Metadata standards for climate impact indicators for publication in the CLIPC portal

Version 1.4: 25th July 2016

Ruth Petrie[1], Milka Radojevic[4], Lars Bärring[2], Martin Juckes[1], Andrej Mihajlovski[3], Grigory Nikulin[2], Christian Page[4]

[1]STFC, Harwell Science Campus, UK
[2]SMHI, Rossby Centre, Sweden
[3]KNMI, De Bilt, Netherlands
[4]CERFACS, Toulouse, France

1. Introduction

This guideline document describes the metadata standards developed for the Climate Information Platform for Copernicus (CLIPC) climate impact indicators (CII). Within CLIPC there are three tiers of indicators: Tier 1 are climate statistics, Tier 2 are environmental impact indicators and Tier 3 are socioeconomic impact indicators. The metadata standards in this document represent the outcome of the CLIPC/IS-ENESES2 Joint Workshop on common controlled vocabularies for climate indices (February 2016).

The metadata standards presented here and referred to as CLIPC-CII-v1.4 are a first attempt to provide a metadata structure for multi-sectoral climate impact indicators. Specified in this document are the NetCDF global metadata attributes and a data reference syntax (DRS; which consists of a number of facets that describe specific attributes and provide a unique identifier for each dataset and file) for use by the CLIPC CII data producers and publishers. The correct specification of these metadata will enable data discovery, access and support data provenance (lineage) information which provides visibility to the data analytic pipeline and simplifies the tracing of errors.

All CII data must comply with the standards defined in this document to allow for facetted and free -text searching therefore ensuring data discovery via the portal.

Instructions for data producers:

1. **Provide indicators in a NetCDF file with one indicator per file** (auxiliary variables are permitted),
2. **Set the global attributes within the file** as specified in section 2 (using the appropriate template),
3. **Specify a filename according to the DRS** as detailed in section 3 (some examples are provided),
4. ***Check your metadata using the compliance checker*** *at https://climate4impact.eu/* **
5. **Send your data to KNMI** using ftp ftp opendap.knmi.nl (contact KNMI at if you do not have a username and password)

   KNMI will then use the metadata you supply within your file to publish your data to the CLIPC portal.

** This service is not yet available, KNMI are developing a file compliance checker that will be available via https://climate4impact.eu/ . You will be able to sign in with your google account, upload a file to your personal basket then at the processing section you can select the uploaded file and the checker will generate a metadata compliance report.

Section 2 of this document details the NetCDF global metadata attributes and these are fully compatible with the corresponding DRS facets. Section 3 details all the DRS facets that are required to create a unique filename and dataset identifier.

Important notes the CLIPC-CII-v1.4 metadata specified in this document

1. These metadata are intended as a guide for the CLIPC project and therefore are **only valid within CLIPC,** although it is hoped that future projects can make use of these metadata.
2. This document contains the full metadata guide that is required for the data producer and for the data publishers (KNMI and MARIS). If you are using this document only as a data producer then you need only refer to the sections that are relevant to you as detailed in the data producers guide list above.
3. The tables provided in this document can be accessed at https://docs.google.com/spreadsheets/d/1dRGC_s-rTyaKeoZ52bjy_AghOspeBKlWxcDNWcI2eZk/edit#gid=143572614

2. Global Attributes

The CLIPC-1.0 global metadata attributes follow the fixed structure and the pre-defined attribute names. The list of global attributes can be divided into three parts:

1. Knowledge discovery in a file,
2. Metadata of input datasets used to compute an indicator (attributes with the *invar_* prefix),
3. Simple description of temporal and spatial characteristics of a given indicator.

The given conventions do not represent an attempt to create entirely new standard but to adequately adapt existing standards for purposes of the CLIPC CII. The attributes represent indeed a combination of the CF-1.6 and ACDD-1.3 conventions with auxiliary attributes often inspired by the WRCP's CORDEX and CMIP5 conventions in order to cover for provenance of input data and ensure direct mapping with DRS facets.

Instructions to set the global attributes:

- Insert the whole list of Attribute Names from Table 1 (**mandatory**) to your NetCDF file.
- Associate value to the attribute names. Leave empty field "" if not applicable for your data.
- Attribute values should be clear and concise.

The following table details all the NetCDF global attributes:

