Data checks prior to ESGF publication of CORDEX data.

October 28th, 2013.

The tables below list checks to be made prior to publication of data. The objective of the tests if to detect problems with the data prior to publication. Additional tests may be run, and data providers are advised to run their own tests. These tests will verify syntactical compliance with key parts of the CORDEX data specification. Ensuring that these tests are run on all data will lead to improved consistency across the CORDEX archive. Syntactical compliance should not be confused with validity: for example, setting the value of "plev" (the pressure coordinate) to 84000 for variable cll will make the file syntactically compliant with test T6.1c, but the file will only be compliant with the specification if the data has actually been interpolated correctly to 84000Pa. The data providers are responsible for ensuring that their data complies with the CORDEX archive specifications.[1] Some of the quantities tested are required for data management reasons, others to impose consistency across the archive to facilitate scientific analysis.

The European nodes of the ESGF archive are supported by the FP7 project IS-ENES2[2], and the list of checks below has been developed within that project. Entries with orange backgrounds indicate there are differing views within the project: feedback here is particularly welcome.

1 File name

Table 1.1: File name	
Test id	**Test**
T1.1	File name must consist of 8 or 9 components, separated by "_", followed by ".nc".

Table 1.2: File name components			
Test id	**Position in file name**	**Component id**	**Test**
T1.2a	1	VariableName	Contained in variable list.*
T1.2b	2	Domain	Contained in domain list.*
T1.2c	3	GCMModelName	Contained in driving model list.*
T1.2d	4	CMIP5ExperimentName	Contained in experiment list.*
T1.2e	5	CMIP5EnsembleMember	Of the form "rxiypz", for integers x,y,z.
T1.2f	6	RCMModelName	Contained in regional model list.*
T1.2g	7	RCMVersionId	Can only use characters a-z, A-Z, 0-9 and "-".
T1.2h	8	Frequency	One of: fx, sem, mon, day, 6hr, 3hr.
T1.2i	9	TimeRange	Required if Frequency is not "fx", see also Table 1.3

*: vocabulary lists are described in Appendix 2.

1 http://cordex.dmi.dk/joomla/images/CORDEX/cordex_archive_specifications.pdf Note that this document is under revision in order to clarify the presentation, but there is no intention to change the technical requirements.

2 https://verc.enes.org/ISENES2/ Infrastructure for the European Network of Earth System Modelling.

<table>
<tr><th colspan="4">Table 1.3: Time range</th></tr>
<tr><th>Test id</th><th colspan="3">Test</th></tr>
<tr><td>T1.3a</td><td colspan="3">TimeRange (see T1.2i) must consist of two integers (“Start” and “End”) separated by “-”.</td></tr>
<tr><td rowspan="8">T1.3b</td><td colspan="3">The rules governing “Start” and “End” depend on the Frequency (see T1.2h):</td></tr>
<tr><td rowspan="2">Frequency</td><td colspan="2">Pattern (for integers x,d,h):</td><td rowspan="2">Comment</td></tr>
<tr><td>Start**</td><td>End**</td></tr>
<tr><td>sem</td><td>xxx012</td><td>xxx011</td><td>10 years per file*</td></tr>
<tr><td>mon</td><td>xxx101</td><td>xxx012</td><td>10 years per file*</td></tr>
<tr><td rowspan="2">day</td><td>xxx10101</td><td>xxx512dd***</td><td rowspan="2">5 years per file*</td></tr>
<tr><td>xxx60101</td><td>xxx012dd***</td></tr>
<tr><td rowspan="2">3hr, 6hr</td><td>xxxx0101hh</td><td>xxxx12ddhh***</td><td rowspan="2">One year per file*</td></tr>
<tr><td>xxxx0101hh30</td><td>xxxx12ddhh30***</td></tr>
</table>

*: the amount of data in the file may be shorter for the first and last parts of the time series.
**: Restrictions in first and 2nd column do not apply to the first and last file in a series respectively, but the length (i.e. the number of characters) of both time range elements should be equal.
***: “dd” corresponds to the last day of December in the calendar of the simulation.

