

The SAMD Product Standard (Standardized Atmospheric Measurement Data)

Version 1.2

Andrea Lammert¹

Verena Grützun¹

Erasmia Stamnas²

¹ University Hamburg

² University Cologne

November 30, 2018

The SAMD data-product description-document includes the conventions for file names, variables and NetCDF-files. The standardized XML-file convention is included as well as all necessary abbreviations for institutes, instruments, variables, etc.

Many thanks to Felix Ament, Holger Baars, Andreas Behrendt, Cintia Carbajal Henken, Susanne Crewell, Hartwig Deneke, Jan Handwerker, Anne Hirsikko, Henk Klein Baltink, Martin Kohler, Volker Lehmann, Ulrich Löhnert, Christine Nam, Johannes Quaas, Sonja Reitter, Patric Seifert, and Volker Winkelmann for their valuable comments for this document.

Special thanks to the project HD(CP)², which is founded by the Federal Ministry of Education and Research (BMBF), FKZ:O1LK1211A/C.

Technical background information of SAMD could be found in:

Stamnas, E., A. Lammert, V. Winkelmann, and U. Lang: The HD(CP)2 Data Archive for Atmospheric Measurement Data. *ISPRS Int. J. Geo-Information* 5(7), <https://doi.org/10.3390/ijgi5070124> (2016).

SAMD contact:

Dr. Andrea Lammert
andrea.lammert@uni-hamburg.de
new: lammert@dkrz.de
Meteorological Institute
University of Hamburg
Bundesstrasse 55
D-20146 Hamburg

Erasmia Stamnas
hdcp2admin@uni-koeln.de
Regional Computing Centre
University of Cologne
Weyertal 121
D-50931 Cologne

Version history

HD(CP)² Observation Data Product Standard

Document Version	Date	Responsible	Changes
1.0	2013-07-14	Verena Grützun, Andrea Lammert	Initial Version
1.0.1	2013-08-01	Verena Grützun	included pyranometer network
1.1	2013-16-09	Verena Grützun	<ul style="list-style-type: none"> - changed time unit to Unix time - added ordering of dimensions - removed note to satellite community in level 2 description - added doppler lidar measurement strategies (instrument names) - added knmi to institutes
1.2	2013-11-29	Verena Grützun	<ul style="list-style-type: none"> - standard_name attribute should be left out completely if no standard name present (as opposed to being left empty) - added doppler lidar measurement strategies, pyranometer, pyrgeometer (instrument names) - added knmi and rao to institutes - added some minor notes to clarify existing points
1.2.1	2013-12-17	Verena Grützun	- corrected example file names (Sec.3)
1.3	2014-01-17	Verena Grützun	<ul style="list-style-type: none"> - made underscores in attribute names in netCDF files binding (Sec. 6) - added an order for the dimension time, height, range (Sec. 6) - added data approval Step-by-Step guide in Appendix - introduced Cloudnet as instrument combination in Tables (Sec. 4, < instnn<) - added mwrBL (microwave boundary layer boundary layer scan) to instruments - O1 Standards: added contact persons (Appendix) - added a couple of variables to the variable table. They are marked with new (Sec 5). - added information about the O1 document to appendix - added keywords for the data base search tree (Sec. 7 and Appendix) - updated Compliance check and data upload (Sec. 8) - updated list of approved data sets - added information about the provision of quick looks (Sec. 3)
1.3.1	2014-05-09	Andrea Lammert	- corrected search tree
1.3.2	2014-06-05	Andrea Lammert	<ul style="list-style-type: none"> - added chapter Variable groups - updated and corrected variable table - updated Quick look information
1.4	2014-06-18	Andrea Lammert	<ul style="list-style-type: none"> - added HOPE Melpitz and other campaigns - added Actos, UV-lidar - added quality flag description - updated variable table
1.4.1	2014-10-31	Andrea Lammert	- updated variable table
1.4.2	2015-01-09	Andrea Lammert	<ul style="list-style-type: none"> - add new Variable group Clouds - add new instrument combination for " Full domain"
1.5	2015-02-10	Andrea Lammert	- new global attribute Dependencies (p. 20)
2.0	2015-06-02	Andrea Lammert	Document rewritten

Continued on next page

Document Version	Date	Responsible	Changes
2.1	2015-09-18	Andrea Lammert	- added alphabetical List of var-names (Appendix) - added new satellite synergy product
2.2	2015-10-06	Andrea Lammert	- added alphabetical List of CF-standard_names (Appendix) - added new variables
2.2.1	2015-12-02	Andrea Lammert	- added new variable

SAMD Product Standard

Document Version	Date	Responsible	Changes
1.0	2016-09-30	Andrea Lammert, Erasmia Stamnas	Document rewritten
1.1	2018-02-28	Andrea Lammert	new supersites, new instruments, completion of information about checksums and file pass
1.2	2018-11-29	Andrea Lammert	new campaigns

Contents

1. Introduction	2
1.1. General Information	2
1.2. History	2
2. The SAMD data archive	3
2.1. Long-term measurements	3
2.2. Short-term measurements	4
3. Data policy	6
4. The SAMD Data Portal: samd-archive.org	7
5. Naming conventions	9
5.1. The file naming	9
5.1.1. The Measurement type (<kkk>)	9
5.1.2. The Supersites (<sss>, part 1)	10
5.1.3. The Institutions (<sss>, part 2)	10
5.1.4. The Instruments (<instnn>)	11
5.1.5. The Level (<lll>)	12
5.1.6. The version numbering (<vnn>)	13
5.2. The variable naming	13
6. NetCDF conventions	23
6.1. Global Attributes	23
6.2. Instrument meta-data for ground based instruments	24
6.3. Dimensions	25
6.4. Coordinate system	25
6.5. Variables	27
7. Metadata conventions	28
7.1. The SAMD metadata web interface	28
7.2. PID	30
7.3. Keywords	30
8. Compliance check and data upload	31
A. Data approval – Step-by-Step	32
B. SAMD Metadata	33
C. List of variable names	38
D. List of CF-standard_names	40

1. Introduction

1.1. General Information

The files provided for the SAMD database are in the format of NetCDF, be provided as daily files, contain 1 variable per file and its error (if available). Therefore the files have to obey the standards defined in this document, contain a standardized minimum of meta-data in the NetCDF header defined below, have a unique Metadata file in XML format as described below. We recommend using NetCDF version 4, but version 3 is also accepted. Please find a description of the NetCDF format here:

<http://www.unidata.ucar.edu/software/netcdf/>.

In general we want to provide daily files for the products. For some cases it is more practicable to provide files with a different increment, e.g. for satellite overpasses or scans. We do allow for well justified exceptions, but urge you to check back with the Observation Expert Users in advance.

With regard to the representation of the products in the NetCDF-files we follow the principles given in the NetCDF Climate and Forecast (CF) Metadata Conventions version 1.6 (CF-Standard¹) as far as possible. This especially means that our definition of the coordinate arrays and data arrays bases on this standard. For many variables, an attribute `<standard_name<`, a name and applicable units are defined there, as well as representations of coordinate systems. However, since it was developed for model data, not all observational data are covered. This especially means that we cannot provide the standardized name and attribute `<standard_name>` for all variables covered by this document but defined some variable names for SAMD (see Section 6). Generally, one quantity per file is mandated, which may have ancillary variables like the quantity's error or quality flags. For some instruments however, it is more meaningful to combine several variables (e.g., brightness temperatures for microwave radiometers). We allow for combining variables in level 1 data (definitions for levels see Section 4). If you want to provide a variable group in your files, which is not yet contained in this data description document, please contact the SAMD team.

1.2. History



The SAMD archive has been developed as part of the project "High definition Clouds and Precipitation for advancing Climate Prediction", short HD(CP)², funded by the German Federal Ministry of Education and Research. HD(CP)² is a German-wide research initiative to improve our understanding of cloud and precipitation processes and their implication for climate prediction. The project strives to build and use a model capable of very high-resolution simulations with horizontal grid spacings of 100 m. This model will be run as short hindcasts over a few days, to advance the parametrization of clouds and precipitation, and to reduce uncertainty in climate projections related to these quantities. The second key aspect of HD(CP)² is to use, organize and improve ground, in situ and satellite based observations of cloud and precipitation events on a scale that was not possible before. The third pillar of HD(CP)² is to understand, synthesize and combine the results from modeling and observations to evaluate, modify and improve existing climate models. Please find more information at: www.hdcp2.eu. In the first three years, the HD(CP)² observation data archive is open just for project members. Afterward the archive has to be open for all users, especially for researchers. Actual the SAMD archive includes mainly observations based on HD(CP)², further projects and data sets are welcome.

¹<http://cfconventions.org/cf-conventions/v1.6.0/cf-conventions.html>

2. The SAMD data archive

2.1. Long-term measurements

The long term observations of the different Supersites are one of the key elements of the SAMD data archive. At the moment, there is not really a common definition of the term "meteorological Supersite", it depends on the project and the instrument focus. As an example: For the Supersite RAO (Richard-Assmann-Observatory Lindenberg), the so-called "Lindenberg column", the German Weather Service defines the current objectives and research activities, among other, as the "three-dimensional/four-dimensional long-term climate monitoring of the atmosphere (4D Lindenberg column), including the analysis of long-term trends;" (www.dwd.de/mol)

For the SAMD archive a Supersite means an observatory for long term measurements, including as minimum equipment: a cloud radar, a microwave radiometer and a lidar system (see Fig 1). Each additional instrument, like a meteorological tower, surface meteorology station, radio soundings, and/or radiation measurements, of course is a surplus profit.