Attribute Name	Description	Comment	Analogous DRS facet
Conventions	Convention name and version, as for example CF-1.6	The Unidata's ACDD-1.3 also follows CF-1.6 convention to enable data access and discovery.	
Activity	Name of project or activity within the larger scope.	All CLIPC indicators will have "clipc" as the activity.	activity
title	Dataset title - Human readable	Human readable concise description of the file content, and will be displayed as the data title on the CLIPC portal	
summary	Abstract related to the title attribute. Human readable.	A brief summary of the data that will be displayed on the CLIPC portal describing your data. Please include all essential information.	
variable_name	Short variable name as included in the file and filename.	Due to no CF-1.6-standard names being available for the climate indicators. Tier 1 indices should use accepted acronym for their index using lowercase conventions, Tier 2 and 3 should use sensible short names. The long_name variable attribute can be used to give full details, if you are unfamiliar with variable attributes see http://CF-1.6conventions.org/CF-1.6-conventions/v1.6.0/CF-1.6-conventions.html#table-attributes .	VariableName
product	Specification for the indicator(s) whether derived from observations, reanalysis, climate model output or other.	Examples as given in [2]: gcm-derived, rcm-derived, reanal-derived, obs-derived, multi-derived, nonClimatic-derived, ens-derived. Multi means at least two sources of input dataset were used to derive an indicator and so on. *Introduced by the CORDEX WRCP project.	product
comment	Miscellaneous information about the data/methods.	Optional, to provide additional information not provided in any other attributes.	
references	Reference relating to indicator.	Please provide citation to documentation (published or technical note) on the indicator.	
package_name	Main software package/model/library used to produce the indicator: name and version number appended by a minus sign.	Examples: icclim-3-1-0; nco-4-4-8; cdo-1-6-0; arcgis-10-3-1; etc.	package
package_references	Reference and/or source for package_name	Reference/link to published model or software. Otherwise, if a set of programs, provide library name and version.	
institution_id	Organization at which data is created. If multiple institutions use lead institution.	Analogous found in the WRCPs CMIP5 and CORDEX, the institute_id reserved for scientific or educational organization.	institution
institution_url	URL of the institution_id's home page.	Flag associated to the institution_id	
contact	Contact e-mail address at the institution_id.	E-mail contact(s) (personal/generic) for project and/or institution. If multiple contacts, include comma separated list of emails.	
contributor_name	Name of any individuals, projects, or institutions that contributed to the creation of this data.	Mandatory in case when the created data are result of a collaboration across the organizations. Include comma separated list of all contributors.	

contributor_role	Role of any individuals, projects or institutions contributed to the creation of this data.	Multiple roles should be presented in the same order and number as given in contributor_name.	
date_created	Reference date of dataset creation.	It must be in form YYYYMMDD	
date_issued	Date when data become available for public use.		version
date_modified	Date when data was last verified/revised/corrected.		
realisation_id	Additional facet for flexibility to allow for potential future variations. Potential usage could be e.g. v1 or r1	OPTIONAL, leave empty if not required	IndicatorRealisation
source_data_id	Single source data: supply the name of dataset Multiple source data: supply multi-type[-realization][-id] - type could be: model, obs, platform, sensors, package, mixed - realization is optional for use as required to ensure uniqueness, suggest to use form r1, v1 (no delimiting characters allowed) - id, is an optional additional component facet which could be used for example to distinguish between experiments (no delimiting characters allowed) Ensemble data: supply in the form: ens-type-realization[-id] - type could be: a single model (specify name), multiModel, multiExpts, multiObs, multiMixed - realization: ensemble member (form e.g. r11), mean, median, mode, percentile (form e.g. p10) - id is optional to ensure uniqueness, it could be a version no. or other identifier e.g. cmip5v3, institutions can use this identifier in any way as it free text to ensure uniqueness, an unlimited number of [-id] facets components can be used with **hyphen** separation only within the sourceDataID facet.	Purpose is to ensure the uniqueness of dataset DRS. Vocabulary is inspired by DRS facet from the Climate Change Initiative (CCI) portal.	sourceDataId
source_data_id_comment	If using multi, please provide comment with additional explanatory details. If using ensembles, please provide additional details and comma separated list of ensemble sources	OPTIONAL, leave empty if not required	
invar_platform	Type of platform/resource from which the input data observations-based were issued except the reanalysis.	OPTIONAL, leave empty if not required Example: station/ground, satellite, remote sensing, GIS, etc. If multiple sources, insert a comma separated list.	
invar_platform_id	Name and version of the platform. If multiple, insert a comma separated list.	OPTIONAL, leave empty if not required, platform_id is equal to sourceDataId facet if one platform used. Example: ERAI, EURO4M-MESAN-v11, EOBS, EUM-SAF, etc.	

