Details on netCDF file generation and data submission for CCMI-2022

Data conversion to netCDF files

Model output is to be produced in netCDF version 4 format files that are compliant with the Climate and Forecast (CF) conventions. To reduce data volumes, the zlib compression native to netCDF4 should be used with settings of `deflate=1` and `shuffle=on` (Balaji et al., 2018).

The list of required metadata to be included in the netCDF files, the 'controlled vocabularies' or 'CVs', very closely follows that developed for CMIP6. A list and a brief description of the required global attributes included in each file is given in the table below. For further details on the required metadata, please see the CMIP6 documentation at https://goo.gl/v1drZl.

Note that the only variation from the list of required global attributes for CMIP6 we have instituted is to drop the requirement for the 'further_info_url', which for CMIP6 points to a webpage providing additional details on a specific simulation. In place of 'further_info_url', we ask groups to include information under the 'contact' attribute as used previously. Additionally, a collection of errata for simulations and variables will be collected on the CCMI webpage.

NOTE: An entry of CV under the 'Form' column indicates that the format of the specified attribute must exactly match that given in the CCMI2022_CV.json file, including matching the case (capital and lowercase) of all letters. For CCMI-1 a number of groups had files fail quality checks by using 'RefC1' or 'REFC1' instead of 'refC1'.

Table of required global attributes.

Global Attribute	Form	Options
activity_id	CV	'CCMI2022'
contact	Free-form	
Conventions	CV	'CF-1.7'
creation_date	*YYYY-MM-DD*T*HH:MM:SSZ*	
data_specs_version	CV	'01.00.00' or similar (see variable tables for exact number)
experiment	CV	'Hindcast', 'Baseline projection using…' (see CV.json for permitted options)
experiment_id	CV	'refD1', 'refD2', 'senD2-geo', 'senD2-ssp126', 'senD2-ssp370'
forcing_index	CV	'1', '2', '3',…
frequency	CV	'fx', 'mon', … (see variable entry in applicable table)
grid	Free-form	Briefly describe model horizontal grid and any regridding.

grid_label	CV	'gm', 'gn', 'gnz', 'gr' (see CV.json for permitted options)
initialization_index	CV	'1', '2', '3', …
institution	CV – as registered	'CNRM (Centre National de Recherche …', (see CV.json for permitted options)
institution_id	CV – as registered	'CNRM-CERFACS', (see CV.json for permitted options)
license	CV – with substitutions[1]	'CCMI2022 data produced by …'
mip-era	CV	'CMIP6'
nominal_resolution	CV	'50 km', '100 km', '250 km', … See CV.json for list of permitted options.
physics_index	CV	'1', '2', '3', …
product	CV	'model-output'
realization_index	CV	'1', '2', '3',…
realm	CV	'atmos', 'land'
source	CV – full model name and components as registered	'SOCOL : \natmos: SOCOL (T63…' See CV.json for list
source_id	CV – as registered	'CMAM', 'CNRM-MOCAGE-1-HR', … See CV.json for full list
source_type	CV	'AER', 'AGCM', 'AOGCM', 'CHEM'…. Space separated list of all that apply. See CV.json for full list and description.
sub_experiment	CV	'none'
sub_experiment_id	CV	'none'
table_id	CV	'Aday10Pt', 'Aday', 'AdayZ', 'Amon', 'AmonZ' or 'fx' as applicable to the particular variable.
tracking_id	CV – specified form[2]	'hdl:21.14100/…'
variable_id	CV	'ta', 'o3', 'ch4', … Variable name as specified in the data request.
variant_label	CV – specified form	'r3i1p1f1',… Combining realization_index, initialization_index, physics_index and forcing_index

[1]Following CMIP6, with some modifications, the license attribute should record the following statement with segments in square brackets ([…]) optional and appropriate text entered in segments with angle brackets (<…>):

CCMI2022 model data produced by <***Your centre name***> is licensed under a Creative Commons Attribution-[***loption***]ShareAlike 4.0 International License (https://creativecommons.org/licenses). Consult http://blogs.reading.ac.uk/ccmi/ for terms of use governing CCMI output, including citation requirements and proper acknowledgment. The data producers and data providers make no warranty, either express or implied, including, but not limited to, warranties of merchantability and fitness for a particular purpose. All liabilities arising from the supply of the information (including any liability arising in negligence) are excluded to the fullest extent permitted by law.

loption - The [***loption***] allows institutions to choose to use the Non-commercial version of this license by inserting the words "NonCommercial-" at this point, but note that this may significantly limit the range the use of the data in downstream applications. Please do not simply copy the statement above when writing data; Some text must be entered, some text is optional and the symbols '[ ] < >' should not appear in the licensing text.