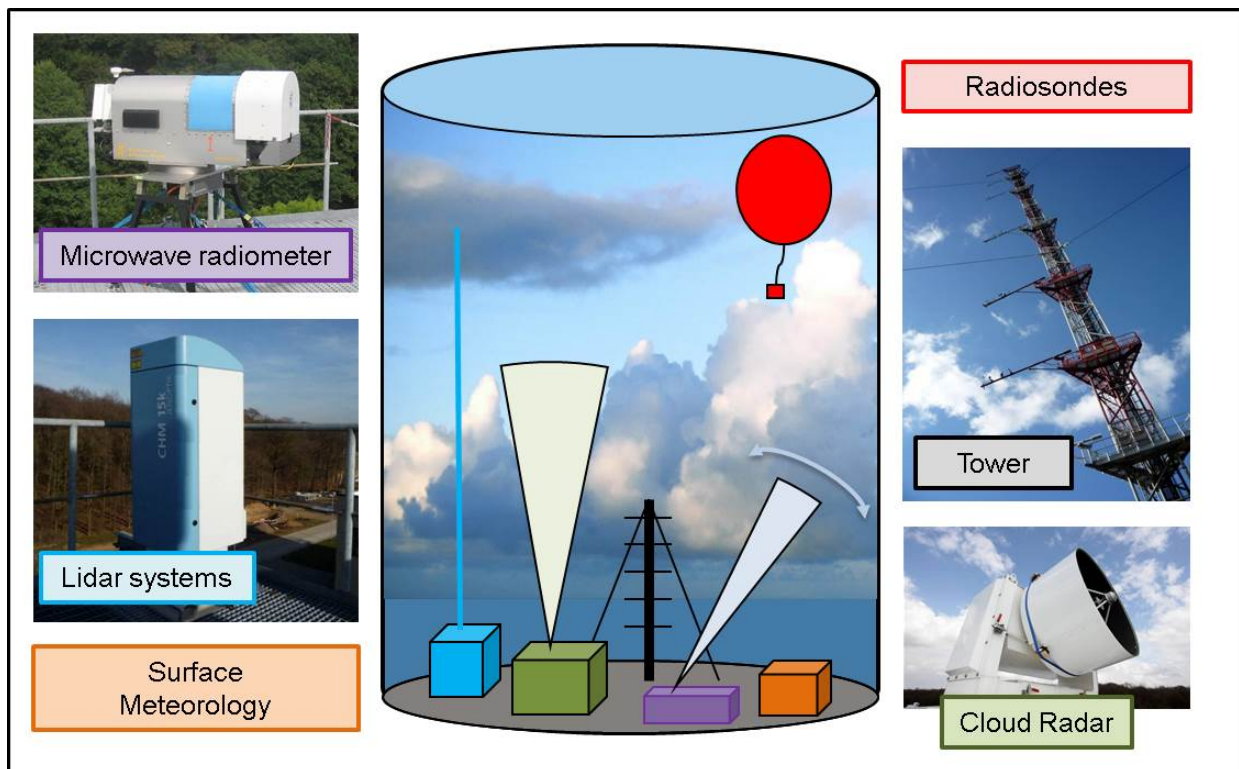


Figure 1: Possible Supersite instrumentation (instrument examples from JOYCE Supersite, <http://www.geomet.uni-koeln.de/forschung/joyce/>)

In addition to the local measurements at the Supersites, the archive includes so called full-domain observations. Actual the archive includes data sets of different satellite based instruments, like MODIS on satellites AQUA and TERRA, and SEVIRI on the MeteoSat second generation (MSG) satellites. Partly the satellite data are (in part freely) available via other data center. The data in the SAMD archive are provided in the SAMD data format, with the focus on the regions of interest.

Second part of the full-domain observations are the data of ground based instrument networks. Actual the archive provides data of:

- the C-band Radar network of the German Weather Service (RADOLAN), with 5 min temporal and $1 \times 1 \text{ km}^2$ horizontal resolution,

- the Ceilometer network, and
- the GNSS (Global Navigation Satellite System) network.

In Figure 2 two examples of full domain data are shown: Precipitation from the C-band Radar network on a regular horizontal grid, and the cloud base altitude from the ceilometer network on an irregular grid, both for the same date and time.

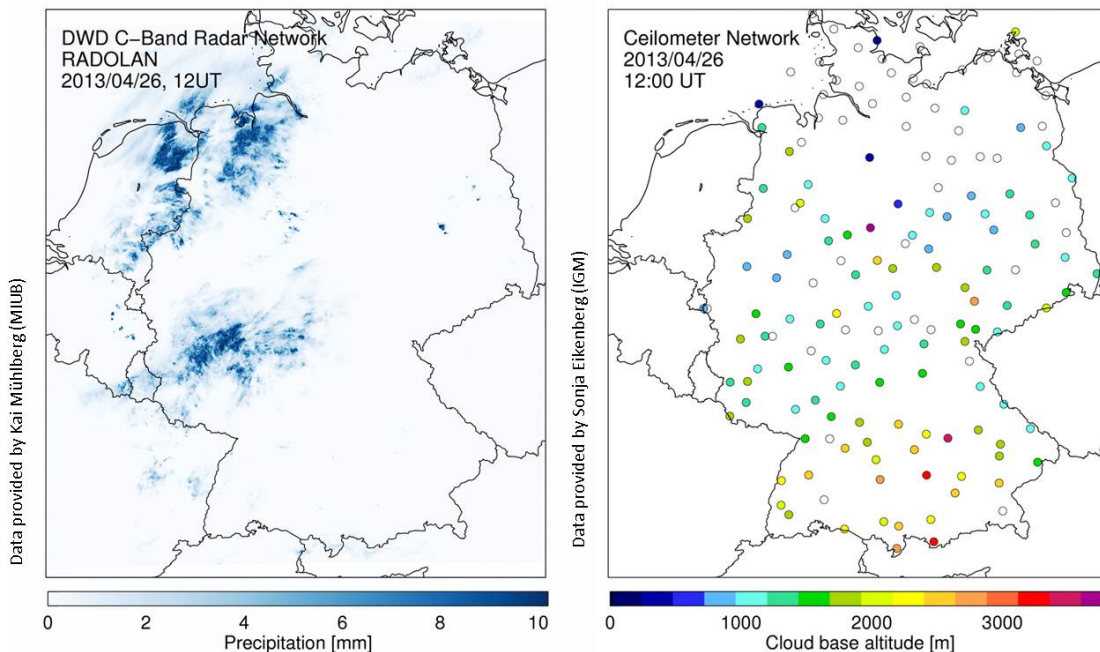


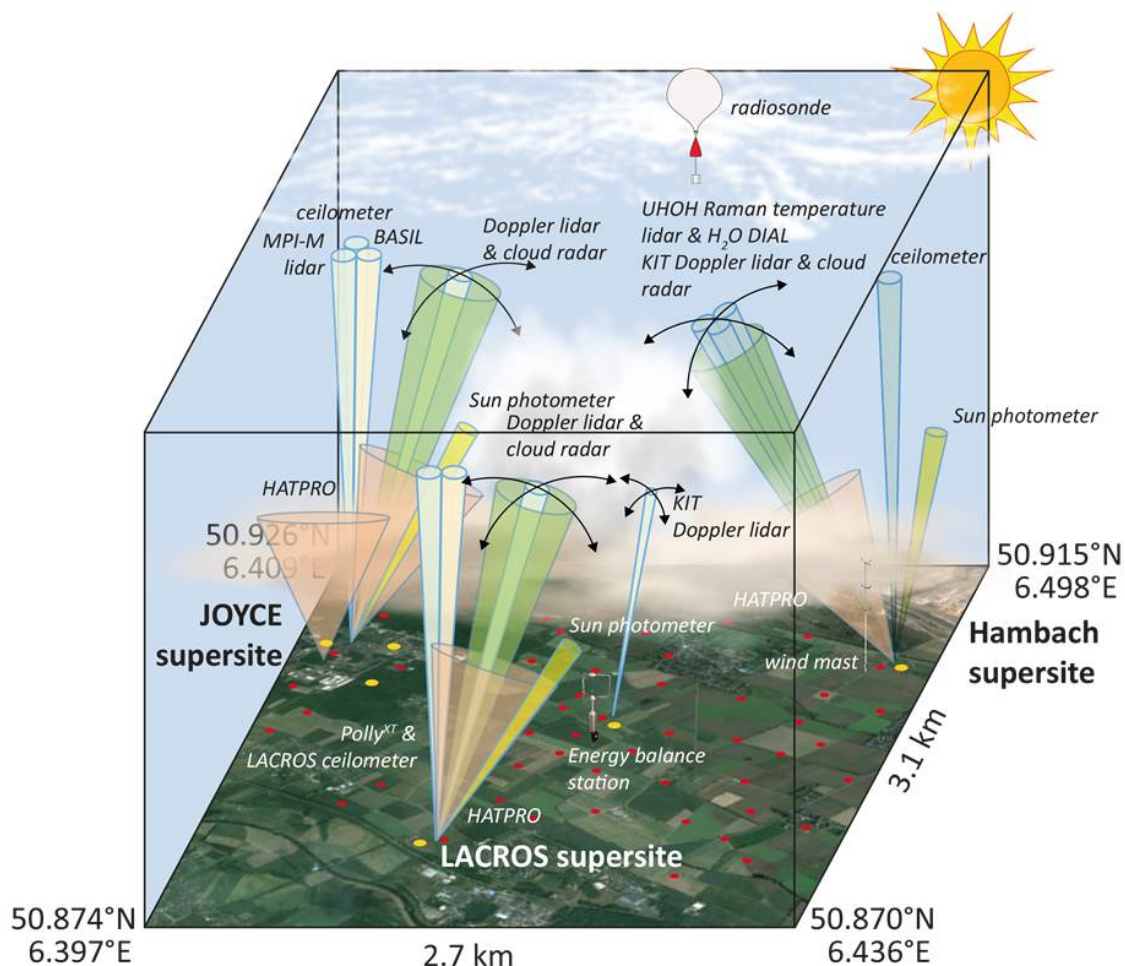
Figure 2: Examples of full-domain observation data. Left: Precipitation from C-band Radar network of the German Weather Service (RADOLAN) and right: Cloud base altitude from Ceilometer network, both for 26 April 2013, 12:00 UT.

2.2. Short-term measurements

In contrast to the long-term observations measurement campaigns and Intensive Observation Periods (IOP's) yield atmospheric data for a short time period and just at a special location, but with a high density of instruments. As part of HD(CP)² the HD(CP)² Observation Prototype Experiment (HOPE) took place in April and May 2013 at and around the Supersite JOYCE in Jülich. "HOPE was designed to provide a critical model evaluation at the scale of the model simulations and further to provide information on sub grid variability and micro-physical properties that are subject to parametrizations even at high-resolution simulations." Therefore the focus was "on the onset of clouds (activation) and precipitation (auto conversion) in the convective atmospheric boundary layer. By capturing the 3D cloud distribution the measurement can support the investigation of cloud-overlap and 3D radiative effects. In the absence of boundary layer clouds, the remote sensing instrument system was ideally suited to retrieve aerosol and cirrus cloud properties." (www.hdcp2.eu)

A second campaign, HOPE-Melpitz, took place in September 2013 in Melpitz, nearby Leipzig. This field experiment concentrated on the closure of the joint ground-based remote sensing techniques with the in-situ observations from ACTOS, the helicopter-based Aerosol and Cloud micro-physics and Turbulence Observation System. In addition, measurements of a large number of devices for the chemical and micro-physical

characterization of the surface-near aerosol were performed that provide additional boundary conditions for the planned coupling of the mobile Supersite LACROS and the ACTOS measurements. Beneath the data of "big" measurement campaigns, the archive includes the possibility of so called IOP's, Intensive Observation Periods. IOP's could be located at Supersites or in the middle of nowhere, important is just the intensive observation of one or more instruments for a given time interval.



Graphic: Patric Seifert & Katja Schmieder, Tropos

Figure 3: HD(CP)² Prototype Experiment (HOPE): 3D instruments map. HOPE took place in Spring 2013 around the Supersite JOYCE.

3. Data policy

Standardized Atmospheric Measurement Data - SAMD: Data Policy for Data Users

This policy holds for all data that is provided by the SAMD archive.

