EUMETSAT Prototype Satellite Data Cube - Methods Document

Release 1.0.2

EUMETSAT

May 06, 2021

LIST OF FIGURES

LIST OF TABLES

CONTENTS

CHAPTER

ONE

DOCUMENT INFORMATION

Doc. id :	usc-psdc/project-docs/methods-document
External version:	1.0.2
Author :	G. Cammarota, M. Cucchi, M. Di Bari (B-Open Solutions)

1.1 Document Change Record

Version	Version Date (as on profile)	DCR* No.	Description of Changes
1.0-doc-rc1	30-11-2020		Initial issue, prepared under contract EUM/CO/17/4600001943/OPN by B-Open Solutions srl.
1.0-doc-rc2	14-12-2020		First review in light of clarifications by EUMETSAT (temporal resolution and aggregation, table changes).
1.0-doc-rc3	24-03-2021		Synchronise content with the actual methods used upon implementation.
1.0.2	13-04-2021		Addressed RIDs, clarifying details, improving readability.

DCR = Document Change Request

CHAPTER

TWO

INTRODUCTION

2.1 Purpose

Climate Data Records (CDRs) and Interim Climate Data Records (ICDRs) provided by EUMETSAT and its network of SAFs are time-series of satellite-derived geophysical variables relevant for climate monitoring, which are inhomogeneous in terms of data format, geographical grid and metadata handling.

The EUMETSAT Prototype Satellite Data Cube for Drought & Vegetation Monitoring (PSDC) aims to gather and convert a set of EUMETSAT data and other data from third parties described in the *Annex I of [PSDC_WP]* in a homogeneous NetCDF4 format following the CF Conventions *[CFMC]* and the Common Data Model in *[CDS]*, making them more readily accessible and attractive to EUMETSAT users.

This document analyses the key issues in generating the PSDC, and shows the methods used to address them, with a specific focus on the re-gridding and re-formatting procedures applied throughout the conversion process.

Such methods, after approval, have been implemented in a dedicated customisation plugin of the EUMETSAT Data Tailor, then applied to the dataset. More details about specific metadata in the chosen data are described in the *Appendix - Input Products* section.

2.2 Scope

The scope of the document is to:

- provide a short description of each problem that was addressed in the generation of the PSDC
- report the strategies used to address these problems
- provide guidance for following evolutions of the project.

In its draft version, this document aimed at:

- EUMETSAT technical representatives, to allow them to review and eventually approve the proposed methods
- PSDC developers, to support them in the development of the prototype plugin.

In its finalised version, the document aims at:

- EUMETSAT personnel in charge of the validation of the prototype data cube
- EUMETSAT technical representatives, to provide guidance for following evolutions
- Users of the PSDC.

2.3 Applicable Documents

Table 2.1: Applicable documents.

#	Title	Reference
[PSDC_WP]	EPCT WP 2020-6: Prototype Satellite Data Cube - Work package description	EUM/USC/WPD/20/1197495

2.4 Reference Documents

Table 2.2: Reference documents.

#	Title	Reference
[CARD4L]	CEOS Analysis Ready Data for Land (CARD4L) Description Document	http://ceos.org/ard/index.html#slide1
[CDO]	Max-Planck-Institute for Meteorology, Climate Data Operators	https://code.mpimet.mpg.de/projects/cdo/
[CDS]	Copernicus Climate Data Store - Common Data Model Specification – v1.0	https://confluence.ecmwf.int/display/COPSRV/CDM%3A+Common+data+model+specification+-+v1.0
[CDS_API]	“Remote API of the Climate Data Store Toolbox”	https://github.com/bopen/cdstoolbox
[CFMC]	NetCDF Climate and Forecast (cf) Metadata Conventions	http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html#_abstract
[CF_STN]	CF Standard Names – Standard Name Table (v73, 23 June 2020)	http://cfconventions.org/standard-names.html
[CF_STN_GUIDE]	Guidelines for Construction of CF Standard Names	http://cfconventions.org/Data/cf-standard-names/docs/guidelines.html
[CMSAF]	CM SAF Website	http://cm-saf.eumetsat.int
[CMSAF_TOOL]	CM SAF R Toolbox	https://www.cmsaf.eu/EN/Products/Tools/R/R_node.html
[CRAN]	Comprehensive R Archive Network	https://cran.r-project.org/
[DATATAILOR]	Data Tailor source code	https://gitlab.eumetsat.int/open-source/data-tailor.
[DR4PSDC]	Data Record PSDC Drought and Vegetation Monitoring	EUM/USC/DOC/20/1197387
[EASEGRID]	EASE-Grid 2.0: Incremental but Significant Improvements for Earth-Gridded Data Sets	https://www.mdpi.com/2220-9964/1/1/32/htm
[EDT]	The Data Tailor	https://www.eumetsat.int/website/home/Data/DataDelivery/NewPilotDataServices/TheDataTailor/index.html
[NETCDF]	Network Common Data Form	https://www.unidata.ucar.edu/software/netcdf/
[NOAA]	NOAA National Centres for Environmental Information server	https://data.nodc.noaa.gov/thredds/catalog.html
[OBS4MIPS]	Obs4Mips (Observations for Model Intercomparison Projects) Data Specifications	https://esgf-node.llnl.gov/projects/obs4mips/DataSpecifications