invar_satellite_algorithm	Name and version of the algorithm for satellite data retrieval.	OPTIONAL, leave empty if not required.	
invar_satellite_sensor	Name of satellite sensor(s) used to provide satellite images.		
invar_rcm_model_id	Name and version of RCM if model output is a: a) source of input data, or b) part of a multi-source indicator.	OPTIONAL, leave empty if not required. If multiple provide comma separated list.	RCMName
invar_rcm_model_realization_id	Realization id for different scenarios as given in the source dataset issued from the CORDEX project.	OPTIONAL, leave empty if not required.	RCMRealisation
invar_rcm_model_driver	Name and version of driving dataset such as GCM or Reanalysis.	OPTIONAL, leave empty if not required.	GCMName
invar_reanalysis_id	Name and version of a Reanalysis project if product is a: a) source of input data, or b) part of a multi-source indicator.	OPTIONAL, leave empty if not required.	GCMName
invar_gcm_model_id	Name and version of GCM if model output is a: a) source of input data, or b) part of a multi-source indicator.	OPTIONAL, leave empty if not required.	GCMName
invar_experiment_name	If CMIP use CMIP standards else use sensible name	Examples: historical, piControl, rcp26, rcp45, sresa2, sresa1b	ExperimentName
invar_ensemble_member	If CMIP use CMIP standards else use "v1" etc.	Examples: r1i1p1, r2i1p1, etc.	EnsembleMember
invar_bc_method_id	Name of the Bias Correction (BC) technique if using bias-corrected output from RCM or GCM.	OPTIONAL, leave empty if not required.	BcName
invar_bc_observation_id	Name of observations-based dataset used for the BC.		BcObsName
invar_bc_period	Reference period of bc_observation_id used for the BC in form start-end format YYYY-YYYY		BcRefPeriod
invar_variable_name	Name of scalar or vector short variable name as used to compute the indicator. If multiple, insert a list comma separated names.	As given in the corresponding resources.	
invar_tracking_id	A UUID (Universally Unique Identifier) can be used for reliably identifying very persistent objects across a network e.g. ESGF. UUIDs are an octet string of 128 bits.	Insert tracking_id from the first input data file. Available for the files as issued by the WRCPs CMIP5 and CORDEX.	
reference_period	Baseline reference period over which the indicator is calculated in reference to. It should be supplied in the format YYYY-YYYY.	OPTIONAL, leave empty if not required.	Reference_period
output_frequency	Time distribution at which the indicator is computed, and is represented by the time variable. It can be mon (monthly), sea (seasonal) or yr (yearly) basis as explained in [1]	In the case of the climate indices, frequency only can be a semantically ambiguous attribute. Defined as the number of times that an event occurs within a given period, frequency is also a generic definition for the indices based on day count when a specific condition is fulfilled.	frequency

time_coverage_start	Time range field: Start time (the first time value for a variable with a time series **or** the start time of a period the indicator represents). Must be in the format YYYYMMDD.	Leave empty if not applicable, i.e. if you have time invariant data that is only valid at one date.	StartTime
time_coverage_end	Time range field: End time (the last time value for a variable with a time series or the end time of a period the indicator represents). Must be in the format YYYYMMDD.		EndTime
cdm_datatype	Unidata's Common Model Scientific Data Types understood by THREDDS: Grid, Image, Point, Radial, Station, Swath & Trajectory. Choice depends on spatial structure of the data.	Selected attribute value is then introduced as a NetCDF variable except Grid: * Instructions are in the CF-1.6 1.6 Convention CF-1.6conventions.org, section: 9.1 Features and feature types. * Examples are at www.nodc.noaa.gov/data/formats/netcdf/v2.0, go to the links in the table Feature Type Templates and Examples.	
domain	Name or short description of horizontal spatial data extent. Or lat-lon bounding box	Example from the CORDEX: EUR-11, EUR-44, etc.	domain
geospatial_bounds	Description of 2D spatial data extent for limited domains.	Example: 'POLYGON (40.26 -111.29, 41.26 -111.29, 41.26 -110.29, 40.26 -110.29)'	
geospatial_lat_min	Northward bounding latitude of the domain	A value followed by the unit: degrees_east is recommended by CF-1.6 convention.	
geospatial_lat_max	Southward bounding latitude of the domain		
geospatial_lat_resolution	A number value followed by the unit.	Examples: '100 meters', '0.1 degree'.	
geospatial_lon_min	Eastward bounding longitude of the domain.	A value followed by the unit: degrees_west is recommended by CF-1.6 convention.	
geospatial_lon_max	Westward bounding longitude of the domain.		
geospatial_lon_resolution	A number value followed by the unit.	Examples: '100 meters', '0.1 degree' Compliant with the INSPIRE standard.	
tile	Geographic tile location for the indicator.	OPTIONAL, leave empty if not required	tile-nnnnn
tracking_id	A UUID (Universally Unique Identifier) can be used for reliably identifying very persistent objects across a network e.g. ESGF. UUIDs are an octet string of 128 bits.	Easy to create on both Linux and Window platforms. On Linux/Unix, use the command: uuid -v 4	
keywords	A comma separated list of keywords based on predefined (Controlled Vocabularies) or commonly used words to describe the subject.	At least two key words to be specified. They must be found as values of other global attributes.	
history	Provides an audit trail for modifications to the created data.	Leave empty string if no modifications have been introduced after date of creation.	

Table 1. The structure of the CLIPC Global Metadata Attributes for the climate change impact indicators (CII) as given in the NetCDF or ASCII formatted files. The table is also available in a spreadsheet format available at: https://docs.google.com/spreadsheets/d/1dRGC_s-rTyaKeoZ52bjy_AghOspeBKlWxcDNWcI2eZk/edit#gid=143572614

2.1 CLIPC global attributes: examples

Two examples (Tier 2 and Tier 1) of the NetCDF global attributes are given. Hence, the corresponding filenames were created according to the DRS facets mapped to the corresponding values of the global attributes.