1. Creative Commons License

All data in the SAMD archive are licensed under the Creative Commons License CC-BY-NC-SA (<https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode>) as far as those conditions are not in any way modified by the following conditions or by any conditions specific to data (especially Cloudnet data, see 7.).

2. For non-commercial use only

Data from the SAMD archive may be used freely for academic research only (non-commercial use).

3. Audit

All downloads from the SAMD archive are auditable. Information will be used for statistical analyses, such as download statistics.

4. Publications

Articles, papers, or written scientific works of any form, based in whole or in parts on data supplied by the SAMD archive, must contain a reference including the title, author, and PID number to each used data set, as given in the meta data file(s).

5. Share alike

Any person extracting data from the SAMD archive will accept responsibility for informing all subsequent data users of these conditions.

6. Liability / Warranty

The data are delivered to the user without a warranty of any kind. The user is aware that the data were generated in keeping with the current state of science and technology.

7. Data originating from other data bases

Here, the data policy of the **source data base** (e.g. Cloudnet, ACTRIS), which is stored in the meta data of the data set, applies in addition to the SAMD policy. Please make sure that you use any SAMD data in agreement with the corresponding conditions of use.

Figure 4: The SAMD data policy (as has to be accepted by data provider and data user).

4. The SAMD Data Portal: samd-archive.org

All positive checked and uploaded data are available via the same Data Portal, independent on the storage server. The Data Portal is located at the Integrated Climate Data Center (ICDC) of the University of Hamburg: <https://icdc.cen.uni-hamburg.de/projekte/samd.html>. The Portal provides, beneath information about the file naming and conventions, the meta data of the different data set and the possibility to browse the archive for specific data:

- **Browse by Category**

Browse through the data sets in Long Term Observations, subdivided into Local and Full Domain Observations, and Short Term Observations, subdivided into different campaigns. Shown are all data sets of the given category, sorted alphabetically.

- **Browse by Selection menu**

According the "easy-to-use" approach, the data sets are sorted into the search tree (see Figure 5). The Selection menu allows the search for special data sets for all users, independent on the users experience with measurement data and instruments.

- **Browse by TDS Catalog**

Alternative you could search directly the data archive catalog to find all data sets of specific instruments, variables and years. It gives you the chance to "root" through the archive, if you are a bit familiar with measurement instruments.

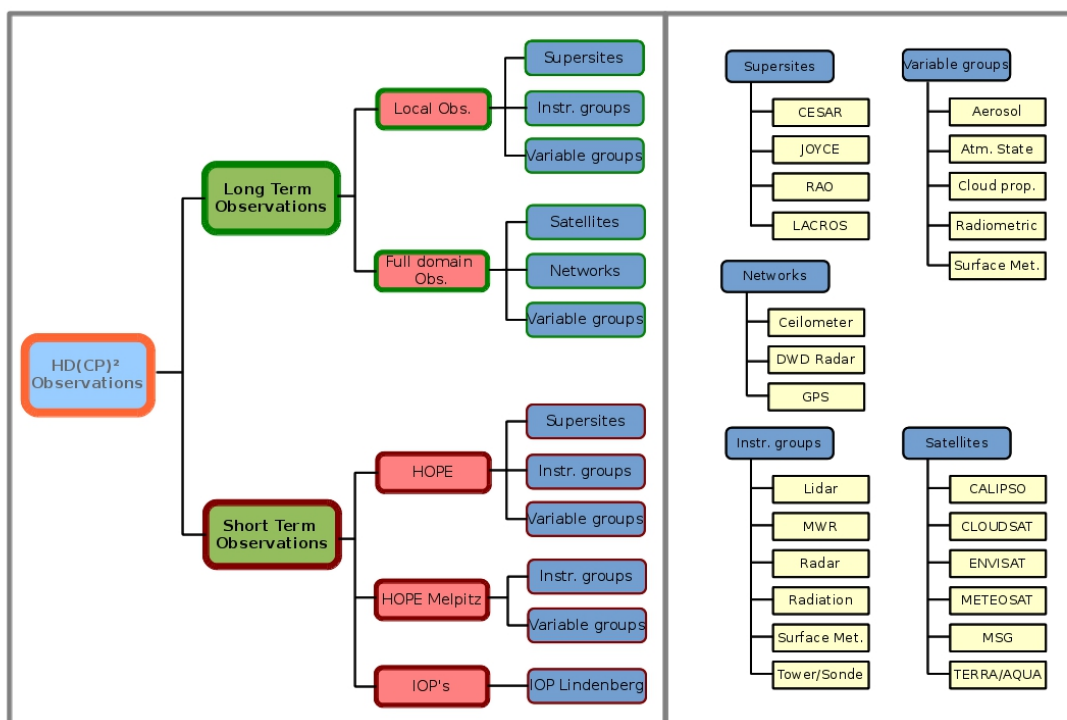


Figure 5: Search tree of HD(CP)² Data Portal

In Figure 6 the compact description of the Ceilometer data set from the Supersite JOYCE is shown as an example, as given by the Data Portal. The information are extracted from the Metadata-file of the data set.

Beneath the standardized compact information for each data set, based on the Metadata-Files, the Data Portal provides the specific link to both, the complete Metadata and the data set itself. The SAMD archive uses an Earth System Grid Federation (ESGF) OpenID as its authentication mechanism. For the download of data from one of the servers an OpenID registration is mandatory; the ESGF-Open-ID registration, e.g. of the German Climate Computing Center (DKRZ), <https://esgf-data.dkrz.de/user/add/?next=https://esgf-data.dkrz.de/projects/esgf-dkrz/>. After the registration and before the first download the user is once asked to accept the SAMD data policy. Hereafter the OpenID is sufficient for the data download.

Jenoptik CHM15k ceilometer data

Jenoptik CHM15k ceilometer data

Access

[View meta data](#)

Get data via HTTP / OPeNDAP

To get all files use this wget command replacing your login credentials:

```
wget -nc -r -l3 -A.nc -R "catalog" -nd -l /thredds/fileServer/ /thredds/catalog/
'http://USERNAME.PASSWORD@data.hdcp2.uni-koeln.de/thredds/catalog/hdcp2/ceilo/00/any/01/sups
/joy/catalog.html'
```

Instrument 1

- Source: Jenoptik CHM15k ceilometer: ID CHM120109, serlom TUB120017, software version 12.12.12.13 0.719
- Descriptive Instrument location: The ceilometer is installed by the JOYCE, on the roof of the Institute for Energy and Climate Research, Research Center Juelich.

Global Information

- Level: 1
- Updated Version: 1 (new software)
- File format: NetCDF3
- Convention: CF-1.6
- Average File Size Uncompressed: 24 Mb
- File name: sups_joy_ceilo00_01_any_v{VV}[YYYYMMDDhhmmss].nc
- Start: 2013-08-09
- End: ongoing

Variables

Name	Dimension	CF stan- dard_name	long_name	Units
beta_raw	time,range		normalized range corrected signal	1
beta_raw_hr	time,range_hr		normalized range corrected signal in high resolution distance from lidar	1

Institution

Research Center Juelich, Institute for Energy and Climate research (IEK-8)

Contact Person(s)

Ulrich Loehnert (Loehnert@meteo.uni-koeln.de)

References

Licence

For non-commercial use only. This data is subject to the HD(CP)2 data policy to be found at hdcp2.zmaw.de and in the HD(CP)2 Observation Data Product Standard.

Figure 6: Meta data of Ceilometer data set from Supersite JOYCE as example for the Data Portal.

5. Naming conventions

5.1. The file naming

Files shall be named according to the following naming convention:

`<kkk>_<sss>_<instnn>_<lll>_<var>_<vnn>_YYYYMMDDhhmmss.nc`

where:

kkk	is the kind of measurement type: supersite, full domain or campaign data,
sss	supersite or owner institute of the instrument or distributor of data,
instnn	instrument or synergy product, retrieval algorithm plus numbering, starting with "00"
lll	level of data processing, starting with l1
var	variable name for level 2 and higher (for level 1 data, which may contain more than one variable use "any"),
vnn	version of data set, starting with v00), and
YYYYMMDD	the start of the data file as YYYY=year, MM=month, DD=day,
hhmmss	hour, minutes, seconds in UTC.

Example1: `sup_sjoy_ceilo00_l1_any_v00_20130101000000.nc`

which is the name for a data set of the supersite ("sup_s") JOYCE ("joy"), from the instrument Ceilometer ("ceilo00"), level 1 ("l1", means raw data), including several variables ("any" - just allowed for l1-data), 1st version of the data set ("v00") for January 1, 2013 0:00:00.

Example2: `hdfd_igmk_gnssnet00_l2_prw_v00_20130101000000.nc`

which stands for a HD(CP)² full domain data set ("hdfd") of the IGM, University of Cologne ("igmk"), measured by the ceilometers network ("gnssnet00"), processing level 2 ("l2"), is the path of integrated water vapor ("prw"), version 1 ("v00") for January 1, 2013 0:00:00.

Note that file names with empty or wrong fields will not be accepted, since the sorting into our database depends on the information given in the file names. The file names are unique, and defined by your information you provide for the Metadata-Files (Section 8). In the following, the different possible entries for the fields are listed.

5.1.1. The Measurement type (<kkk>)

In general the data sets are distinguished between Long-term and Campaign data, and the Long-term data in Local and Full-domain observation.

Table 3: List of abbreviations for the measurement type <kkk>

Kind of Measurement type	Abbreviation
Super site	sup_s
HD(CP) ² full domain observations	hdfd
HOPE campaign	hope
HOPE Melpitz campaign	hopm
IOP (Intensive observation period) + location	iopxxx
TRIPEX (TRIple-frequency and Polarimetric radar Experiment for improving process observation of winter precipitation)	tripex
HALO (High Altitude and LOng range) research aircraft	halo

5.1.2. The Supersites (<sss>, part 1)

Supersite means here an atmospheric measurement site, which is able to yield long-term observations and has as minimum instrumentation a lidar, a microwave radiometer, and a cloud radar (plus basic meteorological measurements, like 2m temperature and humidity, air pressure, precipitation).