continues on next page

Table 2.2 – continued from previous page

#	Title	Reference
[OPSTRR]	OPS/TSS Working Practice on TRR_TRB	EUM/TSS/DOC/13/688793
[PSDC_EXECPLAN]	EUMETSAT Prototype Satellite Data Cube - Execution Plan	usc-psdc/project-docs/execution-plan
[PSDC_TP]	EUMETSAT Prototype Satellite Data Cube - Test Cases and Test Procedures	usc-psdc/project-docs/verification-validation-test-cases-test-procedures
[PSDC_VCD]	EUMETSAT Prototype Satellite Data Cube - Verification Control Document	Verification control tracked in *[PSDC_VVPLAN]*
[PSDC_VVPLAN]	EUMETSAT Prototype Satellite Data Cube - Verification and Validation Plan	project-docs/verification-validation-plan
[THREDDS]	UCAR Community Programs – Unidata – THREDDS Data Sever	https://www.unidata.ucar.edu/software/tds/
[TRS]	Technical Information Security Requirements	EUM/ISMS/REQ/18/966683
[TRST]	Security Patch Management Policy	EUM/ISMS/POL/17/941805
[UNCERT]	Evaluation of Measurement Data—Guide to the Expression of the Uncertainty in Measurement; Technical Report JCGM 100:2008 GUM 1995 with minor corrections; BIPM: Cedex, France, 2008. Page 33 chapter 5.2.2	https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf

2.5 List of TBDs

Table 2.3: Items yet to be defined.

Item	Place in text	Expected definition time
PSDC Verification and Validation Test Cases and Test Procedures Document reference	*Introduction*	Closed
PSDC Verification Control Document reference	*Introduction*	Closed
Details of *bounds* attribute for Pixel Center	*Spatial Resolution*	Closed

CHAPTER

THREE

PROBLEM ANALYSIS

3.1 Overview

In this section we identify and analyse the main issues that have been handled in the generation of the PSDC.

Identified issues fell in the following main categories:

- *Spatial Resolution*
- *Temporal coordinate*
- *Aggregated variables*
- *Provenance*
- *Common data model*.

They are analysed in details in the following paragraphs.

3.2 Spatial Resolution

This section describes the issues handled in manipulating data, precisely how the spatial representation has been managed to meet the PSDC requirements.

3.2.1 Common geographic reference system

When performing operations using geographical coordinates, data must be referred to the same reference system to be appropriately compared or combined. Specifically, data must share the same coordinates in the same reference system.

Most of the variables for the PSDC data set were reported in a geostationary system. Therefore, they have been reprojected to the requested WGS-84 regular longitude-latitude grid as per *REQ 07 in [PSDC_WP]*.

A related issue, which does not affect the PSDC prototype but may concern a future data cube evolution, is the range definition of the *longitude* coordinate. Longitude is usually defined between either -180 and 180 degrees or between 0 and 360 degrees. Both definitions lead to problems in managing data which overlay across:

- the international date line (when longitude is defined between -180 and 180)
- the Greenwich meridian (when longitude is defined between 0 and 360).

3.2.2 Regridding

PSDC data is defined on coordinates grids with spacings congruent with each other. For each spacing, all data must lie on the same grid.

All input products, with the exception of GPCC Precip-Monitoring V6, follow the convention pixel-is-point with a grid offset equal to 0. Two products with the "pixel" label anchor (see Fig. 3.1) and 0 grid offset with different grid spacing are not congruent. On the contrary, two products with an "area" anchor (see Fig. 3.2) and *0* grid offset are, in fact, congruent. Therefore, we have applied a regridding process to make the former congruent.

In addition, the input NDVI products have a spacing of 1/112 degrees. Target spacings are defined by *REQ 08 in [PSDC_WP]* and summarised in the *Appendix - Input Products*. A regridding process has been applied to adapt the original grids to the target ones.

