Level 1b data that have been further processed, e.g. by correcting radiances or by mapping onto a spatial grid, prior to deriving geophysical variables from the data.	SMOS data products definition and ESACCI discussions
Level 2	L2	Retrieved environmental variables at the same resolution and location as the level 1 source	CEOS interoperability handbook[11]
Level 2 Pre-processed	L2P	Geophysical variables derived from Level 1 source data at the same resolution and location as the level 1 data, typically in a satellite projection with geographic information. These data form the fundamental basis for higher level CCI products.	GHRSST

[11]http://ceos.org/document_management/Working_Groups/WGISS/Documents/WGISS_CEOS-Interoperability-Handbook_Feb2008.pdf

Level	<Processing Level> Code	Description	Based on Source
Level 3	L3	Level 2 variables mapped on a defined grid with reduced requirements for ancillary data. Three types of L3 products are defined:	GHRSST
	L3U	Uncollated (L3U): L2 data granules remapped to a space grid without combining any observations from overlapping orbits.	
	L3C	Collated (L3C): Observations combined from a single instrument into a space-time grid.	
	L3S	Super-collated (L3S): observations combined from multiple instruments into a space-time grid.	
Level 4	L4	Data sets created from the analysis of lower level data that result in gridded, gap-free products.	GHRSST
Indicator	IND	Indicators derived from satellite data.	ESACCI

<CCI Project>

This corresponds to GHRSST in the GHRSST filenaming convention.

CCI Project	<CCI Project>
Aerosol	AEROSOL
Antarctic Ice Sheet	AIS[12]
Cloud	CLOUD
Fire	FIRE
Greenhouse Gases	GHG
Glaciers	GLACIERS

Ice Sheets Greenland	GIS[12]
Land Cover	LC
Ocean Colour	OC
Ozone	OZONE
Sea Ice	SEAICE
Sea Level	SEALEVEL
Sea Surface Temperature	GHRSST
Soil Moisture	SOILMOISTURE
Biomass	BIOMASS
High Resolution Land Cover	HRLC
Lakes	LAKES
Land Surface Temperature	LST
Permafrost	PERMAFROST
Salinity	SEASURFACESALINITY
Sea State	SEASTATE
Snow	SNOW
Water Vapour	WATERVAPOUR

Further CCI projects will be added in as the programme continues.

<Data Type>

This corresponds to SST Type in the GHRSST convention and shall contain a short term describing the main data type in the data set. Examples from Phase 1 are given in the following list. Where new

[12] This was formerly part of the Icesheets project and could also have <CCI Project> = ICESHEETS

or different data types are produced in the CCI programme, the new term shall be added to the list by informing the CCI Open Data Portal.

CCI Project	**Parameter**	**<Data Type>**
Aerosol	Aerosol optical depth	AOD
	Absorbing aerosol index	AAI
	Stratospheric aerosol extinction profile	AEX
	Aerosol type	ATY
	Multiple aerosol products	AER_PRODUCTS
Biomass	Above-ground biomass	AGB
Cloud	Cloud cover	CFC
	Cloud top pressure	CTP
	Cloud top height	CTH
	Cloud top temperature	CTT
	Cloud optical thickness	COT
	Cloud effective radius	CER
	Cloud liquid water path	LWP
	Cloud ice water path	IWP
	Joint cloud physical properties	JCH
	Multiple cloud products	CLD_PRODUCTS
Fire	Burned area	BA
GHG	column-averaged dry air mole fraction of CO2	CO2
	column-averaged dry air mole fraction of CH4	CH4
Glaciers	Glacier area	GA

