CLIP-C NetCDF template for the global attributes v2.2 CERFACS

Good metadata is key for research data access and re-use (discovery and identification).
It provides visibility to the analytics pipeline (evaluation) and simplifies tracing errors (communication).
The global attributes support data lineage (provenance of data and where it flows over time) that is important for data mining and knowledge discovery in a file.

- This template does not represent an attempt to create a new standard but already to adapt existing to purposes of the climate change impact indicators within the CLIP-C.
- The template represents indeed a combination of the CF and ACDD conventions with auxiliary attributes related to the specific needs of the project.
- Description, link to the standard conventions and comments are provided. Comments and suggestions are welcome.

Instructions:
1. All global attributes are mandatory across the Tiers except platform and satellite.
2. If no attribute value to specify, leave empty string.
3. Be explicit as much as you can but be explicit and brief.

Table 1. Global, mandatory and recommended, attributes as Data lineage intended to describe the climate change and impact indicators within the CLIP-C project (www.clipc.eu).

Global Attribute	Convention	Description	Comment
Part 1. Knowledge discovery in a file			
Conventions	CF & ACDD	Convention name and version: CF-1.6	Unidata ACDD also follows CF convention to enable dataset access & discovery.
title		A short description of the file content.	It indicates (to humans and search engines) what your data is about, and help decide whether to select it.
activity	new	Name of the project or activity within the larger scope	
product	Not in CF and ACDD conventions, but used in the WCRP programme CORDEX	Type of the created data classified according to their characteristics with specific sector or application.	It is meant to describe kind/type of created data (Example: CLIPC Tier 1 climate change index)
package_name	new	Name and version of the software package/model used to produce the data.	If a set of local programs has been used, leave empty string.
package_references		Reference/source/library for framework_name	Reference/link to published model or software. Otherwise, if a set of local programs, provide library name and version.
references		Published or web-based references to describe data/methods used to create data.	If available.
comment	CF	Miscellaneous information about the data or methods used.	It could be longer than long_name and is intended to provide in greater detail important to interpret data.
date_created	ACDD	Date of completition of the data.	
date_published	new	Date when the data become available for use.	ISO 8601:2004 extended date format (Example: 2010-04-16T00:00:00) CORDEX project uses ISO standard but creation_date naming.]
date_revised		Date when the data was last verified/revised/corrected.	
summary	ACDD	A longer description of dataset.	Insert short versions of other relevant global attributes (creator, geospatial and temporal).
keywords		A comma separated list of keywords.	Based on other relevant global attributes.
institution_id	Not in CF and ACDD conventions.	Institution's acronym/initialism	In CMIP and CORDEX (projects for modeling climate data), institute_id indicates an scientific or educational organization. But, to enable various options and consistency with institution, this might be required.
institution_url		Institution's url	Associated to institution_id
contact_email	contact is from CF	E-mail address of contact	Suggested form for CF-1.6 attribute contact. E-mail tag is to specify type of contact. E-mail contact(s) (personal/generic) for project and/or institution. If multiple contacts, include comma separated list of emails.
creator_name	ACDD	Data creator's name. It could be person/group/institution	
creator_email		Data creator's e-mail	Please specify at least one attribute. The rest might be empty strings.
creator_url		Data creator's url	
contributor_name		Name of any individuals, projects, or institutions that contributed to the creation of this data.	
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_names.	Mandatory in case when data are produced as result of a collaboration.
Mandatory in case when the created data are observations or observation-based:			
platform	ACDD	Identify type of platform from which data is originally collec	Options: satellite, station, ship(s), aircraft(s)
platform_id	new	Name and version of the platform.	Combination of the ACDD's global attributes, platform and id.
satellite_algorithm	ACDD	Name and version of the algorithm for satellite data retriev	To be inserted if satellite-based data. To be inserted if satellite-based data.
satellite_sensor		Name of the satellite sensors used to provide satellite imag	
Part 2. Implicit global metadata structure from the input data as used to create final product			
indata_history	new	Metadata of input data set(s) represented by a Complex attribute that can be broken down into a number of other attributes (elements) and user-defined types. It allows: 1. Attribute(s) explicitly specified for an input data type (climatic/other, observed/modelled, raw/bias-corrected); 2. Multiple input data sets and types; 3. User-defined attributes considered as important. Advice: avoid inserting history attribute from input data.	Complex attributes replacing previously suggested attributes having the suffix in_var. An explicit lineage structure does not exist even for one input dataset issued from different sources of provenance such as observed/modelled. Thus, it is not feasible to store all of the information linked to processing operators and input data neither to find all desired/assumed attributes. Metadata-based approach to data lineage assumes relatively coarse-scale grained information (Woodruff and Stonebraker, 1997).
Part 3: Simple description of temporal and spatial characteristics of created data:			

Attribute	Convention	Definition	Notes/Examples
frequency	Not in CF and ACDD conventions, but used in the WCRP programms such as CMIP and CORDEX	Time step at which data are computed/recorded (as defined by time variable): day/mon/sea/ann	In CMIP and CORDEX, frequency is used to tag time step at which temperature and other measures of physical condition/proprety are given. * For climate indices, frequency can be semantically ambiguous. Frequency is defined as the number of times that an event occurs within a given period (dictionary.reference.com). But this is also an generic definition for the indices based on day count when a condition is fulfilled. However, the attribute is suggested to account for other climate products already published and to facilitate building a semantic web search.
time_coverage_start	ACDD	Time range field: Start time (the first value in time variable)	ISO 8601:2004 extended date format (Example: 2010-04-16T00:00:00). Leave empty
time_coverage_end		Time range field: End time (the last value in time variable)	
time_coverage_duration		Number of years/months/days between start and end date.	
time_coverage_resolutio		Alternative to the frequency attribute.	ISO 8601:2004 extended date format.
cdm_datatype	CF & ACDD	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 attrubute value is then introduced as a NetCDF variable except Grid:** **- The instructions are given in the CF 1.6 Convention [cfconventions.org], section: 9.1** **Features and feature types.** **- Examples can be found on www.nodc.noaa.gov/data/formats/netcdf/v2.0, by** **clicking on the links given in the table Feature Type Templates and Examples.**
domain		Name or short description of horizontal spatial data extent.	Example: 'EUR-11', for the EURO-CORDEX domain having 0.11 degree lat-lon res
geospatial_bounds	ACDD	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	
geospatial_lat_max		Southward bounding latitude of the domain	A value followed by the unit: degrees_east is recommended by CF convention.
geospatial_lat_resolution		A number value followed by the unit.	Example: '100 meters', '0.1 degree'. The INSPIRE standard accepts degrees as spatial scale (see inspire.ec.europe.eu)
geospatial_lon_min		Eastward bounding longitude of the domain.	
geospatial_lon_max		Westward bounding longitude of the domain.	A value followed by the unit: degrees_west is recommended by CF convention.
geospatial_lon_resolution		A number value followed by the unit.	Example: '100 meters', '0.1 degree' The INSPIRE standard accepts degrees as spatial scale (see inspire.ec.europe.eu)
history	CF & ACDD	Provides an audit trail for modifications to the created data.	Leave empty string if no modifications have been introduced after date of creati