Note that, for future evolutions, it may be desirable to have data provided in multiple resolutions so that a communal, less dense grid is used only to combine data that are defined on different grids.

3.2.3 Pixel-wise error measures and quality flags

For pixel-wise error measures and quality flags, resampling has been managed separately. In fact, some special variables like data quality flags, relative errors, and statistical quantities cannot be resampled using standard interpolating functions. Instead, specific interpolating functions have been used, that take into account the way such variables are calculated. For variables indicating pixel-based uncertainty measures, such methods are provided in the *uncertainties reference document* as per requirement *REQ 02 in [PSDC_WP]*.

3.2.4 Pixel center

On the grid, each data value must be associated with either the vertex, or the centre of the cell described by the coordinates. In the metadata, this association is respectively flagged via coordinate attribute anchors "area" or "pixel".

When a value is associated to the centre of the cell, this cell can be thought as representing a point in the real world. However, such grids with cells of different sizes are not congruent (see Fig. 3.1), as a cell does not include a whole number of multiple sub-cells.

On the contrary, if each cell represents instead an area in the real world, cell-vertices make up all possible locations to which the data value can be associated. Therefore, vertices as locations are a preferred choice, as grids of different sizes are congruent when a vertex is used as tie-point (see Fig. 3.2).

Since PSDC data uses grids with different resolutions, choosing the "area" option makes such grids congruent from one to the other. Fig. 3.1 and Fig. 3.2 further clarify the underlying reasons.

The CF conventions try to tackle this problem by adding auxiliary variables containing the vertex coordinates of each cell. Those new variables are thus defining the cells. In this case, the *bounds* attribute is added to the appropriate coordinate and filled with the name of the variable containing the cell vertex values for such coordinate. For PSDC data no *bounds* attribute have been added, since a regridding process was applied to the input products to make them congruent, generating new spatial grids following the pixel-is-area convention.

3.3 Temporal coordinate

This section analyses the issues that have been encountered when transforming data to comply with PSDC requirements, specifically the temporal data representation.

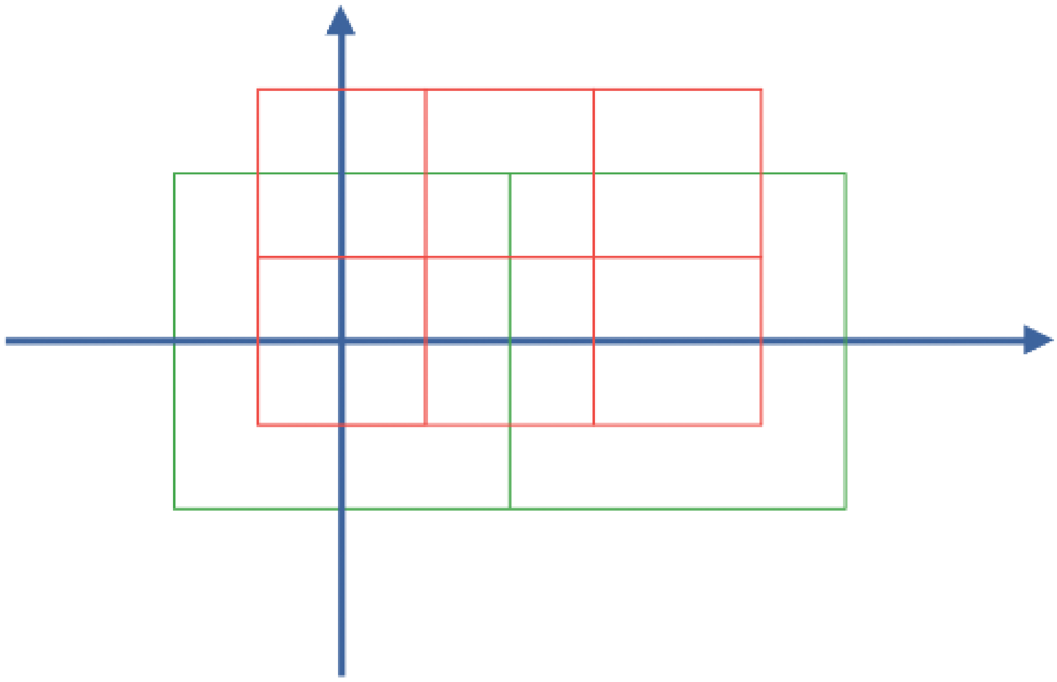

Fig. 3.1: The case when using different grids with the anchor attribute: “pixel”.