2.1.1 Global attributes for a tier 2 CCI: Growing Season Index for olives (GSI)

ncdump -h **gsi_nco-4-4-8_CERFACS_multi-platform-tier2v1_day_19890101-20101231.nc**

:Conventions = "CLIPC-1.0" ;
:activity = "clipc" ;
:title = "Growing Season Index for olives in Mediterranean basin" ;
:summary = "The GSI after Orlandi et al. (2013) is an indicator describing the combined effect of key climatic factors on the olive phenological development in the Mediterranean basin. The annual cycle of a 21-day moving average of the single daily GSI represents the olive growing season curve and follows a bimodal distribution." ;
:variable_name = "gsi" ;
:product = "obs_derived" ;
:comment = "Original approach has been adapted to fit a structured fine-resolution grid of input data. At altitudes above 800 m and over water surfaces, the indicator is set to missing value." ;
:references = "Orlandi et al., 2013: Climatic indices in the interpretation of the phenological phases of the olive in mediterranean areas during its biological cycle. Climatic Change, 116:262-284. DOI:10.1007/s10584-012-0474-9" ;
:package_name = "nco-4-4-8" ;
:package_references = "nco.sourceforge.net" ;
:institution_id = "CERFACS" ;
:institution_url = "cerfacs.fr" ;
:contact = "christian.page@cerfacs.fr" ;
:contributor_name = " " ;
:contributor_role = " " ;
:date_created = "20160217" ;
:date_issued = "20160613" ;
:date_modified = " " ;
:realisation_id = "v1" ;
:source_data_id = "multi-platform-tier2v1" ;
:source_data_id_comment = "Input datasets are issued from two sources. Air temperature input datasets are the regional re-analysis EUR4M-MESAN product while surface downwelling shortwave radiation is published by EUMETSAT CM-SAF" ;
:invar_platform = "satellite" ;
:invar_platform_id = "EUMETSAT CM-SAF" ;
:invar_satellite_algorithm = " " ;
:invar_satellite_sensor = " " ;
:invar_rcm_model_id = " " ;
:invar_rcm_model_realization_id = " " ;
:invar_rcm_model_driver = " " ;
:invar_reanalysis_id = "EURO4M-MESAN_v1" ;
:invar_gcm_model_id = " " ;
:invar_experiment_name = " " ;
:invar_ensemble_member = " " ;
:invar_bc_method_id = " " ;
:invar_bc_observation_id = " " ;
:invar_bc_period = " " ;
:invar_variable_name = "tas,tasmin,SIS" ;
:invar_tracking_id = " " ;
:reference_period = "1989-2010" ;
:output_frequency = "day" ;
:time_coverage_start = " " ;
:time_coverage_end = " " ;
:cdm_datatype = "Grid" ;
:domain = "EUR-05" ;
:geospatial_bounds = "Mediterranean basin bounded on the south" ;
:geospatial_lat_min = " " ;
:geospatial_lat_max = "47.5 degrees_north" ;
:geospatial_lat_resolution = "0.05 degrees" ;
:geospatial_lon_min = "-12.5022 degrees_east" ;

:geospatial_lon_max = "42.7446 degrees_east" ;
:geospatial_lon_resolution = "0.05 degrees" ;
:tile = " " ;
:tracking_id = "e38c163b-ab47-4cf8-b4d0-9fd8428d90ef" ;
:keywords = "climate, mediterranean, olive, growing, season, index, Orlandi, observations, reanalysis, satellite, 0.05 degree" ;
:history = " " ;
}

2.1.2 Global attributes for a tier 1 CCI: TN10p (cold nights)

ncdump -h **tn10p_icclim-4-1-0_CERFACS_SMHI-MESAN-v1_yr_19890101-20101231_19890101-20101231.nc**