Table 4: List of abbreviations for the Supersites <sss>

Supersites	Abbreviation
BCO - Barbados Cloud Observatory	bco
CESAR - Cabauw Experimental Site for Atmospheric Research	ces
JOYCE - Jülich ObservatorY for Cloud Evaluation	joy
KITCube - Mobile facility of the Karlsruhe Institute for Technology	kit
LACROS - Leipzig Aerosol and Cloud Remote Observations System	lac
LMU - Ludwig-Maximilians-Universität München	lmu
RAO - Richard Assman Observatory Lindenberg	rao
UFS - Environmental Research Station Schneefernerhaus	ufs

5.1.3. The Institutions (<sss>, part 2)

Table 5: List of abbreviations for the Institutes <sss>

Owner Institute	Abbreviation
Institute of Energy and Climate, Research Center Juelich	fzj
Institute of Geophysics and Meteorology, University of Cologne	igmk
Institute for Meteorology and Climate Research, Karlsruhe Institute of Technology	kit
Institute for Meteorology, University of Leipzig	lim
Institute for Physics and Meteorology, University of Hohenheim	ipm
Institute for Space Sciences, Free University Berlin	fub
Leibniz Institute for Tropospheric Research, Leipzig	trop
Max-Planck-Institute for Meteorology, Hamburg	mpi
Meteorological Institute, University of Bonn	miub
Meteorological Institute, University of Hamburg	uhh
Meteorological Institute at the Ludwig-Maximilians-University Munich	lmu
Richard Assmann Observatory, Lindenberg	rao
Royal Netherlands Meteorological Institute	knmi
University of Basilicata - Potenza	unibas

Table 6: List of abbreviations for HALO campaign data <sss>

HALO campaign	Abbreviation
NARVAL North	narn
NARVAL South	nars
NARVAL 2	nar2

5.1.4. The Instruments (<instnn>)

If your institute has more than one instrument of the same type, the number (nn) behind the instrument tag is used to distinguish the instruments. E.g., if you have two ceilometers, the <inst>-tags would be ceilo00 and ceilo01. If only one instrument is present the number is 00 per default.

If your instrument has a specific name/acronym it is commonly known under and you would like to name your files accordingly, please contact the Observation Expert Users, so we can eventually include it.

Table 7: List of abbreviations for Instruments for Local observations <instnn>

Instrument or synergy/retrieval	Abbreviation
ACTOS (Airborne Cloud Turbulence Observation System)	actos
Ceilometer	ceilo
Cloud radar	cr
Cloud camera	ccam
DIAL (Differential Absorption Lidar)	dial
Disdrometer	dm
Doppler lidar	dlid
scanning with varying elevation and azimuth (custom)	dlidCUST
measured vectors from where wind profiles can be estimated	dlidDBS
scanning with varying elevation (range height indicator)	dlidRHI
scanning with varying azimuth	dlidVAD
zenith only	dlidST
GNSS (Global Navigation Satellite System)	gnss
Infrared thermometer	irt
scanning	irtS
Infrared spectrometer	irs
Meteorological data (near surface)	mets
Meteorological tower data	mett
Micro Rain Radar	mrr
Microwave radiometer	mwr
boundary layer scan new	mwrBL
Pyranometer	pyr
Pyranometer network	pyrnet
Pyrgeometer	pyrg
Radar wind profiler	rwp
Radiation data (surface energy balance system)	sebs
Radio sounding	sonde
Raman lidar	rlid
Sodar	sodar
UV lidar new	uvlid
Weather Radar, S-Band	wrs
Weather Radar, X-Band	wrx
Predefined instrument combinations	
Cloudnet (MWR + CR + Lidar)	cln
DIAL + Raman Lidar + Doppler Lidar	drdl
MWR + CR + xxx	ipr
MWR + CR + IPT	ipt
Raman Lidar + MWR	rlmwr

Instrument or synergy/retrieval	Abbreviation
Triple frequency Cloud Radar	trcr

Table 8: List of abbreviations for Instruments for Full-domain observations <instnn>

Instrument or synergy/retrieval (on satellite)	Abbreviation
AATSR/MERIS synergy	aatrs
AMSU-B/MHS on METOP/NOAA series	amsu
CALIOP on CALIPSO	caliop
Ceilometer network	cmnet
CPR on CloudSat	cpr
DWD C-Band Doppler radar network	drnet
GNSS network	gnssnet
MERIS on ENVISAT	meris
MODIS on TERRA and AQUA	modis
MSG Radar composite	msg
SEVIRI on Meteosat	seviri
Predefined instrument combinations	
Ceilometer Network + Seviri	cmnetsev
CALIOP + CPR	calcpr

5.1.5. The Level (<lll>)

Several versions/stages of processing at one data level may be indicated with the version tag <vnn> (see below). Note: In some exceptional cases it is reasonable to provide more than one level in your data put the highest level into the file name (e.g. CloudSat data). Please contact the Observation Expert Users before providing more than one level in one file.

Table 9: Level description <lll>

Level	Abbreviation
Level 0: Raw data as produced by the instrument (NOT in SAMD).	l0
Level 1: Raw data in standardized form (processed). Processing may include quality controls or calibration factors.	l1
Level 2: Quality-tested data. Value-added products, means quantities which are derived from the instrument data and not directly measured (e.g. wind vector from doppler lidar system).	l2
Level 3: Quality-tested data which combine different instruments (without model data).	l3
Level 4: Quality-tested data regridded to a regular grid and/or synergy with model data.	l4

Several versions/stages of processing at one data level may be indicated with the version tag <vnn> con-

tained in the file name. The first version has the tag v00. For each subsequent version of a dataset the tag will be increased by 1. For example, if you time- or space average your level 1 data, or if you reprocess your level 2 data with an improved processing routine the version tag will be increased by one for the resulting dataset.

The <var>-tag in the file name corresponds to the output variable name of the quantity (see Section 5.2). Level 1 data files may contain more than one variable. In this case, please use "any" for the tag. Note that the "any"-tag is exclusively reserved for level 1 data. For all other levels one variable/product per dataset is aimed at.

5.1.6. The version numbering (<vnn>)

Each data set starts with version number "v00". Please note that each change in your data like a new processing, a fixed bug in the processing, a new position of the instrument, a significant new instrument software, and so on, implicates a new data set with a new file name, means a new version number. Changes has to be explained in the NetCDF-header as global attribute "History" (see Sec. 6.1).

5.2. The variable naming

In the following, output variables/products are defined. Given is the CF standard name if it is existent, the output variable name, which is identical with the <var>-tag for the file name, and the canonical unit for the variable. Note that there are some level 1 variables, that must not appear in the file name, they are marked accordingly. If no standard_name is defined, the long_name attribute becomes mandatory for the description of the variable in the NetCDF-file (see Section 6). In that case, the standard_name attribute should be left out completely because the data set will fail the automated checking if an empty standard_name is present. For some variables long_names have already been suggested as common long_names within the project. Please find a alphabetical list, sorted according the <var>-tag name (variable abbreviation) in Appendix C and sorted according the CF-standard_name and long_name attribute respectively in Appendix D.

Table 10: Variable description <var>

Variable	CF-Standard_name or, if not in CF: (suggested long_name)	Abbreviation / filename <var>	Unit
Absolute humidity	mass_concentration_of_water_vapor_in_air	hua	kg m-3
Aerosol Angström exponent	angstrom_exponent_of_ambient_aerosol_in_air	aae	1
Aerosol layer heights	(aerosol layer heights)	zal	m
Aerosol number concentration	number_concentration_of_ambient_aerosol_in_air	conccn	m-3
Aerosol optical thickness	atmosphere_optical_thickness_due_to_ambient_aerosol	aot	1
Aerosol particle extinction coefficient	volume_extinction_coefficient_in_air_due_to_ambient_aerosol	apec	m -1
Altitude above mean sea level	altitude	zsl	m
Atmospheric boundary layer height	atmosphere_boundary_layer_thickness	zmla	m
... derived from wind profile	(mixing layer height from vertical wind profile)	zmlaw	m
... derived from aerosol profile	(mixing layer height from aerosol profile)	zmlaa	m
Attenuated backscatter coefficient	volume_attenuated_backwards_scattering_function_in_air	beta	m-1 sr-1
Backscatter intensity	(backscatter intensity)	intensity	1
Brightness temperature	brightness_temperature	tb	K
Buoyancy	(buoyancy)	buoy	N

Variable	CF-Standard_name or, if not in CF: (suggested long_name)	Abbreviation / filename <var>	Unit
CAPE	specific_convective_available_potential_energy	cape	J kg-1
CIN	atmosphere_convective_inhibition	cin	J kg-1
Cloud base altitude	cloud_base_altitude	zcb	m
Cloud base pressure	air_pressure_at_cloud_base	pcb	Pa
Cloud base temperature	(cloud base temperature)	tcb	K
Cloud fraction (total)	cloud_area_fraction	clt	1
Cloud ice content (height resolved)	mass_fraction_of_cloud_ice_in_air	cli	1
Cloud liquid water content (height resolved)	mass_concentration_of_cloud_liquid_water_in_air	cllw	kg m-3
Cloud mask	(cloud mask)	clm	1
Cloud optical thickness	atmosphere_optical_thickness_due_to_cloud	otc	1
Cloud optical thickness due to ice water	(atmosphere optical thickness due to cloud ice water)	otcli	1
Cloud optical thickness due to liquid water	(atmosphere optical thickness due to cloud liquid water)	otclw	1
Cloud thickness (liquid clouds)	thickness_of_liquid_water_cloud	cth	m
Cloud top altitude (geometric height above geoid)	cloud_top_altitude	zgct	m
Cloud top height (distance above surface)	height_at_cloud_top	zth	m

Variable	CF-Standard_name or, if not in CF: (suggested long_name)	Abbreviation / filename <var>	Unit
Cloud top pressure	air_pressure_at_cloud_top	pct	Pa
Cloud top temperature	air_temperature_at_cloud_top	tct	K
Cloud water content (height resolved)	mass_fraction_of_cloud_liquid_water_in_air	clw	1
Convective precipitation	convective_precipitation_amount	prcon	kg m-2
Doppler velocity	<i>(doppler velocity away from instrument)</i>	dv	m s-1
Drizzle liquid water (height resolved)	mass_concentration_of_drizzle_in_air	drlw	kg m-3
Drop number concentration	<i>(drop number concentration)</i>	dnc	m-3 mm-1
Effective radius of cloud ice particles	<i>(effective radius of cloud ice particles)</i>	reffcli	m
Effective radius of cloud liquid particles	<i>effective_radius_of_cloud_liquid_water_particle</i>	reffclw	m
Frequency band of sensor	<i>sensor_band_central_radiation_frequency</i>	freq_sb <i>(is not valid as <var>-tag)</i>	s-1
Frozen phase water content (height resolved)	<i>(mass concentration of frozen water in air)</i>	iwc	kg m-3
Geopotential height	geopotential_height	zg <i>(is not valid as <var>-tag)</i>	m

Variable	CF-Standard_name or, if not in CF: (suggested long_name)	Abbreviation / filename <var>	Unit
Height	height	height or zag (is not valid as <var>-tag)	m
Humidity mixing ratio	humidity_mixing_ratio	humr	1
Ice crystal number concentration	number_concentration_of_ice_crystals_in_air	dnccli	m-3
Latent heat flux	surface_upward_latent_heat_flux	hfls	W m-2
Latitude	latitude	lat	degree_north
Liquid water content (height resolved; cloud + rain)	mass_concentration_of_liquid_water_in_air	lwc	kg m-3
Longitude	longitude	lon	degree_east
LW broadband downwelling radiation (surface)	surface_downwelling_longwave_flux_in_air	rlds	W m-2
LW broadband upwelling radiation (surface)	surface_upwelling_longwave_flux_in_air	rlus	W m-2
Normalized Difference Vegetation Index	normalized_difference_vegetation_index	ndvi	1
Normalized range corrected signal (lidar)	(normalized range corrected signal) OR (range corrected backscatter signal)	beta_raw (is not valid as <var>-tag)	1
Nyquist frequency	(nyquist frequency)	nqf	s-1
Nyquist velocity	(nyquist velocity)	nqv	m s-1