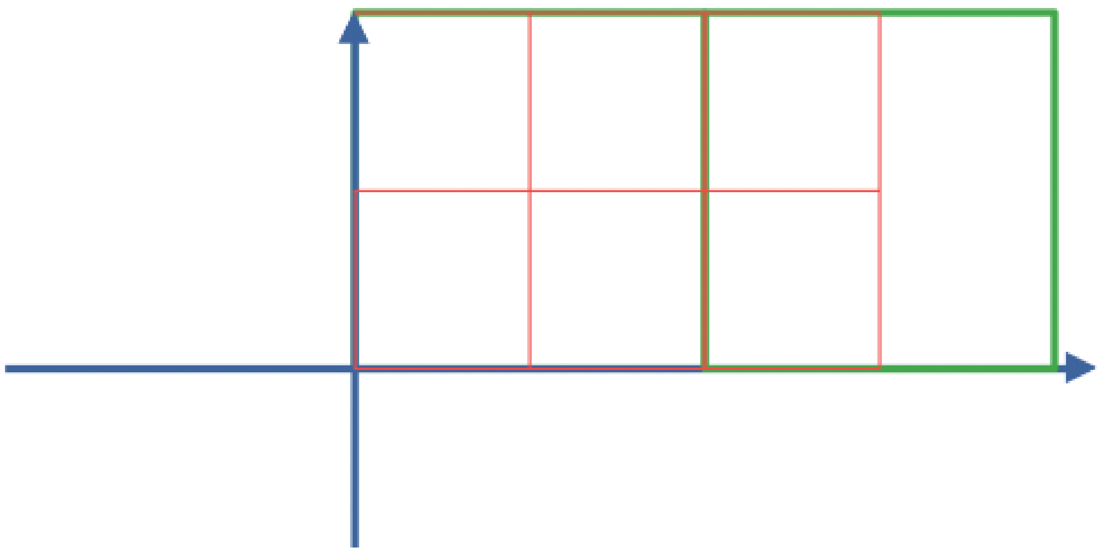

Fig. 3.2: The case when using different grids with the attribute: “area”.

3.3.1 Reference time

Reference time is the time from when the coordinate starts. In fact, time is indicated as a certain amount of units counted starting from the reference time. The reference time has been set equal for each variable, explicitly indicated in the *units* attribute of the time coordinate.

This issue is strictly related with the one concerning time units.

3.3.2 Temporal resolution

As per requisite *REQ 06 in [PSDC_WP]*, the indicated temporal resolution has been used. The intervals are *month*, *10-day*, *day* and *hour* (*10-day* starting at 1st, 11th, and 21st of the month). The original data were already in the requested temporal resolution, except for LST data which is provided with a temporal resolution of 15 minutes.

3.4 Aggregated variables

In the Data Cube, it must be possible to compare and combine variables. This section analyses some outstanding issues in this respect.

3.4.1 Time aggregation

When dealing with time aggregated variables, such as a monthly mean, it is important to specify which point in the time axis the values are associated with. For example, the single values of a monthly mean can be associated either to the first or the last (or even, in selected cases, the fifteenth) day of each month.

Associating the value to the first day ensures uniformity among different months, while losing the concept that such value is associated with the whole month. Instead, using the last day produces inhomogeneity. The last day indeed differs from month to month, leading to other issues when performing statistical analyses.

3.5 Provenance

Any of the transformation processes discussed so far, has the potential to alter the data.

In order to keep track of changes and inform the final user about all processing operations the data has gone through, data origin is saved and exposed as data attribute for each manipulation applied.

3.6 Common data model

As per requisite *REQ 04 in [PSDC_WP]*, all data records included in the PSDC follow a common data model. This common data model follows the CF-Conventions (version 1.8), implementing all requirements and as many of the recommended guidelines as is practical. Supplementing the CF Conventions data also complies with the Copernicus Climate Data Store (CDS).

CHAPTER

FOUR

METHODS

4.1 Overview

In this section we detail the methods and tools that have been used to overcome the problems exposed in chapter *Problem Analysis*.

4.2 Spatial Resolution

The section describes the methods and tools that have been used to perform the spatial manipulations analysed in *Spatial Resolution*.

4.2.1 Common geographic reference system

Per requisite *REQ 07 in [PSDC_WP]*, the WGS-84 system with a longitude and latitude regular grid has been used.

Therefore, all the original data projected in a geostationary system has been reprojected to a WGS-84 projection (see *Appendix - Input Products* for details).

One of the main tools to perform reprojection operations on common reference systems is available in the Climate Data Operator (CDO) suite. Another reliable and tested tool to compute such operations is GDAL. Although GDAL is already used in other Data Tailor customisation plugins, we have adopted the CDO suite since it comprises a collection of standard operators used in the climate and meteorological scientific community.