2 Required global attributes

Table 2: Required global attributes		
Test Id	**Global attribute name**	**Test**
T2.1	institute_id	Contained in institute list.*
T2.2	contact	Free text
T2.3	rcm_version_id	Equal to filename component RCMVersionId
T2.4	product	Equal to “output”
T2.5	CORDEX_domain	Equal to filename component Domain
T2.6	creation_date	Should specify the date of creation of the file. If a file which has previously published in ESGF is modified and re-published, this attribute should be updated.
T2.7	frequency	Equal to filename component Frequency
T2.8	model_id	Equal to filename component RCMModelName
T2.9	driving_model_id	Equal to filename component GCMModelName
T2.10	driving_experiment_name	Equal to filename component CMIP5ExperimentName
T2.11	driving_model_ensemble_member	Equal to filename component CMIP5EnsembleMember
T2.12	experiment_id	Equal to filename component CMIP5ExperimentName
T2.13	Conventions	“CF-1.x”, where x=4,5 or 6 – OR – make this attribute optional, but test its value if present.

*: See Appendix 2 for details on the institute list.

3 Optional Global Attributes

Table 3.1: Optional Global Attributes		
Test id	**Global attribute**	**Test (if attribute is present)**
T3.1	tracking_id	Should be changed when a file is modified*
T3.2	driving_experiment	Equal to "<driving_model_id>; <driving_experiment_name>; <driving_model_ensemble_member>" (or with "," as separator). – OR – omit from this list, as a bad value here will not cause archiving problems or problems of scientific interpretation.

*: subject to availability of a working service to execute this test.

4 Dimensions

Table 4: Dimensions		
Test id	**Dimension name**	**When required:**
T4.1	time	if frequency not "fx";
T4.2	plev	for variables clh, clm, cll , ua850, va850, ta850, hus850, ua500, va500, ta500, zg500, ua200, va200, ta200, zg200
T4.3	height	for variables tas, tasmax, tasmin, huss, sfcWind, sfcWindmax, wsgsmax, uas, vas;
T4.4	lat	for interpolated grids;
T4.5	lon	

5 Dimension attributes

Table 5: Dimension attributes		
	Attribute name	**Value or rule**
Test id	**time attributes (if dimension time present)**	
T5.1a	units	"days since 1949-12-01 00:00:00Z" or equivalent (e.g. "days since 1949-12-01") – OR – must be precisely "days since 1949-12-01 00:00:00Z"
T5.1b	standard_name	time
T5.1c	long_name	time
T5.1d	calendar	Must be a valid CF Convention calendar name.
T5.1e	time_bnds	Required for non-instantaneous fields (time means, sum and extrema), must equal "time_bnds" if present. See also T8.3.
	plev attributes (if dimension plev present)	
T5.2a	units	Pa
T5.2b	standard_name	air_pressure
T5.2c	long_name	pressure

T5.2d	positive	down
T5.2e	axis	Z
T5.2f	bounds	Required for variables clh, clm, cll; must equal “plev_bnds” if present. See also T8.3.
	height attributes (if dimension height present)	
T5.2a	units	m
T5.2b	standard_name	height
T5.2c	long_name	height
T5.2d	positive	up
T5.2e	axis	Z
	lat attributes (if dimension lat present)	
T5.4a	units	degrees_north
T5.4b	standard_name	latitude
T5.4c	long_name	latitude
	lon attributes (if dimension lon present)	
T5.5a	units	degrees_east
T5.5b	standard_name	longitude
T5.5c	long_name	longitude

6 Dimension and Bounds Values

Table 6a: Dimension Values			
Test Id	**Dimension**	**Value**	**Applicable for variables:**
T6.1a	plev	22000	clh
T6.1b		56000	clm
T6.1c		84000	cll
T6.1d		85000	ua850, va850, ta850, hus850
T6.1e		50000	ua500, va500, ta500, zg500
T6.1f		20000	ua200, va200, ta200, zg200
T6.2a	height	About 2	tas, tasmax, tasmin, huss
T6.2b		About 10 – OR – exactly 10	sfcWind, sfcWindmax, wsgsmax, uas, vas
		Rule	**When applicable**
T6.3	lat, lon	Must include domain specific grid (Table 9) – OR – must be exactly the domain specific grid (Table 9).	For interpolated domains

Table 6b: Values for plev_bnds (if present)

Test Id	Variable	Values
T6.4a	clh	[44000, 0]
T6.4b	clm	[68000, 44000]
T6.4c	cll	[100000, 68000]

7 Variable name and attributes

Table 7a: Variable name and attributes		
Test id	**Quantity tested**	**Value or rule**
T7.1	Variable name	Same as variableName component of file name (see T2.1a).
Test id	**Variable attribute**	**Value or rule**
T7.2	standard_name	As in VR or CMOR.***
T7.3	units	As in VR or CMOR.***
T7.4	long_name	As in VR or CMOR.***
T7.5	positive**	Required for a flux when the direction of the flux is otherwise ambiguous – OR – required for all fluxes
T7.6	cell_methods	Must contain “time: mean” for time averaged fields, or “time: point” for instantaneous fields. See also Table 8 for special cases of this attribute.
T7.7	missing_value	1.e+20f, if present* – OR – should always be present with this value.
T7.8	_FillValue	1.e+20f, if present*– OR – should always be present with this value.