[2]To allow for possible future publication to the ESGF, tracking_id should be of the form "hdl:21.14100/<uuid>" (e.g., "hdl:21.14100/02d9e6d5-9467-382e-8f9b-9300a64ac3cd"). The tracking_id should be unique for each file. The <uuid> should be generated using the OSSP utility which supports a number of different DCE 1.1 variant UUID options. Here version 4 (random number based) is required. Download the software from http://www.ossp.org/pkg/lib/uuid/.

Filename construction

Again, very closely following CMIP6, filenames are to be constructed following the template

filename =
<variable_id>_<table_id>_<source_id>_<experiment_id>_<variant_label>_<grid_label>[_<time_range>].nc

Except for invariant fields (fx), the time_range is required and is constructed as 'N_1-N_2', where N_1 and N_2 are integers of the form yyyy[MM[dd[hh[mm[ss]]]]]. Here yyyy, MM, dd, hh, mm, ss are integers specifying the year, month, day, hour, minute and second, respectively. The integers N_1 and N_2 specify the time of the first and last time sample in the file and are constructed with sufficient precision to unambiguously differentiate between two consecutive time samples. For example, a file of 3-D monthly average ozone on the original horizontal grid of the model:

o3_Amon_SOCOL_refD1_r1i1p1f1_gn_196001-201812.nc

and a file of zonal average temperature on the original latitude grid of the model:
ta_AmonZ_SOCOL_refD2_r1i1p1f1_gnz_196001-210012.nc

Data submission to the Centre for Environmental Data Analysis

Instructions for uploading model data to the Centre for Environmental Data Analysis (CEDA) can be found at https://help.ceda.ac.uk/article/4660-depositing-data-at-ceda-a-step-by-step-guide. Note that if you have been involved with data submission to CEDA (or BADC) in the past there is a chance you already have an account.

When you have gone to https://arrivals.ceda.ac.uk/intro/ and logged on, click on the button '+ New delivery' where you will be asked to create a name for the new data submission. Please name the delivery 'ccmi-2022' and targeted for delivery to CEDA. This will create a subdirectory in your account named 'ccmi-2022' and you will be directed to the 'Upload' page.

From there, please create a new directory below ccmi-2022 that gives the version number for this data submission. The version number should be constructed from a representative date for the current data submission, with the form 'vYYYYMMDD (e.g. 'v20210224'). The date used to construct the version number will not necessarily be the date on which the files were created or submitted, but should be representative of the current data submission and will be used to denote a consistent version of the data. This version number will be used to track any necessary resubmission of data to correct for errors and will be used as the <version> component in the final directory structure of the data archive in an identical manner as used for CMIP6.

Given the large number of files that each group will likely need to transfer, it is recommended that the transfer be done using rsync or ftp and not through the web browser. Details on how to proceed to upload data with rsync or ftp can be found following the 'Other upload methods' button on the 'Upload' page. Both the rysnc and ftp tabs have the ability to generate passwords. Use the password generated here when logging on through rsync or ftp to transfer the data. **Note that the new password generated for the ftp or rsync connection will not affect the password you use to log on to CEDA through a web browser.**

Please ensure that the data is placed into the ccmi-2022/vYYYYMMDD subdirectory. Users can choose to place all the uploaded files directly into that subdirectory, or create a directory tree below that point. Whichever is easiest.

Once the data has been uploaded, you do not need to proceed to 'Review submission' on the arrivals.ceda web page. The files will be checked and if they pass they will be moved into the ccmi-2022 archive. If there are problems, you will be notified.

> **NOTE: Please inform us of the CEDA username under which data is being submitted. This will allow us to target these accounts for automated monitoring, allowing for a more rapid quality checking and transfer of data into the archive.**

Balaji, V., K. E. Taylor, M. Juckes, B. N. Lawrence, P. J. Durack, M. Lautenschlager, C. Blanton, L. Cinquini, S. Denvil, M. Elkington, F. Guglielmo, E. Guilyardi, D. Hassell, S. Kharin, S. Kindermann, S. Nikonov, A. Radhakrishnan, M. Stockhouse, T. Weigel and D. Williams, Requirements for a global data infrastructure to support CMIP6, Geosci. Model. Dev., 11, 3659 – 3680, doi:10.5194/gmd-11-3659-2018, 2018.