As long as the PSDC variables are concerned, defining the longitude range between -180 and 180 degrees did not cause any issue since the requested bounding box was entirely included in such range.

Information on the reference system has been stored in the *crs* variable. The following solutions are adopted:

- the CM-SAF, LSA_SAF, H-SAF, GPCC, and ERA5-Land products are confirmed by the providers to be georeferenced to WGS84, so this information has been used in *crs* even if it is not declared in the input data
- according to the provider, the value of *inverse_flattening* attribute in the *crs* variable of LSA-SAF products was wrong. The correct value of *inverse_flattening* has been reported in the output products instead.

4.2.2 Regridding

Per requirement *REQ 08 in [PSDC_WP]*, the final resolution has been kept as close as possible to the original one and specifically:

- all data that followed the convention pixel-is-point with a grid offset equal to 0 is regridded because two products with “pixel” label anchor and 0 grid offset with different grid spacing are not congruent. On the contrary, two products with “Area” anchor and grid offset 0 are congruent
- all NDVI data were in the correct projection, but using a spacing of 1/112 degrees. They have been regridded using a spacing of 0.01 degrees.

Coordinates definitions are:

- 1 degree as grid spacing: longitude is defined in a range from -180 to 179 and latitude from -90 to 89
- 0.1 degrees as grid spacing: longitude is defined in a range from -180 to 179.9 and latitude from -90 to 89.9
- 0.05 degrees as grid spacing: longitude is defined in a range from -180 to 179.5 and latitude from -90 to 89.5
- 0.01 degrees as grid spacing: longitude is defined in a range from -180 to 179.99 and latitude from -90 to 89.99.

The CDO suite provides standard tested tools to perform both downsampling and oversampling operations with different choices among available resampling functions. CDO *remapbil* operator has been used to regrid all data following the convention pixel-is-point with a grid offset equal to 0. Whereas, in order to regrid the LSA SAF NDVI input variables from the input regular grid with 1°/112 of grid spacing to the output regular grid with a grid spacing of 0.01°, the *scipy.ndimage.map_coordinates* function has been used to perform a nearest neighbour resampling.

4.2.3 Ancillary variables

For variables indicating pixel-based uncertainty measures, as per requisite *REQ 02 in [PSDC_WP]*, methods for correct interpolation are provided in the *Evaluation of Measurement Data—Guide to the Expression of the Uncertainty in Measurement*.

The following approach has been adopted to regrid pixel-based uncertainty measures. The re-gridding implies that the value of each “new” cell is in the middle of a square made of 4 input cells, as shown in Fig. 4.1, where the output cells are colored in light-blue and the input cells are colored in green. So, a sliding window of size 2x2 has been considered to estimate the value of each output cell, according to the following rules:

- if the ancillary variable represents an error estimate: we have used the same method adopted for the resampling of the main variable to which the error refers, i.e. CDO *remapbil*
- if the ancillary variable is a quality flag: the resampling has been made according to the following rules:
  1. if in the sliding window 4 values are identical, the output value is that value
  2. if in the sliding window the values are discordant, the output value is either an already existent value, or a new value that stands for “undefined”, “uncertain”.

  With regard to rule 2 (i.e. discordant values in the sliding window), to resample the quality flag variables, we have adopted the following values:

  – LSA-SAF METREF quality flag: the new value, *-5*, stands for “undefined”

  – LSA-SAF LST quality flag: the new value, *1*, stands for “undefined”

  – LSA-SAF LAI, FVC, FAPAR quality flags: the existent value, *0*, is associated to the first bit. The actual meaning is “ocean”, the new meaning is “ocean/undefined”.
- for LSA-SAF NDVI products: resampling also implies a change in spatial resolution from 1°/112 to 0.01°. In this case, we have used, also for the *ndvi* main variable, the same interpolation method chosen for the ancillary variables (*day_in_decad*, *total_clear_observation* and *status_flag*), i.e. the nearest neighbour interpolation. In this way, missing values are consistent in all variables. Note that the resampling of the