:Conventions = "CLIPC-1.0" ;
:activity = "clipc" ;
:title = "TN10p: Count of days when daily minimum of 2 m air temperature is below the 10th percentile" ;
:summary = "TN10p is a climate change index defined by the ETCCDI using the percentile-based threshold. The indicator measures the maximum number of cold nights during a year for a given location. To calculate the corresponding percentile with respect to a given reference period, the bootstrap method has been applied (Zhang et al., 2005)." ;
:variable_name = "tn10p" ;
:product = "reanal_derived" ;
:comment = "ETCCDI stands for the joint CCl/CLIVAR/JCOMM Expert Team (ET) on Climate Change Detection and Indices." ;
:references = "Karl, T.R., N. Nicholls, and A. Ghazi, 1999: CLIVAR/GCOS/WMO workshop on indices and indicators for climate extremes: Workshop summary. Climatic Change, 42, 3-7 \n",
"Peterson, T.C., and Co-authors: Report on the Activities of the Working Group on Climate Change Detection and Related Rapporteurs 1998-2001. WMO, Rep. WCDMP-47, WMO-TD 1071, Geneve, Switzerland, 143pp. \n",
"Zhang, X., G. Hegerl, F.W. Zwiers and J. Kenyon, 2005: Avoiding Inhomogeneity in Percentile-Based Indices of Temperature Extremes, J. of Climate, Volume 18, 1641-1651" ;
:package_name = "icclim-4-1-0" ;
:package_references = "https://github.com/cerfacs-globc/icclim" ;
:institution_id = "CERFACS" ;
:institution_url = "cerfacs.fr" ;
:contact = "christian.page@cerfacs.fr" ;
:contributor_name = " " ;
:contributor_role = " " ;
:date_created = "20160417" ;
:date_issued = "20160614" ;
:date_modified = " " ;
:realisation_id = "v1" ;
:source_data_id = "SMHI-MESAN-v1" ;
:source_data_id_comment = "SMHI-HIRLAM European high-resolution surface reanalysis published as a CLIPC activity." ;
:invar_platform = " " ;
:invar_platform_id = " " ;
:invar_satellite_algorithm = " " ;
:invar_satellite_sensor = " " ;
:invar_rcm_model_id = " " ;
:invar_rcm_model_realization_id = " " ;
:invar_rcm_model_driver = " " ;
:invar_reanalysis_id = "SMHI-MESAN" ;
:invar_gcm_model_id = " " ;
:invar_experiment_name = " " ;
:invar_ensemble_member = " " ;
:invar_bc_method_id = " " ;
:invar_bc_observation_id = " " ;
:invar_bc_period = " " ;
:invar_variable_name = "tasmin" ;
:invar_tracking_id = "7be0ca35-16b3-42b3-b1fe-fe53e84c2c48" ;
:reference_period = "1989-2010" ;
:output_frequency = "yr" ;
:time_coverage_start = "19890101" ;
:time_coverage_end = "20101231" ;
:cdm_datatype = "Grid" ;
:domain = "EUR-11" ;
:geospatial_bounds = " " ;

:geospatial_lat_min = "21.98783f" ;
:geospatial_lat_max = "72.585f" ;
:geospatial_lat_resolution = "0.11 degrees" ;
:geospatial_lon_min = "-44.59386f" ;
:geospatial_lon_max = "64.96438f" ;
:geospatial_lon_resolution = "0.11 degrees" ;
:tile = " " ;
:tracking_id = "23ac1fd1-2370-4295-81e8-bab1285c4b5c" ;
:keywords = "ETCCDI, climate, index, TN10P, year, cold, temperature, below, threshold, percentile, reference, bootstrap, reanalysis, EUR-11" ;
:history = " " ;
}

3. Data Reference Syntax (DRS)

In this section the data reference syntax (DRS) for the climate impact indicators for CLIPC are described in detail. A DRS is constructed through a list of facets that describe dataset and file attributes. The facets have a corresponding global attribute, however the usage is different, therefore there is some repetition between this section and the global attributes section. This is intentional so that each section can be understood if read independently. The aim of the DRS is to provide a unique identifier for each dataset and file.

Since climate impact indicators may be calculated from a variety of sources, either climate model data, reanalysis data, observational data, non-climatic data or from multiple sources the DRS for climate impact indicators are non-trivial to construct. Here two DRS are provided: one for indicators/indices that are solely derived from model (or reanalysis) output and one general DRS for indicators derived from observations, non-climatic information it allows for more complex indicators derived from multiple sources (models, observations or combinations) and ensembles.

For the purposes of the CLIPC project the following DRS has been agreed. All those producing, Tier 1, 2, or 3 climate impact indicators should follow the DRS conventions detailed below. **If you are using this guide as a data producer it is essential that you provide a file with one indicator that has a filename in the format as detailed below.** KNMI will use the global attributes specified in your file to generate the dataset DRS and therefore the information and examples relating to dataset DRS are for the data publishers use and for information only for the data producer.

Section 3.1 lists all the facets used to construct the filename and dataset DRS, section 3.2 details the construction of the filename and dataset DRS and section 3.3 provides some example filenames and dataset identifiers.

3.1 DRS Facets for CLIPC climate impact indicators

In Table 1 all the facets that are required to construct both the model-derived and general climate impact indicator filename and dataset identifiers are listed. The table indicates where facets are required or optional, either in the filename or dataset and whether the facet value is fixed or flexible. A fixed facet is one that has a fixed format, e.g. a date that must be supplied in a specific format or to be selected from the relevant controlled vocabulary, for example in this case the *product* facet may only have certain values that are specified in the table. A flexible facet can be constructed freely, though please use commonly accepted community acronyms where possible.

The facets used for model derived indicators loosely follows the DRS used in CORDEX, however the CORDEX standard is not flexible enough to cover the needs of the climate impact indicator communities, therefore some facets have been made optional or are not constrained in the same way as CORDEX (see table for full details).

Climate impact indicators that are not derived solely from model data, e.g. are derived from observations, non-climatic information, multiple sources (models, observations or both) or ensembles should be described by the General DRS. This DRS has been adapted from the ESA Climate Change Initiative (CCI) DRS that has been developed as part of CLIPC see D5.2[1].