Variable	CF-Standard_name or, if not in CF: (suggested long_name)	Abbreviation / filename <var>	Unit
Particle backscatter coefficient	(particle backscatter coefficient)	pbc	m-1 sr-1
Particle depolarization ratio	(particle depolarization ratio)	pdr	1
Particle extinction coefficient	(particle extinction coefficient)	pec	m-1
Particle lidar ratio	(particle lidar ratio)	plr	sr
Path integrated attenuation	(path integrated attenuation)	pia	dB
Path of integrated cloud liquid water	atmosphere_mass_content_of_cloud_liquid_water	clwvi	kg m-2
Path of integrated ice water	atmosphere_mass_content_of_cloud_ice	ciwvi	kg m-2
Path of integrated water vapor	atmosphere_mass_content_of_water_vapor	prw	kg m-2
Precipitation	precipitation_amount	precip	kg m-2
Pressure	air_pressure	pa	Pa
Radar co-polar correlation function	(radar co-polar correlation function)	rhohv	1
Radar differential reflectivity	(radar differential reflectivity)	zdr	dBZ
Radar integrated differential phase	(radar integrated differential phase)	phidp	deg
Radar linear depolarization ratio	(radar linear depolarization ratio)	ldr	dB

Variable	CF-Standard_name or, if not in CF: (suggested long_name)	Abbreviation / filename <var>	Unit
Radar reflectivity factor	equivalent_reflectivity_factor	dbz	dBZ
Radar specific differential phase	(radar specific differential phase)	kdp	deg km-1
Radar spectral skewness	(radar spectral skewness)	rsk	1
Radar spectral width	(radar spectral width)	sw	m s-1
Radial velocity	radial_velocity_of_scatterers_away_from_instrument	rv	m s-1
Rain rate	rainfall_rate	rr	m s-1 or mm s-1
Range	(distance from sensor to center of each range gates along the line of sight)	range	m
Range corrected backscatter signal	(range corrected backscatter signal) OR (normalized range corrected signal)	beta_raw (is not valid as <var> -tag)	1
Reflected solar spectral radiance	downwelling_radiance_per_unit_wavelength_in_air alias: downwelling_spectral_radiance_in_air	rssr	W m-2 m-1 sr-1
Relative humidity	relative_humidity	hur	1
Sea surface temperature	sea_surface_temperature	sst	K
Sensible heat flux	surface_upward_sensible_heat_flux	hfss	W m-2
Sensor azimuth angle	sensor_azimuth_angle	azi	degree
Sensor azimuth angle velocity	(sensor azimuth angle velocity)	aziv	degree s-1
Sensor elevation angle	(sensor elevation angle)	ele	degree
Sensor elevation angle velocity	(sensor elevation angle velocity)	elev	degree s-1

Variable	CF-Standard_name or, if not in CF: (suggested long_name)	Abbreviation / filename <var>	Unit
Soil heat flux (downward)	downward_heat_flux_in_soil	hfsoil	W m-2
Solar radiation flux (atmosphere)	downwelling_shortwave_flux_in_air	rsd	W m-2
Specific humidity	specific_humidity	hus	kg kg-1
Station number new	(station number)	stat (is not valid as <var>-tag)	1
Surface albedo	surface_albedo	surfalb	1
Surface longwave emissivity	surface_longwave_emissivity	sle	1
Surface temperature	surface_temperature	ts	K
SW broadband downwelling radiation (surface)	surface_downwelling_shortwave_flux_in_air	rsds	W m-2
SW broadband upwelling radiation (surface)	surface_upwelling_shortwave_flux_in_air	rsus	W m-2
Temperature	air_temperature	ta	K
Time	time	time (is not valid as <var>-tag)	seconds since 1970-01-01 00:00:00
Total downwelling radiation	(total downwelling radiation)	rtd	W m-2
Total upwelling radiation	(total upwelling radiation)	rtu	W m-2

Variable	CF-Standard_name or, if not in CF: (suggested long_name)	Abbreviation / filename <var>	Unit
Vertical velocity	upward_air_velocity	w	m s-1
Visibility	visibility_in_air	vis	m
Volume depolarization ratio	(volume depolarization ratio)	vdr	1
Volume extinction coefficient	(volume extinction coefficient)	vec	1
Wavelength of radiation	radiation_wavelength	wl	m
Wind direction	wind_from_direction	wdir	degree
Wind speed	wind_speed	wspeed	m s-1
Wind speed maximum (gust)	wind_speed_of_gust	wspeed_max (is not valid as <var> -tag)	m s-1
Zenith angle of beam direction	zenith_angle	zenith	degrees

For some combinations of instruments and variables (e.g. wind components) it makes sense to allow an exception of the rule of just one variable per file for higher level data. For those cases the files should be named with a variable group instead of "any" or one variable.

Table 11: Conventions for variable groups

Variable group	Comments	filename <var>
Aerosol classification	for Cloudnet; file includes additionally model data	aclass
Target classification	for Cloudnet; file includes additionally model data	class
Wind	file includes wind components, direction, etc.	wind
Cloud	Cloud related variables (e.g. for cmnetsev: zcb+zgct)	cloud

6. NetCDF conventions

As described before, the HD(CP)² observation data archive provides just one file format: NetCDF, both in version 3 and 4. In the following the minimum meta-data, which have to be provided in each NetCDF-file as so called "Global Attribute" are described. Also, the technical structure of the coordinate system and variables is described in this chapter, which is widely identical with the structures found in the CF Standard 1.6 (<http://cf-pcmdi.llnl.gov/>).

6.1. Global Attributes

The global attributes apply for the whole NetCDF rather than any particular variable. The SAMD standard defines some standardized global attributes, given in Table 12. Please note that **the order as well as the spelling of the global Attributes is binding**, including upper/lower case, and the usage of underscores instead of spaces in the attributes name. All attributes has to be filled, except "Comments" (in the case of no content, please fill in "none"). The attribute **History** is important for the versioning of the datasets. In the case of the first version of a dataset information about used programs and routines has to be filled in. If a higher version number is necessary (please see Sec. 5.1.6) the information about the reason has to be characterized.

Table 12: Global Attributes contained in the NetCDF-file header

Attribute	Content and Explanation (has to be adjusted to the data set!)
Title	Short Title including Instrument and content of data set. (The title will be used as a description of the data set in the web portal of the archive. So, please chose the title carefully.)
Institution	Owner Institution or distributor of data set
Contact_person	<Name>, <email>
Source	Instrument(s) (and its software version, if important for interpretation of the data)
History	First version: How is the data set processed? Programs or routines with version number; For higher version numbers: Differences to the last version of the dataset!
Dependencies	just in case of higher level products: <file name> (without date) of the depending data set or "external" (for all data sets not archived in the data base)
Conventions	basing on CF-1.6
Processing_date	<YYYY-MM-DD, hh:mm:ss > (time the file was produced)
Author	<name>, <email>
Comments	Miscellaneous Information about your dataset, otherwise none
Licence	For non-commercial use only. This data is subject to the SAMD data policy to be found at www.icdc.cen.uni-hamburg.de/projekte/samd.html and in the SAMD Observation Data Product standard.

As an example, the global attributes of the data file `sup_s_joy_ceilo00_l1_any_20130422000000.nc`, so

for the Ceilometer data, level 1, at Supersite JOYCE for 22 April 2013, are (as given in NetCDF header):

```
// global attributes:
:Title = "Jenoptik CHM15k ceilometer data" ;
:Institution = "Research Center of Jülich, Institute for Energy and Climate (IEK-8)" ;
:Contact_person = "Ulrich Loehnert (Loehnert@meteo.uni-koeln.de)" ;
:Source = "Jenoptik CHM15k ceilometer: ID CHM120109, serlom TUB120017, software version
12.03.1 2.13 0.559" ;
:History = "Data processed with readin_jenoptik_ceilodata_hdcp2_0.1.0 and
write_jenoptik_ceilonc_hdcp2_earlysoftware_0.1.0 by University of Cologne" ;
:Dependencies = "external" ;
:Conventions = "CF-1.6 where applicable" ;
:Processing_date = "2014-03-14 15:03:06 (CET)" ;
:Author = "Anne Hirsikko (anne.hirsikko@fmi.fi)" ;
:Comments = "none" ;
:Licence = "For non-commercial use only. This data is subject to the HD(CP)2 data policy to be
found at www.hdcp2.eu and in the HD(CP)2 Observation Data Product standard." ;
```

6.2. Instrument meta-data for ground based instruments

For each instruments the following information has to be included into the file. In the case of just one instrument the information has to be stored as variables without dimension, otherwise, the variables are stored as one dimensional fields, length equal number of instruments (e.g. ground based lidar network).

- longitude of instrument position in degree_east (float)
- latitude of instrument position in degree_north (float)
- height of instrument above ground (float)
- altitude of instrument above mean sea level in meter (float)

E.g., the instrument meta-data for rain radar on Geomatikum's roof in Hamburg (as given in NetCDF header) is given as :

```
variables:
float lat ;
    lat:standard_name = "latitude" ;
    lat:comments = " Latitude of instrument location" ;
    lat:units = "degrees_north" ;
lat = 53.569 ;
float lon ;
    lon:standard_name = "longitude" ;
    lon:comments = " Longitude of instrument location" ;
    lon:units = "degrees_east" ;
lon = 9.974 ;
```

```
float zsl ;
    lat:standard_name = "altitude" ;
    lat:comments = " Altitude of instrument above mean sea level" ;
    lat:units = "m" ;
zsl = 124.8 ;
```

6.3. Dimensions

A variable may have any number of dimensions and the dimensions must all have different names. The dimensions of the variable define the axes of the quantity it contains. Dimensions other than those of space and time may be included. Especially the dimensions for the variables time, height, range, lon, and lat should be named like the variable itself.

6.4. Coordinate system

The structures of coordinates and variables have to follow the CF Standard 1.6. We only provide some example coordinates here, a variety of example coordinate systems can be found in the CF Standard and for a detailed description we refer to its documentation at <http://cfconventions.org/documents.html>. The dimensions, as far as present, should appear in the following order: time, then height, then latitude, then longitude (or time, height, range if applicable):

```
time [, height or range] [, lon] [, lat]
```

Please see the example for variables below for details, or have a look into the CF convention: <http://cfconventions.org/Data/cf-conventions/cf-conventions-1.6/build/cf-conventions.html> Please note that time is in any case the leftmost dimension!