CCI Project	Parameter	<Data Type>
Ice Sheets	Ice sheet surface elevation change	SEC
	Ice sheet velocity	IV
	Glacier calving front location	CFL
	Glacier grounding line location	GLL
Lakes	Multiple lake products	LK_PRODUCTS
Land Cover	Land cover map	Map
	Condition fire (burned area)	BA
	Condition water (water bodies)	WB
	Condition snow	Snow
	Condition normalised difference vegetation index	NDVI
	Condition albedo	Alb
	Leaf Area Index	LAI
	Surface reflectance	SR
Ocean Colour	Multiple products (chla, nlw, IOPs, etc)	OC_PRODUCTS
	Phytoplankton Chlorophyll-a concentration	CHLOR_A
	Normalised water leaving radiance	NLW
	Remote Sensing Reflectance	RRS
	Spectral attenuation coefficient for downwelling irradiance	K_490
	Total absorption	ATOT
	Total backscattering	BB
	Absorption by coloured dissolved organic matter	ADG

CCI Project	Parameter	<Data Type>
	Backscattering by particulate matter	BBP
	Absorption by phytoplankton	APH
Ozone	Ozone total column	TC
	Ozone nadir profile	NP
	Ozone limb profile	LP
Permafrost	Permafrost extent	PFR
	Ground temperature	GTD
	Active layer thickness	ALT
Sea Ice	Sea Ice Concentration	SICONC
	Sea Ice Thickness	SITHICK
	Sea Ice Extent	SIEXTENT
Sea Level	Corrected sea surface height	SSH
	Sea level anomaly	SLA
	Absolute dynamic topography	ADT
	Maps of sea level anomalies	MSLA
	Mean sea level	MSL
	Mean sea level trends	MSLTR
	Mean sea level amplitude and phase	MSLAMPH
Sea State	Significant wave height	SWH
Sea Surface Salinity	Sea surface salinity	SSS
Snow	Snow water equivalent	SWE
SST	Sea surface temperature	SSTint

CCI Project	Parameter	<Data Type>
	Sea surface skin temperature	SSTskin
	Sea surface subskin temperature	SSTsubskin
	Sea water temperature	SSTdepth
	Sea surface foundation temperature	SSTfnd
Soil Moisture	Surface soil moisture volumetric absolutes	SSMV
	Surface soil moisture volumetric anomalies	SSMVA
	Surface soil moisture degree of saturation absolute	SSMS
	Surface soil moisture degree of saturation anomalies	SSMSA
	Soil water index volumetric absolute	SWIV
	Soil water index volumetric anomalies	SWIVA
	Soil water index degree of saturation absolute	SWIS
	Soil water index degree of saturation anomalies	SWISA

<Product String>

Each ECV team shall define the Product Strings they will use for their data and make this information available in their documentation.

The Product String field must not include any hyphens but can include underscores.

Examples:

Example Product String	Description
MERGED	Data from more than one platform and/or sensor
AVHRR_MERGED	Data from AVHRR, more than one platform
AATSR_ENVISAT	Data from AATSR on Envisat
ORAC	Data from Oxford RAL Aerosol and Cloud algorithm

OPERA	Data from OPERA analysis

<Additional Segregator>

This is an optional part of the filename. It must be used if otherwise different data sets would generate the same filename. It can also be used to include in the filename information that doesn't fit elsewhere in the filename convention, but which projects feel is useful for easy identification of different data sets.

Each ECV team shall define the Additional Segregators they will use for their data and make this information available in their documentation.

More than one element may be included, separated by an underscore, not a hyphen.

Examples:

Information	Example <Additional Segregator> element	Description
Composite data	<n><Y|M|D>[<n><y|m|d>]	nY or nM or nD corresponds to the period over which data are aggregated of n number of years or months or days. An optional ny or nm or nd refers to the 'season' covered by the composite data set. For example 15Y7d indicates 15 year composite data covering a specific 7 days of the year. The start date and time of the earliest data included in the composite data product will be given by the <Indicative Date>[<Indicative Time>]
Orbit identifier	<id>	where id is an orbit identifier
Ground or sea area	<code>|<tile number>	code identifying the area of land or sea covered by the data or tile number. Projects expecting to produce regional products have generally described codes to identify them.
Length of time period covered.	[<n>]<DAILY | MONTHLY | YEARLY>	Identify that data set holds n days, months or years data, where the filenames might otherwise coincide.
Resolution	<n>m	where n is the number of metres resolution.
Processing Centre	<Processing centre abbreviation>	name of processing centre
Algorithm	<Algorithm>	name of algorithm

v<GDS version>

Including the version number of the GHRSST Data Specification is optional for the CCI filenaming convention. If used it shall be 02.0

fv<File Version>

File version number in the form n{1,}[.n{1,}] (That is 1 or more digits followed by optional . and another 1 or more digits.)