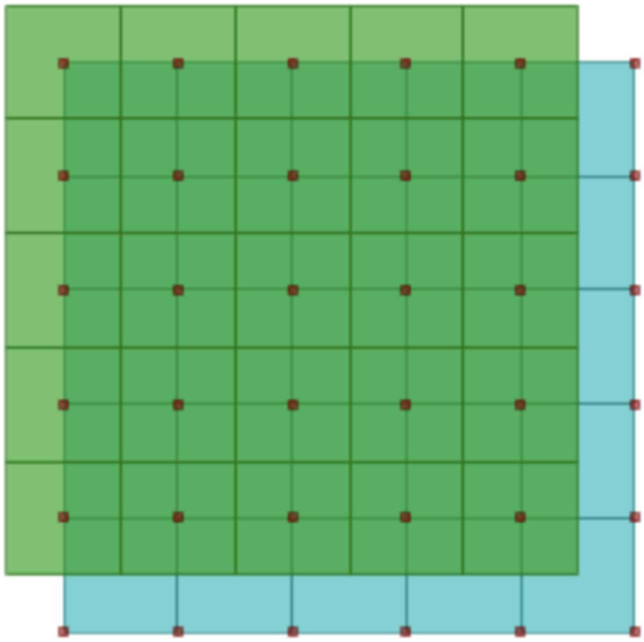

Fig. 4.1: Grid with data associated with the centre of cells (green) versus grid with data associated with a vertex of the cells (blue).

associated quality flag *status_flag* does not follow the rules described above because the nearest neighbour interpolation is already enough. Knowing that CDO *remapnn* operator is very slow, we have used *scipy.ndimage.map_coordinates* function.

4.2.4 Pixel center

To achieve congruency between different grids, coordinates have been re-labeled so that each couple of coordinate values indicates the south-west corner of a cell. Moreover, the point indicated by the longitude/latitude couple (0, 0) corresponds to the lower left corner of the cell having as north-east corner:

- (1, 1) when grid spacing is 1 degree
- (0.1, 0.1) when grid spacing is 0.1 degrees
- (0.05, 0.05) when grid spacing is 0.05 degrees
- (0.01, 0.01) when grid spacing is 0.01 degrees.

4.3 Temporal coordinate

This section describes the methods and tools that have been used for the temporal coordinates manipulation analysed in *Temporal coordinate*.

4.3.1 Reference time

In UNIX systems, the *de-facto* standard for this value is 01/01/1970. This date is also the reference used in the Copernicus Climate Data Store (CDS) common data model. Such reference time has been explicitly indicated in the *units* attribute of the time coordinate i.e. "seconds since 1970-01-01". The temporal reference system UTC has been employed. The units is *seconds* from the reference time. An example is reported in Fig. 4.2

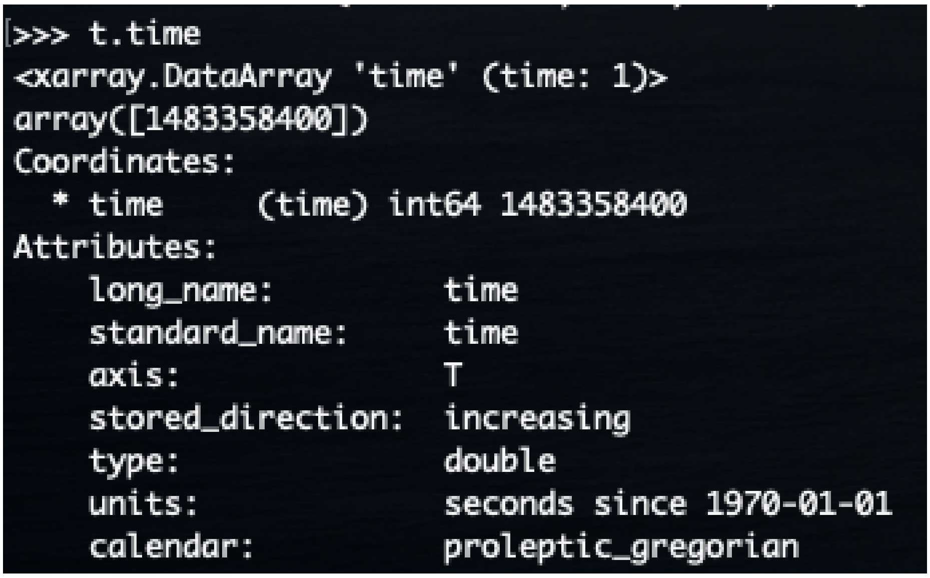

Fig. 4.2: Example of the time coordinate defined in the CDS common data model.

This issue is strictly related to the time units one, and they have been handled together using the tested library *xarray*.

4.3.2 Temporal resolution

Time intervals are *month*, *10-day*, *day* and *hour* (*10-day* starting at 1st, 11th and 21st of the month). The original data were already in the requested temporal resolution, except for LST data, provided with a temporal resolution of 15 minutes. For LST data only, full-hour values have been considered so that the final temporal resolution is *hourly*.

4.4 Aggregated variables

In this section, the methods and tools that have been used to handle the data aggregation issues are reported.