Time

Time has to be given as seconds (minutes, hours, days) since 1 January 1970, 00:00:00, given as UTC (e.g., "minutes since 1970-01-01 00:00:00"). As usual, time has to be stored with as much precision as necessary, not more! (E.g., for a measurement with 10 s resolution it doesn't make sense to give time as double variable with several decimal places.) A boundary array is mandatory for the time to have a unique description of the data's time averaging. The boundary array has one dimension more than your time array. For time(i) you have time_bnds(i,0) and time_bnds(i,1) containing the start and end of the averaging interval, respectively. Please note, the time boundary variable corresponds to cell boundaries in CF standard 1.6. Thus, in the NetCDF-file time has the following structure:

dimensions:

```
time = 12345; ("UNLIMITED" as possible)
nv = 2 ; (Number of boundaries, in this case start and end of time interval => 2)
```

variables:

```
double time (time) ;
    time:standard_name = "time" ;
    time:units = "seconds since 1970-01-01 00:00:00" ; (in UTC)
    time:bounds = "time_bnds" ;
double time_bnds (time,nv)
```

In case of swath-based satellite data please define the first pixel time as variable `base_time` without `time_bnds` and give the difference of each pixel to `base_time` as variable `dttime`:

dimensions:

```
time = 12345; ("UNLIMITED" as possible)
```

variables:

```
double base_time ;
```

```
base_time:long_name = "reference time" ;
```

```
base_time:units = "seconds since 1970-01-01 00:00:00"; (in UTC)
```

```
base_time:description = "earliest time stamp of data field for both day and night" ;
```

```
float (or double) dttime (time) ;
```

```
dttime:long_name = "difference time" ;
```

```
dttime:units = "s" ;
```

```
dttime:description = "actual relative time of each pixel corresponding to the reference time  
(see variable base_time): dttime=pixel_time-reference_time" ;
```

IMPORTANT: For level 3 and higher level data sets, time has to be a **continuous** variable with **equidistant steps** for the whole day, means full 24 hours! Gaps in the measurement variable must be filled with the defined fill value! So, if you have e.g. a temporal resolution of one hour, your data set must include 24 time steps per file, always! This is an important step towards our "easy-to-use" approach, e.g. for the validation of model data.

Geographical Grid

The following coordinates might vary, depending on the instrument. The given examples show the most common ones. A boundary array for spatial coordinates is not requested, but recommend to provide if possible. This will specify the spatial resolution of the data and facilitate the comparison with model data.

dimensions:

```
lon=360;      to be adjusted to the instrument!
```

```
lat=180;     to be adjusted to the instrument!
```

```
nv = 2;
```

variables:

```
float lon (lon);
```

```
lon:standard_name = "longitude";
```

```
lon:units = "degrees_east";
```

```
lon:bounds = "lon_bnds"      may be skipped, if not:
```

```
float lon_bounds (lon,nv)
```

```
float lat (lat);
```

```
lat:standard_name = "latitude";
```

```
lat:units = "degrees_north";
```

```
lat:bounds = "lat_bnds"      may be skipped, if not:
```

```
float lat_bounds (lat,nv);
float zsl (lat,lon);
    zsl:standard_name = "altitude" ;
    zsl:long_name = "altitude above mean sea level" ;
    zsl:units = "m" ;
```

6.5. Variables

Variables must have a name, dimension, and the CF standard name (if standard name is existent, if not, please leave out the attribute completely), units, and a fill value. If no CF standard name is present, the `long_name` attribute is mandatory to describe the variable. We strongly suggest that common `long_name` attributes for the same variables are used within the HD(CP)² project! As a long-term aim it may be possible to include new standard names into the CF standard, which also match the requirements of the observational community. Suggestions following the guidelines given in the CF-standard² are welcome. Flags may be added to the variables using the attributes "flag_values", "flag_meanings", "flag_masks" and marked as "ancillary_variables" attribute to the variable itself, as described in the CF-standard³. Exemplary a variable may look like this (for the air temperature):

dimensions:

```
time = "UNLIMITED";
height=100
lat = 360;
lon = 360;
```

variables:

```
double ta (time,height,lat,lon); Note dimensions order!
    ta:standard_name = "air_temperature";
    ta:long_name = "height resolved air temperature";
    ta:_FillValue=-999.d ;
    ta:units = "K";
    ta:ancillary_variables = "ta_flag" ;
byte ta_flag (time,height,lat,lon)
    ta_flag:standard_name = "air_temperature status_flag" ;
    ta_flag:flag_values = 0b,1b,2b ;
    ta_flag:flag_meanings = "no_measurement good_quality bad_quality" ;
```

The order of dimensions, given for `ta` in the example, is how it should appear! Variables are listed in Section 5.2. If a variable is missing, please contact the Observation Expert User so it can be added to this document. We strongly suggest using the canonical units given above. However, according to the CF Standards, also units supported by UNIDATA's Uduunits⁴ package may also be accepted.

²<http://cfconventions.org/Data/cf-standard-names/docs/guidelines.html>

³<http://cfconventions.org/cf-conventions/v1.6.0/cf-conventions.html>

⁴<http://www.unidata.ucar.edu/software/udunits/udunits-2-units.html>

7. Metadata conventions

Well-documented data should enable other researchers to understand, use and reuse the data correctly. For this reason, every data set which is published by a SAMD archive must be associated with metadata in an appropriate form. We designed about 30 metadata descriptive elements to define our own standard, amongst others SAMD has adapted some well-established attributes from the data file's NetCDF header. The SAMD metadata files are encoded in the eXtensible Markup Language (XML).

7.1. The SAMD metadata web interface

All data producers should generate a separated file with detailed information about measurement and instruments. This file has to be created via the **SAMD web interface** for data providers: <http://hdcp2.uni-koeln.de>, which is hosted by the Regional Computing Center Cologne (contact: Erasmia Stamnas, hdcp2-admin@uni-koeln.de). Each data provider has to answer several questions in a specifically designed web form. The form mostly presents lists of selectable items, displayed within drop-down menus. Some fields allow free text. As an example, the SAMD-approved variables (see 5.2) can be picked from a list. At the end of the questionnaire the underlying application will transform the filled-in information into a valid XML document and provide it with a digital signature (hash key). Only signed metadata files are accepted for the data archive, ensuring they have been edited by the metadata editor. This guarantees compliance with the SAMD standard. The metadata editor can also be used for addition, modification, and removal of metadata, choose "upload/edit File" at the web interface to edit an old metadata file (with a valid hash).

Table 13: SAMD Metadata standard Part 1: XML elements at the SAMD metadata web interface

XML element	sub-element	Definition	Form
project		Project under which the data were obtained.	drop down menu
region		Measurement region (used for the generation of the PID of the data set).	drop down menu
dsName		Name of data set (5 parts).	drop down menu
fileAverageSize	size [unit] status	Average size of the dataset files in MB, uncompressed.	free text
datatype		Daily data file or exceptions (16-day, event).	drop down menu
provenance		Data provenance (description of the origin and purpose).	free text
contactPersons	list number		automated
	institution forename surname code phone email	Contact data of person responsible for data, person in permanent position if possible.	drop down menu and free text
keywordLists	list number		automated
	experimentType measurementType mainGroup variableGroup	Keywords describing the data according to the SAMD classification tree (tree supports 4 search levels currently - see Section 7.3).	drop down menu
temporalExtent	startDate endDate	Day the measurement started respectively ended, or if the measurement is continuous, ongoing.	drop down menu
Continued on next page			

XML element	sub-element	Definition	Form
resolution	temporal [unit] horizontal [unit] vertical [unit]	Data resolution	arithmetical value
location		Short description of measurement location and type of measuring instrument (single, group or network). In prose text, no acronyms. In addition may be used to give information about flight paths, orbits, etc.	free text
productDescription		Short description of the data product.	free text
limitations		Limits of measurement, environmental conditions under which the data were acquired (optional).	free text
references	list number		automated
	referenceType publicationDate	Specify references with form, date, author, title, abstract, and publishing media (optional).	drop down menu
	author title abstract publish		free text
instruments	list number		automated
	source	Name and type of instrument	free text
	instrumentLocation: - latitude [unit] - longitude [unit] - altitude [unit] - height [unit]	Attributes characterizing the measuring facilities: - declaration of instrument used for the measurement, - relative position or location.	numerical value
	instrument-Specification	Additional information for measuring facility or instrument, e.g. wavelengths of LIDAR, frequency of radar, etc. (optional)	free text

After the data set is uploaded to the checking server, the meta-information, which is already present in the NetCDF header of the data (global attributes and variable information), will be extracted and add to the metadata file automatically.

Table 14: SAMD Metadata standard Part 2: XML elements generated automatically from NetCDF header

XML element	sub-element	Definition	NetCDF header
subtitle		Additional title, short headline.	Title
dsAuthor	forename surname email	Person who processed the data. Name and email-address of person primarily responsible for making the resource.	Author
institution		Instrument institution or distributor of data.	Institution
conventions		Based on the climate and forecast metadata conventions (current CF-1.6).	Conventions
dependencies		Name of the parent data set, for dependent data sets archived in SAMD, otherwise 'external'.	Dependencies
versionHistory		Description of data processing, software used, what has changed compared to previous versions, etc.	History
licence		SAMD data policy (potential additional agreements).	Licence
comments		Other useful information about the data, information describing aspects of the primary data.	Comments
Continued on next page			

XML element	sub-element	Definition	NetCDF header
dimensions	time height nv	Dimensions of data contained in NetCDF file, see chapter 6.4	<i>dimensions</i>
variables	name dimension standard_name long_name units	Variables contained in NetCDF file.	<i>variables</i>

The SAMD metadata scheme provides mandatory, optional and automatically generated attributes. The final metadata file contains all elements listed in the table (see below). Please note that the **metadata element spelling** you find in the table is **binding!** Including upper case and lower case. Because of the future cooperation with the CERA database we have to limit the free text fields to 1024 characters (exceptions: versionHistory and productDescription, 2048 chars each). Also note that a significant change in data processing, a fixed bug, a new position of the instrument and so on, implicates a new dataset version with a new metadata file. You have to go through the process of creating new metadata and NetCDF header again and the version number of your data set is then increased by 1.

If you would like to suggest a new variable name, miss your institution/anything else in the menu lists, or if you are in doubt whether you should create a new data set version please contact the SAMD team (andrea.lammert@uni-hamburg.de) any time.

The final XML file includes all information and elements from Tables 13 and 14 and some automated generated elements. Please find the complete metadata attributes in Appendix B.

7.2. PID

Regarding sustainability one important part of the metadata is the Persistent Identity (PID), which is automated generated. The PID allows the unique identification of each single data set. Therefore the PID consists of the following information: The first group includes the institution, which assigns the PID. The second group characterizes the data type and the measurement region, whereas the last group includes the main part of the data set name. An example for the PID is:

de.koeln.rrzk/amd.de.sups/joy.cr00.l1.beta

with

- de.koeln.rrzk: the Regional Computing Center of the University of Cologne as the PID assigning institution,
- amd.de.sups: Atmospheric Measurement Data (AMD) in Germany (de) at a Supersite (sups), and
- joy.cr00.l1.beta: abbreviation of the Supersite JOYCE (joy), the instrument Cloud radar (cr), the level (l1), and the variable name (beta).