[1] http://www.clipc.eu/content/content.php?htm=102

IMPORTANT NOTES FOR DATA PRODUCERS:

- If a filename facet is optional then simply omit it from your specified filename.

- Note that filenames are underscore ("_") delimited and dataset identifiers are dot-delimited (".") therefore these are prohibited for use within a facet name. If for example if a software package has a dot delimited version (or any other character than letters, digits or hyphens) this **must** be altered, e.g. cdo v.1.6.2 becomes cdo-1-6-2.

- In the <*VariableName*> facet do not use a theme name, use the indicator acronym (Tier1) or short name if there is no acronym (Tiers 2 & 3). Also note that **only one indicator per file is permitted**, although auxiliary variables can be included in the file.

- The <*frequency*> facet is the temporal frequency of the data in the file. However, many indicators are time invariant. For a time invariant temporal resolution of output follow the CMIP5 standard (http://cmip-pcmdi.llnl.gov/cmip5/output_req.html p7).

  If your indicator is time invariant use e.g. "monClim" or "dayAvg". If your data does still does not fit within these parameters, simply use "clim" or "avg" for time invariant output and specify details of the calculation within the comment field of the global attributes.

  If you have time invariant data then within the facet <StartTime-*EndTime*>:
  - You can provide the period over which your indicator is valid i.e. the averaging period in the format YYYYMMDD-YYYYMMDD or
  - if your data is only valid at one specific date then only provide an end time at which your data is valid in the format YYYYMMDD
  - It maybe that you include the facet <*Reference_period*> and in some cases this may be equal to the <*StartTime-EndTime*> this is unfortunate, but necessary repetition of information.

  Some variables such as growing season length have multiple and competing definitions. The definition should be given in the variable "long_name" of the file, the global attribute "title". Each indicator should have individual records in the climate impact indicators database there.

- **The facet *<reference_period>* is optional.** Most indicators will have a reference period from which they were calculated, however there are exceptions, if a reference period is not applicable to the indicator then leave this facet empty, see examples in 3.3.1.1 and 3.3.1.2.

- The <*version*> facet is used in the dataset identifier it is the ESGF version **not the dataset version**. The version can be specified by the person publishing the data OR picked up as the current date by the ESGF-publisher tool. Either way, it must always have the precise form "**v<YYYYMMDD>**". If you need a dataset version use the <*IndicatorRealisation*> facet.

- **Ensembles** if you have ensemble data, you need to specify this in the <*product*> facet using the guide notes there and then in the <*sourceDataID*> you need to follow the facet name construction for ensembles.

Facet Name	Fixed or flexible	Usage	Required (Req) / Optional (Opt)			
			Dataset		Filename	
			Model	General	Model	General
activity	Fixed	Within the CLIPC project the fixed string "clipc" will be used.	Req	Req	---	---
product	Fixed	One of "gcm", "rcm", "obs", "reanal", "multi", "nonClimatic", immediately followed by "-derived" (without any space). In the case of ensembles use either "ensemble" or "ens-stat"	Req	Req	---	---
package	Fixed	Software package used to generate the climate indicator. Include version by appending "-" directly after package name, e.g. icclim, R-package, cdo-1-6-2 (note some versions may be dot delimited, these or any other character should be translated to "-"	Req	Req	Req	Req
domain	Flexible	**Preference is to use CORDEX domain names** as specified here: http://www.cordex.org/index.php?option=com_content&view=article&id=241&Itemid=488 Alternatively, a lat/lon bounding box may also be used. For example for a northern Pacific domain could be 170E-170W-40N-70N or 40N-70N-170E-170W the N,S,E,W components can be given in any order.	Opt	Req	Opt	Req
institution	Flexible	Lead Institution or organisation that generated the data. Where data has been generated through a multi-institutional collaboration, list others as contributors in the global attributes).	Req	Req	Req	Req
GCMName	Fixed	Name of the model used to derive the index could be global model, a reanalysis product or the driving model in the case of indicators calculated from regional models. Conventions for this facet should follow that of the program (i.e. CMIP5 or CORDEX) used to derive the indicator.	Req	---	Req	---
ExperimentName	Fixed	Project experiment name, where project is e.g. CMIP5 it would be of the form rcp{*xx*}, historical, etc.	Req	---	Req	---
EnsembleMember	Fixed	Project ensemble reference number, where project is e.g. CMIP5 then this is in the form r12i1p1	Req	---	Req	---
IndicatorRealisation	Flexible	Additional facet for flexibility to allow for potential future variations. Potential usage could be e.g. v1 or r1	Opt	---	Opt	---
RCMName	Fixed	Regional climate model name and version id (hyphen delimited)	Opt	---	Opt	
RCMRealisation	Fixed	Regional climate model downscaling realisation id as in CORDEX	Opt	---	Opt	
BcObsName	Flexible	An acronym for the observation/reanalysis datasets used as a reference for bias adjustment	Opt	---	Opt	
BcRefPeriod	Fixed	Reference or calibration period given in the form "bcref-YYYY-YYYY" format where "YYYY-YYYY" specifies the start and end years.	Opt	---	Opt	
sourceDataID	Flexible	Single source data: supply the name of dataset Multiple source data: supply multi-type[-realization][-id] - type could be: model, rcm, obs, platform, sensors, package, mixed - realization is optional for use as required to ensure uniqueness, suggest to use form r1, v1 (no delimiting characters allowed) - id, is an optional additional component facet which could be used for example to distinguish between experiments (no delimiting characters allowed)	---	Req	---	Req