2.8 Version Numbering

The version of any given data product is specified via both the 'product_version' in the global attributes and the 'file_version' identifier in the filenames. To meet the differing requirements of the project teams, these two version identifiers may be used in different ways, either:

a) used to represent the same version id (i.e. with identical numbering), or

b) used separately to represent different versioned actions.

For instance, in the case of the addition of new metadata to an existing data product, this could be represented either by a new version, with both the 'product_version' attribute and the 'file_version' in the filename incremented, or alternatively, may be considered the same 'product_version' but with an incremented 'file_version'.

The following general principles shall be adhered to:

- Version numbering shall always increase.
- The file version used in the filename shall uniquely identify the particular instance of the dataset, and should always increase with subsequent product versions.

R-10	The file version used in the filename shall uniquely identify the particular instance of the dataset, and should always increase with subsequent product versions.

2.9 Support for checking compliance with the data format

CCI data producers are encouraged to liaise with the CCI Open Data Portal when producing their data formats. The Open Data Portal can provide support for checking the compliance of any datasets to these standards.

Files can be checked for compliance with the CF convention using the on-line checking tool: http://cfconventions.org/compliance-checker.html

APPENDIX A – EXAMPLE GLOBAL ATTRIBUTE FIELDS

For fields coloured blue, please select terms from the CCI vocabulary at http://cci.esa.int/working-groups.

Global Attribute	Example content
title	ESA CCI aerosol product level 3 daily
institution	Plymouth Marine Laboratory
source	MERIS FR L1B version 4.02, MERIS RR L1B version 4.02, SPOT VGT P version 3.0[13]
history	2011-02-14 12:22:43 - Product generated from ORAC-SW dual-view version 2.0
references	http://www.esa-landcover-cci.org/
tracking_id	0c9e9570-cd44-102f-8001-0050c28e1010
Conventions	CF-1.8
product_version	1.3
format_version	CCI Data Standards vx.x
summary	This dataset contains Level-3 monthly global sea surface temperature products from satellite observations. Level-3 data are raw observations processed to geophysical quantities, and placed onto a regular grid.
keywords	satellite,observation,ocean
id	filename.nc
naming authority	uk.ac.pml
keywords_vocabulary	NASA Global Change Master Directory (GCMD) Science Keywords
cdm_data_type	Swath
comment	These data were produced at ESACCI as part of the ESA SST CCI project.
date_created	20120131T120000Z
creator_name	Plymouth Marine Laboratory Remote Sensing Group
creator_url	http://rsg.pml.ac.uk
creator_email	rsghelp@pml.ac.uk
project	Climate Change Initiative - European Space Agency
geospatial_lat_min	-90.0
geospatial_lat_max	90.0
geospatial_lon_min	-180.0
geospatial_lon_max	180.0
geospatial_vertical_min	0.0
geospatial_vertical_max	0.0
time_coverage_start	20120131T070000Z
time_coverage_end	20120131T083000Z
time_coverage_duration	P1Y
time_coverage_resolution	P1D

[13] For assimilation datasets the model could also be included here, e.g. “TM5 model, MIPAS L2 data (CCI_ozone product version 1.5)”

standard_name_vocabulary	'CF Standard Name Table v71'
license	ESA CCI Data Policy: free and open access
platform	Envisat
sensor	AATSR, MERIS
spatial_resolution	1.1km at nadir
geospatial_lat_units	degrees_north
geospatial_lon_units	degrees_east
geospatial_lon_resolution	0.1
geospatial_lat_resolution	0.1
key_variables	sea_surface_temperature,sea_surface_temperature_depth