7.3. Keywords

As part of the meta data web interface, data provider are able to choose between different keywords to characterize and sort their data set and make it as easy as possible to find the data in the archive. The entire search tree is shown in Figure 5. Multiple entries (different keyword "paths") are explicitly welcome. Please keep in mind; by means of these keywords, the user of the SAMD archive is able to find your data at the data base (or not)!

8. Compliance check and data upload

Before you can provide your data to the databases you have to provide one example data file with a complete Metadata-file (XML format) which describes your data set and a check sum. The example data file and the XML-file will undergo a compliance check (see Fig. 7). The checks consists of two parts. The first one, the compliance check is done automatically by the Regional Computing Center in Cologne, organized by Erasmia Stamnas. The second part is the scientific check, which is done manually. Once the files pass the checks you can start uploading files to the SAMD archive. After a successful upload your data will be available via the SAMD Data Portal hosted at the Integrated Climate Data Centre of the University of Hamburg (<https://icdc.cen.uni-hamburg.de/projekte/samd.html> and in the data browser <http:samd-archive.org>). Thus the steps for data upload are:

1. Please read this product standard and the "Data approval – Step-by-Step" (See Appendix A) carefully!
2. Upload test data (1 netCDF file, 1 web-xml file, 1 checksum file - together) to the test directory of Cologne's server.
3. Write a short email with subject "data upload" to hdc2-admin@uni-koeln.de for the checks.
4. Wait for results of checking, correct your files according to the results.
5. Once the files have passed the quality checks and the Observation Expert User gives you a final "YES", you may start uploading your data into the upload directory. Write a short email with "final data upload" to hdc2-admin@uni-koeln.de before the data upload!

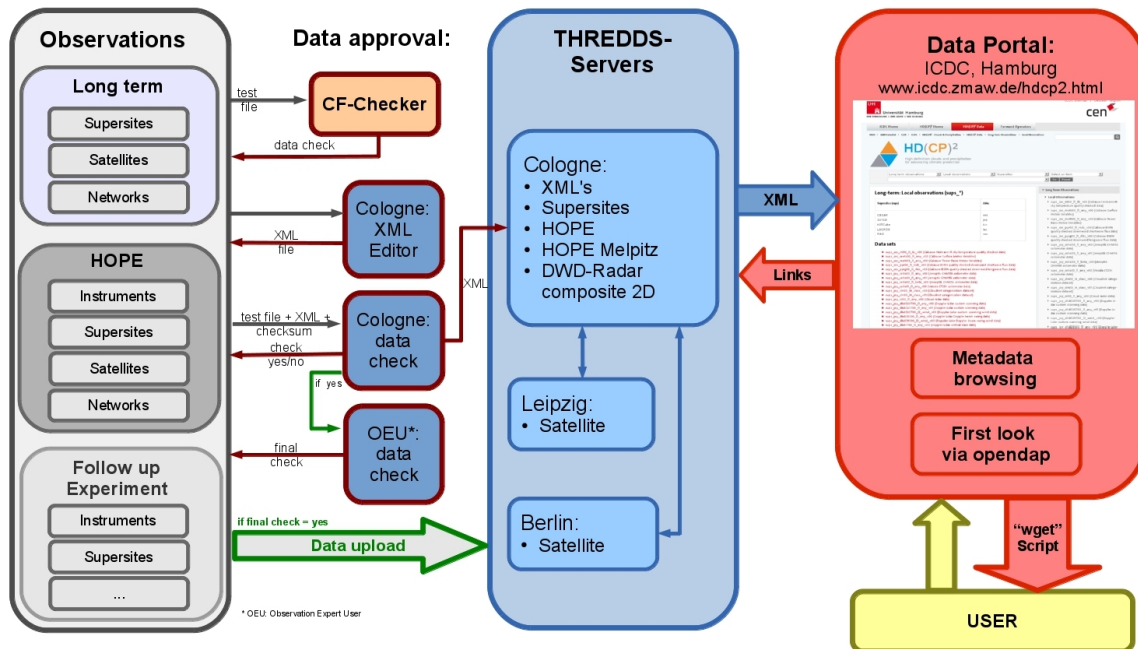


Figure 7: Overview of the data flow from the different providers, via the checking routine, the data storage (on distributed THREDDS-Servers), and the HD(CP)² observation data archive Data Portal to the user.

A. Data approval – Step-by-Step

by Verena Grützun, Andrea Lammert, Erasmia Stamnas, Volker Winkelmann

1. **Check your netCDF file with ncdump and ncview, and then with the CF-Convention Compliance Checker.**

There is an online checker, which you can find here:

<http://puma.nerc.ac.uk/cgi-bin/cf-checker.pl?cfversion=auto> If that's ok, proceed:

2. **Register (if you haven't done it already) at the University of Cologne.**

For registration follow the instructions mentioned in the SAMD Provider's Guide for Data Upload, which you find at <http://hdcp2.uni-koeln.de/>. (We perform some more checks for compliance with the SAMD Standard before you can start uploading your data sets).

3. **Create web-xml file for one data set at <http://hdcp2.uni-koeln.de/>.**

The password is available in our Observation Module Redmine at code.zmaw.de (all of you should have access, if not please contact us).

4. **Create SHA 256 checksum file for each datafile.**

The checksum files must have the name `<myfile>.sha256` (Linux command: `sha256sum`). Please regard: the file should only contain the checksum, no file name and no path!

5. **Upload: one data file per data set with its checksum (!) via secure copy (SCP) or secure file transfer (SFTP) into test-directory.**

```
scp -P 2222 <myfile>.nc <accountname>@hdcp2data.rrz.uni-koeln.de:/test
```

```
sftp -P 2222 <accountname>@hdcp2data.rrz.uni-koeln.de:/test <myfile>.nc
```

(files go into the test directory), remember to upload the corresponding Metadata file (XML) and the checksum file, too. They are crucial for the rest of the procedure and without them your data will not be checked. Please send a short email (subject: data upload) to hdcp2-admin@uni-koeln.de stating the names of the testfiles.

6. **Wait for the results from the SAMD check.**

You will get an email whether your data file passed the checks. **IF NO:** correct your errors, repeat the process from step 3 with the corrected netCDF file (same name, unless the name had to be corrected), you may have to create a new xml file, depending on your error (step 2).

The final "YES" will come from the observation expert users – please wait for our feedback. If you have it, proceed with step 7:

7. **Data Upload** Upload your data set (netcdf + checksum) via secure copy (SCP) or secure transfer (SFTP) into your hdcp2 directory at the cologne server. Please note that the storage can take a maximum of 250 GB at once. If you want to upload more, then please contact the SAMD admins (hdcp2admin@uni-koeln.de) before (see SAMD Provider's Guide for Data Upload provided at <http://hdcp2.uni-koeln.de/>).

All documents are available at the SAMD Data Portal (<https://icdc.cen.uni-hamburg.de/projekte/samd.html>). Among other the detailed **SAMD Provider's Guide for Data Upload** is given there. This documentation explains how to get your data on the servers of the Regional Computing Center in Cologne. It is also available via <http://hdcp2.uni-koeln.de/>. You can get the access data for Cologne's site by contacting estamnas@uni-koeln.de.

B. SAMD Metadata

Table 15: Overview of all SAMD Metadata attributes

No	XML element	Sub-element	Definition	Obligation	Source
General					
1	PID	(region)	Persistent identifier	automated	Script
2	language		Metadata describing language: english (en)	automated	Script
3	samd		Data Product Standard Version	automated	Script
4	project		Project under which the data were obtained	mandatory	Form: drop down menu
Dataset - global					
5	dsName		Name of data set (6 parts)	mandatory	Form: drop down menu
6	dsVersion		Version of data file	automated	Script
7	dsTitle		Short title of dataset including instrument/variable name	automated	Script
8	subtitle		Additional title, short headline	optional	NetCDF header: title
9	dsAuthor	forename surname email	Person who processed the data. Name and email-address of person primarily responsible for making the resource.	mandatory	NetCDF header: author
10	institution		Instrument institution or distributor of data	mandatory	NetCDF header
11	conventions		Based on the climate and forecast metadata conventions (current CF-1.6)	mandatory	NetCDF header
12	dependencies		Name of the parent data set, for dependent data sets archived in SAMD, otherwise 'external'	mandatory	NetCDF header
13	versionHistory		Description of data processing, software used, what has changed compared to previous versions, etc	recommended	NetCDF header: history
14	fileFormat		Specified format of data (e.g. NetCDF 3 or 4)	automated	Script
Continued on next page					

No	XML element	Sub-element	Definition	Obligation	Source
15	fileAverageSize	size [unit] status	Average size of the dataset files in MB, uncompressed	optional	Form: free text
16	datatype		Daily data file or exceptions (16-day, event)	mandatory	Form: drop down menu
17	level		Data processing level, with the lowest level of data being designated as O (raw data)	automated	Script
18	licence		SAMD data policy (potential additional agreements)	mandatory	NetCDF header
19	provenance		Data provenance (description of the origin and purpose)	recommended	Form: free text
20	contactPersons	list number [unlimited]		automated	Script
		institution		mandatory	Form: drop down menu
		forename surname code phone email	Contact data of person responsible for data, person in permanent position if possible		Form: free text
21	keywordLists	list number [unlimited]		automated	Script
		experimentType measurementType mainGroup variableGroup	Keywords describing the data according to the SAMD classification tree (tree supports 4 search levels currently)	mandatory	Form: drop down menu
		startDate endDate	Day the measurement started respectively ended, or if the measurement is continuous, ongoing	mandatory	Form: drop down menu
23	resolution	temporal [unit] horizontal [unit] vertical [unit]	Data resolution	optional	Form: arithmetical value

Continued on next page

No	XML element	Sub-element	Definition	Obligation	Source
24	location		Short description of measurement location and type of measuring instrument (single, group or network). In prose text, no acronyms. In addition may be used to give information about flight paths, orbits, etc.	recommended	Form: free text
25	productDescription		Short description of the data product	recommended	Form: free text
26	limitations		Limits of measurement, environmental conditions under which the data were acquired	optional	Form: free text
27	comments		Other useful information about the data, information describing aspects of the primary data	optional	NetCDF header
28	references	list number [unlimited]		automated	
		referenceType publicationDate author title abstract publish	Specify references with form, date, author, title, abstract, and publishing media.	optional	Form: drop down menu Form: free text
Datatype - specific					
29	instruments	list number [unlimited]		automated	
		source instrumentLocation: - latitude [unit] - longitude [unit] - altitude [unit] - height [unit] instrumentSpecification	Attributes characterizing the measuring facilities: - declaration of instrument used for the measurement, - relative position or location. Additional information for measuring facility/instrument, e.g. wavelengths of LIDAR, frequency of radar, etc.	mandatory	Form: free text Form: numerical value Form: free text
Continued on next page					

No	XML element	Sub-element	Definition	Obligation	Source
30	dimensions	time height nv	Coordinates contained in NetCDF file, see chapter 6.3	mandatory	NetCDF dimensions
31	variables	name dimension standard_name long_name units	Variables contained in NetCDF file, see chapter 6.5	mandatory	NetCDF variables

C. List of variable names

The table shows the list of variables allowed for <var>-tag in the file name in alphabetical order.