*: these attributes are optional if no data is present.
: listed in VR*, but not required where standard name makes the sense of the flux unambiguous. Data should, of course, be archived in a way which is consistent with the CF standard_name definitions.
***: for details on VR and CMOR, see Appendix 2.

Table 7b: Special cases for cell_methods attributes		
Test id	**Variable**	**cell_methods string**
T7.9a	tasmin	time: minimum within days time: mean over days
T7.9b	tasmax, sfcWindmax	time: maximum within days time: mean over days
T7.9c	sund	time: sum within days time: mean over days

8 General rules

Table 8: General rules	
Test id	**Rule**
T8.1	Variables must be single precision
T8.2	Dimensions must be double precision.
T8.3	“plev_bnds” and “time_bnds” variables, if present, must follow the CF Convention rules for bounds variables.

9 Appendix 1: Domains.

Table 9: Names and boundaries of interpolated grids.						
Domain description	**Domain name**	**Grid spacing (degrees).**	**Grid Boundaries***			
			Longitudinal (degrees East)		**Latitudinal (degrees North)**	
			West	**East**	**South**	**North**
South America	SAM-44i	0.5	-106.25	-16.25	-58.25	18.75
Central America	CAM-44i	0.5	-124.75	-21.75	-19.75	35.25
North America	NAM-44i	0.5	-171.75	-22.25	12.25	76.25
Europe	EUR-44i	0.5	-44.75	65.25	21.75	72.75
Africa	AFR-44i	0.5	-25.25	60.75	-46.25	42.75
South Asia	WAS-44i	0.5	19.25	116.25	-15.75	45.75
East Asia	EAS-44i	0.5	62.75	175.75	-18.75	59.25
Central Asia	CAS-44i	0.5	10.75	140.25	17.75	69.75
Australasia	AUS-44i	0.5	88.75	207.25	-53.25	12.75
Antarctica	ANT-44i	0.5	-179.75	179.75	-89.75	-55.25
Arctic	ARC-44i	0.5	-179.75	179.75	48.75	89.75
Mediterranean	MED-44i	0.5	-20.75	51.75	25.25	57.25
Middle East and North Africa	MNA-44i	0.5	-26.75	75.75	-7.25	45.25
Middle East and North - high res.	MNA-22i	0.25	-26.625	75.625	-6.875	45.125
Europe high res.	EUR-11i	0.125	-44.8125	65.1875	21.8125	72.6875

*: position of grid points (i.e. grid cell centres).

10 Appendix 2: Vocabulary lists

List id	Label	Information
L1.1	institute	http://cordex.dmi.dk/joomla/images/CORDEX/RCMModelName.txt – second column.
L1.2	variable	Part of VR or CMOR tables. E.g. Column 2 of sheet “all” of VR (see

		L2 below).
L1.3	domain	“Domain name” column of Table 9 for interpolated domain, or the same values without the final “i” for native grid domains.
L1.4	driving model	http://cordex.dmi.dk/joomla/images/CORDEX/GCMModelName.txt
L1.5	experiment	Either “evaluation” (for experiments driven be re-analyses) or a valid CMIP5 experiment name.
L1.6	regional model	http://cordex.dmi.dk/joomla/images/CORDEX/RCMModelName.txt – first column.
L1.7	frequency	The following are valid: fx, sem, mon, day, 6hr, 3hr.
L2	VR	https://madwiki.dkrz.de/farm/CORDEXDataManagement?action=AttachFile&do=view&target=CORDEX_variables_requirement_table.xls or https://madwiki.dkrz.de/farm/CORDEXDataManagement?action=AttachFile&do=view&target=CORDEX_variables_requirement_table.pdf
L3	CMOR	CMOR tables for CORDEX: git://uv-cdat.llnl.gov/gitweb/cordex-cmor-tables.git

This document created by Martin Juckes, 28th Oct. 2013, derived from https://www.enes.org/data/projects/documents/quality-control-checks-for-the-cordex-archive