4.4.1 Time aggregation

The start of the time intervals has been used as value for time coordinate (e.g. first day of the month for monthly data). Since all data were already in the final time resolution and no aggregation was requested, when the reference time was changed, *xarray* automatically managed time labeling accordingly.

4.5 Provenance

A 'history' attribute has been used to keep track of the data origin and all the applied processing operations. This attribute records the related information in human-readable metadata, and is appended as a global attribute to the data.

CF Convention recommendations on *history* attribute have been applied preserving the original provenance stored in the attribute, and appending information on a new line.

In this case, the only call is to the EUMETSAT Data Tailor and such call is appended to the history metadata using the following template:

```
<ISO format date and time of call>: epct run-chain <call-parameters>
```

4.6 Common data model

The PSDC common data model is structured to be compliant to the CF-Conventions (version 1.8) and to the common data model of Copernicus Climate Data Store (CDS).

In particular, to enforce compliance with the CDS, when input files contain multiple variables having an obvious meaning even on their own, those are split into separated files. Therefore, each PSDC dataset contains one geophysical variable (together with its associated ancillary variables), except for the input products having geophysical variables strictly related to each other. Those are kept in the same file, deviating from the CDS common data model standard.

4.6.1 Variable names

For variable names, the standard names from the CF conventions are used. The most recent version of those are found in document *NetCDF Climate and Forecast (cf) Metadata Conventions*, applying the official *Guidelines for Construction of CF Standard Names*.

A detailed investigation of the input dataset revealed, in some cases, the need for new variable names. In these cases, we have escalated the request to add the variable name to the CF conventions. A proposal for the variable name has been added in the *tentative_standard_name* attribute of the output dataset, to make it visible to the users, who may then provide feedback.

Long names are written with no capital letters.

Where short names do not exist, an attempt has been made to generate meaningful ones by analogy with existing cases. An overview of the proposed tentative standard names is provided in *Appendix - Tentative standard names*.

4.6.2 Attribute names and values

For the *platform*, *sensor*, *product_version* and *doi* attributes:

- when the value in the input product was *NA*, default values from configuration files were used
- for *CMSAF_CDR_** products, the *product_version* has been read from original metadata, and the attribute *doi* from the *id* attribute; note that in some input products the *id* was in the form *DOI:10.5676/EUM_SAF_CM/SARAH/V002_01* which has been kept without modifications in the output product
- for *CMSAF_ICDR_** products, the value of the *sensor* attribute has been read from original attribute *instrument*
- for all other products, *platform*, *sensor* and *product_version* has been read from the original metadata.

The Region of Interest (or bounding box) in the original metadata was expressed in different ways (i.e. different attribute names and different value types). As there is no uniform prescription in the CF Conventions on how to express them, the bounding box attributes have been removed from the output products, and the information has been served as metadata of the whole web catalog (i.e. not in the dataset, but in the THREDDS definitions).

CHAPTER

FIVE

APPENDIX - INPUT PRODUCTS

The following table summarizes temporal and spatial details for each type of input products supported:

Table 5.1: Input products details.

Variable	Temporal Resolution	Stats Value	Temporal Resolution in Cube	Projection	Anchor	Regridding method (main variables)*	Spatial Resolution in Cube	Format
SIS	monthly	mean	monthly	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
SIS	monthly	mean	monthly	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-1.4
SIS	daily	mean	daily	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
SIS	daily	mean	daily	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
DNI	monthly	mean	monthly	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
DNI	monthly	mean	monthly	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
DNI	daily	mean	daily	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
DNI	daily	mean	daily	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
SDU	monthly	sum	monthly	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
SDU	monthly	sum	monthly	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
SDU	daily	sum	daily	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
SDU	daily	sum	daily	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
LST	15 min.	Instantaneous	hourly	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
LST	15 min.	Instantaneous	hourly	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
METREF	daily	Instantaneous	daily	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
METREF	daily	Instantaneous	daily	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
NDVI	10-daily	composite	10-daily	geographic	pixel	nearest neighbour using scipy.ndimage.map_coordinates function	0.01°	IMG +header - XML metadata
NDVI	10-daily	composite	10-daily	geographic	pixel	nearest neighbour using scipy.ndimage.map_coordinates function	0.01°	IMG +header - XML metadata
FVC	daily	composite	daily	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
FVC	daily	composite	daily	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
LAI	daily	composite	daily	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
LAI	daily	composite	daily	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
fAPAR	daily	composite	daily	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
fAPAR	daily	composite	daily	geographic	pixel	CDO *remapbil* operator	0.05°	NetCDF-4
Soil Wetness	daily	Valid 00UTC	daily	geographic	pixel	CDO *remapbil* operator	0.1°	NetCDF-4 (GRIB2 also available)