Table 16

<var>	Variable
aae	Aerosol Angström exponent
aot	Aerosol optical thickness
apec	Aerosol particle extinction coefficient
azi	Sensor azimuth angle
aziv	Sensor azimuth angle velocity
beta	Attenuated backscatter coefficient
buoy	Buoyancy
cape	CAPE
cin	CIN
ciwvi	Path of integrated ice water
cli	Cloud ice content (height resolved)
cllw	Cloud liquid water content (height resolved) in kg m ⁻³
clm	Cloud mask
clt	Cloud fraction (total)
clw	Cloud water content (height resolved) in kg kg ⁻¹
clwvi	Path of integrated cloud liquid water
conccn	Aerosol number concentration
cth	Cloud thickness (liquid clouds)
dbz	Radar reflectivity factor
dnc	Drop number concentration
dnccli	Ice crystal number concentration
drlw	Drizzle liquid water (height resolved)
dv	Doppler velocity
hfls	Latent heat flux
hfsoil	Soil heat flux (downward)
hfss	Sensible heat flux
hua	Absolute humidity
humr	Humidity mixing ratio
hur	Relative humidity
hus	Specific humidity
intensity	Backscatter intensity
iwc	Frozen phase water content (height resolved)
kdp	Radar specific differential phase
lat	Latitude
ldr	Radar linear depolarization ratio
lon	Longitude
lwc	Liquid water content (height resolved; cloud + rain)
ndvi	Normalized Difference Vegetation Index
otc	Cloud optical thickness
otcli	Cloud optical thickness due to ice water
otclw	Cloud optical thickness due to liquid water
pa	Pressure
pbc	Particle backscatter coefficient
pct	Cloud top pressure
pdr	Particle depolarization ratio
pec	Particle extinction coefficient
phidp	Radar integrated differential phase

Continued on next page

<var>	Variable
pia	Path integrated attenuation
plr	Particle lidar ratio
prcon	Convective precipitation
precip	Precipitation
prw	Path of integrated water vapor
reffcli	Effective radius of cloud ice particles
reffclw	Effective radius of cloud liquid particles
rhohv	Radar co-polar correlation function
rlds	LW broadband downwelling radiation (surface)
rlus	LW broadband upwelling radiation (surface)
rr	Rain rate
rsd	Solar radiation flux (atmosphere)
rsds	SW broadband downwelling radiation (surface)
rsk	Radar spectral skewness
rssr	Reflected solar spectral radiance
rsus	SW broadband upwelling radiation (surface)
rtd	Total downwelling radiation
rtu	Total upwelling radiation
rv	Radial or fall velocity (away from instrument)
sle	Surface longwave emissivity
sst	Sea surface temperature
stat	Station number new
surfalb	Surface albedo
sw	Radar spectral width
ta	Temperature
tb	Brightness temperature
tcb	Cloud base temperature
tct	Cloud top temperature
ts	Surface temperature
vdr	Volume depolarization ratio
vec	Volume extinction coefficient
vis	Visibility
w	Vertical wind velocity
wdir	Wind direction
wl	Wavelength of radiation
wspeed	Wind speed
zal	Aerosol layer heights
zcb	Cloud base altitude
zdr	Radar differential reflectivity
zgct	Cloud top altitude (geometric height above geoid)
zmla	Atmospheric boundary layer height
zmlaa	... derived from aerosol profile
zmlaw	... derived from wind profile
zsl	Altitude above mean sea level
zth	Cloud top height (distance above surface)

D. List of CF-standard_names

The table shows the list of used CF-standard_names and long_names, respectively.

Table 17

CF-standard_name OR long name	Variable	In CF?
aerosol layer height	Aerosol layer heights	n
air_pressure	Pressure	y
air_pressure_at_cloud_base	Cloud base pressure	y
air_pressure_at_cloud_top	Cloud top pressure	y
air_temperature	Temperature	y
air_temperature_at_cloud_top	Cloud top temperature	y
altitude	Altitude above mean sea level	y
angstrom_exponent_of_ambient_aerosol_in_air	Aerosol Angström exponent	y
atmosphere optical thickness due to ice water	Cloud optical thickness due to ice water	y
atmosphere_boundary_layer_thickness	Atmospheric boundary layer height	y
atmosphere_convective_inhibition	CIN	y
atmosphere_mass_content_of_cloud_ice	Path of integrated ice water	y
atmosphere_mass_content_of_cloud_liquid_water	Path of integrated cloud liquid water	y
atmosphere_mass_content_of_water_vapor	Path of integrated water vapor	y
atmosphere_optical_thickness_due_to_ambient_aerosol	Aerosol optical thickness	y
atmosphere_optical_thickness_due_to_cloud	Cloud optical thickness	y
atmosphere optical thickness due to cloud liquid water	Cloud optical thickness due to liquid water	n
atmosphere optical thickness due to cloud ice water	Cloud optical thickness due to ice water	n
backscatter intensity	Backscatter intensity	n
brightness_temperature	Brightness temperature	y
buoyancy	Buoyancy	n
cloud base temperature	Cloud base temperature	n
cloud mask	Cloud mask	n
cloud_area_fraction	Cloud fraction (total)	y
cloud_base_altitude	Cloud base altitude	y
cloud_top_altitude	Cloud top altitude (geometric height above geoid)	y
convective_precipitation_amount	Convective precipitation	y
distance from sensor to center of each range gates along the line of sight	Range	n
doppler velocity	Doppler velocity	n
downward_heat_flux_in_soil	Soil heat flux (downward)	y
downwelling_radiance_per_unit_wavelength_in_air <i>Alias: downwelling_spectral_radiance_in_air</i>	Reflected solar spectral radiance	y
downwelling_shortwave_flux_in_air	Solar radiation flux (atmosphere)	y
drop number concentration	Drop number concentration	n

Continued on next page

CF-standard_name OR long name	Variable	In CF?
effective radius of cloud ice particles	Effective radius of cloud ice particles	n
effective_radius_of_cloud_liquid_water_particle	Effective radius of cloud liquid particles	y
equivalent_reflectivity_factor	Radar reflectivity factor	y
geopotential_height	Geopotential height	y
height	Height	y
height_at_cloud_top	Cloud top height (distance above surface)	y
humidity_mixing_ratio	Humidity mixing ratio	y
latitude	Latitude	y
longitude	Longitude	y
mass concentration of frozen water in air	Frozen phase water content (height resolved)	n
mass_concentration_of_cloud_liquid_water_in_air	Cloud liquid water content (height resolved)	y
mass_concentration_of_drizzle_in_air	Drizzle liquid water (height resolved)	y
mass_concentration_of_liquid_water_in_air	Liquid water content (height resolved; cloud + rain)	y
mass_concentration_of_water_vapor_in_air	Absolute humidity	y
mass_fraction_of_cloud_ice_in_air	Cloud ice content (height resolved)	y
mass_fraction_of_cloud_liquid_water_in_air	Cloud water content (height resolved)	y
mixing layer height from aerosol profile	Atmospheric boundary layer height derived from aerosol profile	n
mixing layer height from vertical wind profile	Atmospheric boundary layer height derived from wind profile	n
normalized range corrected signal OR range corrected backscatter signal	Normalized range corrected signal (lidar)	n
normalized_difference_vegetation_index	Normalized Difference Vegetation Index	y
number_concentration_of_ambient_aerosol_in_air	Aerosol number concentration	y
number_concentration_of_ice_crystals_in_air	Ice crystal number concentration	y
nyquist frequency	Nyquist frequency	n
nyquist velocity	Nyquist velocity	n
particle backscatter coefficient	Particle backscatter coefficient	n
particle depolarization ratio	Particle depolarization ratio	n
particle extinction coefficient	Particle extinction coefficient	n
particle lidar ratio	Particle lidar ratio	n
path integrated attenuation	Path integrated attenuation	n
precipitation_amount	Precipitation	y
radar co-polar correlation function	Radar co-polar correlation function	n

Continued on next page

CF-standard_name OR long name	Variable	In CF?
radar differential reflectivity	Radar differential reflectivity	n
radar integrated differential phase	Radar integrated differential phase	n
radar linear depolarization ratio	Radar linear depolarization ratio	n
radar specific differential phase	Radar specific differential phase	n
radar spectral skewness	Radar spectral skewness	n
radar spectral width	Radar spectral width	n
radial_velocity_of_scatterers_away_from_instrument	Radial velocity or fall velocity	y
radiation_wavelength	Wavelength of radiation	y
rainfall_rate	Rain rate	y
range corrected backscatter signal OR normalized range corrected signal	Range corrected backscatter signal	n
relative_humidity	Relative humidity	y
sea_surface_temperature	Sea surface temperature	y
sensor azimuth angle velocity	Sensor azimuth angle velocity	n
sensor elevation angle	Sensor elevation angle	n
sensor elevation angle velocity	Sensor elevation angle velocity	n
sensor_azimuth_angle	Sensor azimuth angle	y
sensor_band_central_radiation_frequency	Frequency band of sensor	y
specific_convective_available_potential_energy	CAPE	y
specific_humidity	Specific humidity	y
station number	Station number	n
surface_albedo	Surface albedo	y
surface_downwelling_longwave_flux_in_air	LW broadband downwelling radiation (surface)	y
surface_downwelling_shortwave_flux_in_air	SW broadband downwelling radiation (surface)	y
surface_longwave_emissivity	Surface longwave emissivity	y
surface_temperature	Surface temperature	y
surface_upward_latent_heat_flux	Latent heat flux	y
surface_upward_sensible_heat_flux	Sensible heat flux	y
surface_upwelling_longwave_flux_in_air	LW broadband upwelling radiation (surface)	y
surface_upwelling_shortwave_flux_in_air	SW broadband upwelling radiation (surface)	y
thickness_of_liquid_water_cloud	Cloud thickness (liquid clouds)	y
time	Time	y
total downwelling radiation	Total downwelling radiation	n
total upwelling radiation	Total upwelling radiation	n
upward_air_velocity	Vertical velocity	y
visibility_in_air	Visibility	y
volume depolarisation ratio	Volume depolarization ratio	n

Continued on next page

CF-standard_name OR long name	Variable	In CF?
volume extinction coefficient	Volume extinction coefficient	n
volume_attenuated_backwards_scattering_function_in_air	Attenuated backscatter coefficient	y
volume_extinction_coefficient_in_air_due_to_ambient_aerosol	Aerosol particle extinction coefficient	y
wind_from_direction	Wind direction	y
wind_speed	Wind speed	y
wind_speed_of_gust	Wind speed maximum (gust)	y
zenith_angle	Zenith angle of beam direction	y