		Ensemble data: supply in the form: ens-type-realization[-id] - type could be: a single model (specify name), multiModel, multiExpts, multiObs, multiMixed - realization: ensemble member (form e.g. r11), mean, median, mode, percentile (form eg. p10) - [-id] is optional to ensure uniqueness, it could be a version no. or other identifier e.g. cmip5v3, institutions can use this identifier in any way as it free text to ensure uniqueness, an unlimited number of [-id] facets components can be used with hyphen separation only within the sourceDataID facet. Use global attribute source data id comment to specify full details.				
frequency	Fixed	Temporal resolution at which output is given following CMIP conventions e.g yr, mon, day, 6hr, 3hr, subhr. Time invariant data can have a frequency of e.g. monClim (climatological monthly mean), yrClim (climatological annual mean). If this is still not applicable to your data then "clim" and "avg" are also accepted.	Req	Req	Req	Req
StartTime-EndTime	Flexible	For time series data this is the temporal range of output data in the format: YYYYMMDD-YYYYMMDD. For time-invariant data this is the temporal range of when the data is valid over in the form YYYYMMDD-YYYYMMDD (or just YYYYMMDD if data is valid at only one time).	Req	---	Req	---
Reference_period	Fixed	Baseline reference period over which the indicator is calculated. It should be supplied in the format YYYY-YYYY; leave blank if not applicable to your data. For example, for climate anomalies relative to the 1986 to 2015 climate, use 'reference_period=1986-2015'.	Opt	Opt	Opt	Opt
VariableName	Flexible	Climate indicator acronym (Tier 1) or a short name (Tier 2 & 3) e.g. for stormSurgeFlood use ssfld not the impact theme flooding, a "long_name" can be specified in the **variable** attributes).	Req	Req	Req	Req
version	Fixed	**ESGF version: vYYYYMMDD** for the dataset. It is either given by provider or auto generated using the date when the dataset is published in ESGF.	Req	Req	---	---
tile-nnnnn	Fixed	Geographic location tile for indicator. Coordinates of the tile should be specified within the attributes of the file and it should be uniquely numbered. (Required due to high resolution input data.)	---	---	Opt	Opt

Table 2 Climate impact indicators DRS facets

3.2. Constructing the DRS

Note that the facet names are given between angle brackets "<>", where there are square brackets "[ ]" surrounding a facet or group of facets it denotes that these facets are optional. Optional facets should be just be ignored and left blank required facets must not be left blank (see examples).

If you are a data producer you only need to provide the filename using one of the formats below.

3.2.2 Model-derived impact indicators DRS

Any climate impact indicator of any tier that is SOLEY & SINGLY derived from model data should use these DRS.

Filename:

```
<VariableName>_<package>_<institution>_<GCMName>_<ExperimentName>_
<EnsembleMember>_[IndicatorRealisation_][<RCMName>_<RCMRealisation>_<domain>_]
[<BcName>_<BcObsName>_<BcRefPeriod>_] <frequency>_<StartTime-EndTime>_
[<Reference_period>_][tile-nnnnn].nc
```

Dataset:

```
<activity>.<product>.<package>.<institution>.<GCMName>.<ExperimentName>.
<EnsembleMember>.[<RCMName>.<RCMRealisation>.<domain>.][<BcName>.<BcObsName>.
<BcRefPeriod>.]<frequency>.[<Reference_period>.]<VariableName>.<version>
```

3.2.3 General DRS

Any climate impact indicator of any tier that is not derived solely and singly from model data should use this DRS as it covers observationally-derived, non-climatic, multi-source (models or observations) and ensembles.

Filename:

```
<VariableName>_<package>_<institution>_<sourceDataID >_<frequency>_
<StartTime-EndTime>_[<Reference_period>_][tile-nnnnn].nc
```

Dataset:

```
<activity>.<product>.<package>.<domain>.<institution>.<sourceDataID>.<frequency>.[<
Reference_period>].<VariableName>.<version>
```

3.3 Example DRS usage

Recall that for CLIPC you only need to provide the correct filename – KNMI will create the dataset id from information supplied in the global attributes.