continues on next page

Table 5.1 – continued from previous page

Variable	Temporal Resolution	Stats Value	Temporal Resolution in Cube	Projection	Anchor	Regridding method (main variables)*	Spatial Resolution in Cube	Format
Soil Wetness	daily	Valid 00UTC	daily	geographic	pixel	CDO *remapbil* operator	0.1°	NetCDF-4 (GRIB1 also available)
Precip	monthly	mean	monthly	geographic	area	not regridded	1°	NetCDF-4
T2m	monthly	mean	monthly	geographic	pixel	CDO *remapbil* operator	0.1°	GRIB

* Note that the Regridding method used for Ancillary variables is detailed in *Ancillary variables*.

CHAPTER

SIX

APPENDIX - TENTATIVE STANDARD NAMES

The following table summarizes the tentative standard names for each product:

Table 6.1: Proposed standard names.

Product-ID	Original variable short-name	Proposed standard-name for PSDC
CMSAF_CDR_DNI / CMSAF_ICDR_DNI	DNIC/DNIc	surface_normalized_direct_shortwave_flux_in_air_assuming_clear_sky
CMSAF_CDR_DNI / CMSAF_ICDR_DNI	DNI	surface_normalized_direct_shortwave_flux_in_air
GPCC_PRECIP	s	number_of_stations
GPCC_PRECIP	interpolation_error	interpolation_error
GPCC_PRECIP	solid_p	proportion_of_solid_precipitation
GPCC_PRECIP	liquid_p	proportion_of_liquid_precipitation
GPCC_PRECIP	corr_fac	correction_factor
GPCC_PRECIP	abs_gauge_err	absolute_gauge_error
GPCC_PRECIP	rel_gauge_err	relative_gauge_error
HSAF_CDR_WET/HSAF_ICDR_WET	var40	soil_wetness_index_in_layer_1
HSAF_CDR_WET/HSAF_ICDR_WET	var41	soil_wetness_index_in_layer_2
HSAF_CDR_WET/HSAF_ICDR_WET	var42	soil_wetness_index_in_layer_3
HSAF_CDR_WET/HSAF_ICDR_WET	var43	soil_wetness_index_in_layer_4
LSASAF_CDR_LST / LSASAF_NRT_LST	LST	land_surface_temperature
LSASAF_CDR_LST / LSASAF_NRT_LST	standard_error	land_surface_temperature standard_error
LSASAF_CDR_METREF / LSASAF_NRT_METREF	METREF	reference_evapotranspiration
LSASAF_CDR_NDVI / LSASAF_ICDR_NDVI	DAY	number_of_days_in_decad
LSASAF_CDR_NDVI / LSASAF_ICDR_NDVI	TCO	number_of_clear_observation

CHAPTER

SEVEN

GLOSSARY

This section contains the main definitions and acronyms used throughout the document.

Acronyms not specific to Data Cube can be found at https://www.eumetsat.int/website/home/Satellites/Glossary/Acronyms/index.html

7.1 Definitions

Customisation plugin A Python software package which contains one or more customisation backends.

Customisation process The process executed by the EUMETSAT Data Tailor to apply the customisation functions specified by the user to one or more input products. A customisation process only generates one output.

ROI Extraction Data Tailor Customisation Function which allows to extract a Region of Interest according to a predefined bounding box from the input product.

Validation The process that ensures that EUMETSAT (user) requirements are met, and the system behaves as intended in an "operational" environment.

Verification The process that ensures that the system is correctly designed and implemented with respect to the system requirements.

7.2 Acronyms

C3S Copernicus Climate Change Service

CARD4L CEOS Analysis Ready Data for Land

CDR Climate Data Records

CDS Climate Data Store

CEOS Committee of Earth Observation Satellites

DR4PSDC Data Records for Prototype Satellite Data Cube

FAT Factory Acceptance Tests

FCDR Fundamental Climate Data Records

FDR Fundamental Data Records

GPCC Global Precipitation Climatology Centre

HDF Hierarchical Data Format

ICDR Interim Climate Data Records

ICSI Information Centric Service Infrastructure

MET METeosat

NetCDF Network Common Data Form

OSAT On-Site Acceptance Test

PSDC Prototype Satellite Data Cube

ROI Region Of Interest

TBC To Be Confirmed

TBD To Be Defined

TRB Test Review Board

TRR Test Readiness Review

USC User Support and Climate Services Division of EUMETSAT

VCD Verification Control Document

VG Validation Goal

WMS Web Map Services and Visualization Systems