3.3.1 Model-derived indicators(indices) examples

3.3.1.1 Tier-1 index from a global climate model that does not require a reference period

Filename: `cdd_icclim-4-1-2_SMHI_ICHEC-EC-EARTH_historical_r12i1p1_yr_19500101-20493112.nc`
Dataset: `clipc.gcm-derived.icclim-4-1-2.SMHI.ICHEC-EC-EARTH.historical.r12i1p1.yr.cdd.v20141010`

3.3.1.2 Tier-1 index from a global climate model that does require a reference period

Filename: `cdd_icclim-4-1-2_SMHI_ICHEC-EC-EARTH_historical_r1i1p1_yr_19500101-20493112.nc`
Dataset: `clipc.gcm-derived.icclim-4-1-2.SMHI.ICHEC-EC-EARTH.historical.r1i1p1.yr.1950-1980.cdd.v20141010`

3.3.1.3 Tier-1 index from a regional climate model, without bias correction or reference period

Filename: `cdd_icclim-4-1-2_SMHI_ICHEC-EC-EARTH_historical_r1i1p1_SMHI-RCA4_v1_EUR-11_yr_19500101-20493112.nc`
Dataset: `clipc.rcm-derived.icclim-4-1-2.SMHI.ICHEC-EC-EARTH.historical.r1i1p1.SMHI-RCA4_v1.EUR-11.yr.cdd.v20141010`

3.3.1.4 Tier-1 index from a regional climate model, with bias correction specifying bias correction reference period and the dataset reference period

Filename: `cdd_icclim-4-1-2_SMHI_ICHEC-EC-EARTH_historical_r1i1p1_SMHI-RCA4_v4_EUR-11_DBS42_EURO4M-Mesan_bcref-1989-2010_yr_19800101-20493112_1981-2010_.nc`
Dataset: `clipc.rcm-derived.icclim-4-1-2.SMHI.ICHEC-EC-EARTH.historical.r1i1p1.SMHI-RCA4.v4.EUR-11.DBS42.EURO4M-Mesan.bcref-1989-2010.yr.1981-2010.cdd.v20141010`

3.3.2 General indictors examples

3.3.2.1 Tier 1 indicator of Standardized Snow Pack Index with no reference period

Filename: SSPI-10_fmi-assimilation-algorithm-1-3_FMI_multi-obs-01_day_19790101-20160501.nc
Dataset: clipc.multi-derived.fmi-assimilation-algorithm-1-3.11W-35E-35N-72N.FMI.multi-obs-01.day.SSPI-10.v20160521

3.3.2.2 Tier 1 indicator of cdd (consecutive dry days) but derived from observations with no reference period

Filename: cdd_icclim-4-1-2_SMHI_EOBS10_yr_19790101-20160501.nc
Dataset: clipc.obs-derived.icclim-4-1-2.EUR.SMHI.EOBS10.yr.cdd.v20150101

3.3.2.3 Tier 2 observationally derived olives indicator

Filename: gsi_nco-4-4-8_CERFACS_multi-mixed_dayClim_19810101-20101231_1981-2010.nc
Dataset: clipc.obs-derived.nco-4-4-8.EUR-05-Med.CERFACS.multi-mixed.dayClim.1989-2010.gsi.v20160217

3.3.2.4 Tier 2 multiple-source derived storm surge flood indicator, using a regional identifier

Filename: ssfld_R-3-1_PIK_multi-mixed-02_clim_20000101-20500101_1980-2000_tile-05536.nc
Dataset: clipc.multi-derived.R-3-1.EUR.PIK.multi-mixed-02.clim.1980-2000.ssfld.v20140404

3.3.2.5 Tier 2 multiple-source derived Moth phenology indicator, using a bounding box for region, no reference period

Filename: MothPhenology_R-nlme-3-1-1-2-8_SYKE_multi-model_clim_20010101-20151231.nc
Dataset: clipc.multi-derived.R-nlme-3-1-1-2-8.19E-32W-58N-71N.SYKE.multi-model.clim.2001-2015.MothPhenology.v20160512

3.3.2.6 Tier 3 multiple-source derived potential economic damages from storm surges and sea-level rise, using a bounding box for region, valid at one time only

Filename: econdamsurge_R-raster-2-5-6_PIK_multi-model-03_clim_20500101_2005-2050.nc
Dataset: clipc.multi-derived.R-raster-2-5-2.24E-41W-34N-71N.SYKE.multi-model-03.clim.2005-2050.econdamsurge.v20140404

**** Please be aware that every effort has been made to ensure that the examples given here are correct, however there may be small errors therefore please take care when constructing your filename**

References

[1] CLIPC MS 19: Extended controlled vocabularies (http://www.clipc.eu/media/clipc/org/documents/milestones/ms19_drsvocabs_april2015_final.pdf)

[2] CLIPC D5.2: Metadata and controlled vocabularies specification for data, quality control and uncertainties (http://www.clipc.eu/content/content.php?htm=102)

[3] Guidelines for Construction of CF Standard (http://cfconventions.org/Data/cf-standard-names/docs/guidelines.html)

[4] NODC NetCDF Templates 1.1 (http://www.nodc.noaa.gov/data/formats/netcdf/v1.1/)

[5] Unidata Dataset Discovery (http://wiki.esipfed.org/index.php/Attribute_Convention_for_Data_Discovery).