List of data aspects that have to be checked at data digestion time into an ESGF CORDEX archive

(based on the CORDEX archive design document (ADD[3]).

Additional tests that can be performed with the DKRZ programme are hidden in the spread sheet and are not mandatory.

Section 1	Tests	Comments, References to ADD[3] or VR[2]
File format	NetCDF4 classic data model with CF convention 1.4 or higher	ADD section 2
	deflated (compressed) with zlib	"
Directory structure (DRS)	Directory level names have to match the corresponding global attributes (see Section 2 below)	ADD section 5.3
	Directory level names have to comply with CORDEX CV as described in ADD or VR	"
Filename (DRS)	Filename elements[1] have to comply with CORDEX CV; i.p. <variable> element (1st element) has to be in the VR or ADD variable name list	ADD section 5.2, 6
	Filename elements have to match the corresponding global attributes (see Section 2 below)	ADD section 5.1 last §
	Files with time dependent variables have to have a period element (9th)	ADD section 5.1
	Separator in the period (9th) element of the filename has to be a dash	"
	Period element range has to be valid as described in ADD	"
	Period element dates have to be correctly formatted [YYYY[MM[DD[HH]]]]	"
	Period element dates have to be correctly cut-off as described in ADD [YYYY[MM[DD[HH]]]]	"
Variables	Variables have to have data type "float" (single precision)	ADD section 2
	long_name has to be as in VR	ADD section 6; VR sheet *'all'*
	Name of a variable has to remain unchanged across chunks	
Variable attributes	*standard_name* has to be as in VR	VR sheet *'all'*
	Unit attribute has to be as in VR	"
	positive = *"up"* or *"down"* (if present it must be as in VR regarding existence and value)	"
	Time dependent variables have to have a "*cell_methods = time* : <value>" attribute	VR sheets *'frequeny'*
	missing_value = 1.e+20f	ADD section 6; VR sheet *'all'*
	_FillValue = 1.e+20f	ADD section 6; VR sheet *'all'*

[1] Filename elements are the strings in the filename separated by "_"

[2] CORDEX_variable_requirements.xls (http://cordex.dmi.dk)

[3] CORDEX Archive Design document (http://cordex.dmi.dk/joomla/images/CORDEX/cordex_archive_specifications.pdf)

[5] http://wcrp-cordex.ipsl.jussieu.fr/images/pdf/cordex_regions.pdf

[6]File name DRS elements and **built rule**:
<VariableName>_<Domain>_<GCMModelName>_<CMIP5ExperimentName>_<CMIP5EnsembleMember>_
<RCMModelName>_<RCMVersionID>_<Frequency>[_<StartTime-EndTime>].nc (see ADD section 5.2)

(i) may pass if unmistakably formulated (example 'Near-Surface Specific Humidity' instead of 'Near-Surface Temperature' will not be accepted; however,
 'Specific Humidity at 2 m height' instead of 'Near-Surface Specific Humidity' may be accepted.

(ii) may pass if no grid cells with missing value exist

(iii) any format may be accepted

(iv) may pass if if not clh,clm,cll

(v) may pass if the pressue value is included in the variable short name

(vi) may pass if the flux direction is implied by the standard_name and the sign of data is correct

(vii) if present in old and new ESGF version

Section 2	Tests	Comments Reference to ADD[3] or VR[2]
Mandatory global attribute constraints	*model_id* = <RCMModelName>	correspondence between NetCDF attributes and DRS elements required ! see ADD section 5.1
	institute_id =<Institution>	
	driving_model_id = <GCMModelName>	
	CORDEX_domain = <Domain>	
	project_id = "CORDEX"	
	driving_model_ensemble_member = <CMIP5 Ensemble_member> or *'r1i1p1'* for *'evaluation'*	
	driving_experiment_name = <CMIP5ExperimentName> or *'evaluation'*	
	rcm_version_id = <RCMVersionID>	
	frequency = <Frequency>	
Mandatory global attributes	*institute_id*[7] from CV	ADD 2.1, 5.1; http://cordex.dmi.dk
	contact = "free text" valid e-mail	ADD section 2.1, 5.1
	rcm_version_id = "free text"; valid characters only	"
	product = "output"; single value	"
	CORDEX_domain from CV	ADD table 1,2; sheet *'CV'* in VR
	creation_date = "YYYY-MM-DDTHH:MM:SSZ" (or equivalen according to the UNIDATA conventions)	ADD section 2.1, 4
	creation_date value must change if the file is replacing a file already published in the ESGF	
	frequency: CV	sheet *'CV'* in VR
	model_id[7]: CV	see CV at http://cordex.dmi.dk
	driving_model_id : CV	"
	driving_experiment_name : CV	sheet 'CV' in VR; same as of the driving data or *'evaluation'*
	driving_model_ensemble_member : CV	sheet 'CV' in VR; same as of the driving data or *'r1i1p1'* for *'evaluation'*
	experiment_id : CV	sheet *'CV'* in VR
Constraints of optional recommended global attributes	*<driving_experiment >* = "*<driving_model_id>; <driving_experiment_name>; <driving_model_ensemble_member >*"	comma as separator is possible
	<experiment> = "description of experiment, free text"	
	tracking_id =<UUID>	"
	tracking_id must change if the file is replacing a file already published in the ESGF	"
	Conventions = "CF-1.4" or later	"
	other attributes belonging to the CMIP5 requirements are recommended to have valid CMIP5 values, e.g. *realm*	"

Section 3: Coordinate variables	Tests	Comments Reference to ADD[3] or VR[2]
time coordinate	has to be present for all but time constant variables; (double precision)	CF convention
time attributes	*unit* = "*days since 1949-12-01 00:00:00Z* " (or equivalent format according to the UNIDATA conventions)	ADD section 4
	calendar attribute has to be present	"
	calendar value has to be the same as that of the driving data/model	"
	time_bnds has to be present for non-instantaneous variables (i.e. if *cell_method* NOT = point; see VR)	"
plev coordinate	required for *clh, clm, cll , ua850, va850, ta850, hus850, ua500, va500, ta500, zg500, ua200, va200, ta200, zg200*	ADD section 3.3; CF convention
plev attributes	*standard_name* = "*air_pressure*"	"
	long_name = "*pressure*"	"
	units = "*Pa*"	"
	positive = "*down*"	"
	axis = "*Z*"	"
	bounds = "*plev_bnds*" required for *clh,clm,cll* only	"
plev value	22000, 56000, 84000 for *clh, clm, cll* respectively	"
	85000 for *ua850, va850, ta850, hus850*	"
	50000 for ua500, va500, ta500, zg500	"
	20000 for *ua200, va200, ta200, zg200*	"
plev_bnds	required for *clh, clm, cll* ; (double precision)	"
plev_bnds attributes	optional; if present must be the same as those of *plev*	"
plev_bnds values	[44000, 0], [68000, 44000], [100000, 68000] for *clh, clm, cll* respectively	"
height coordinate	required for *tas, tasmax, tasmin, huss, sfcWind, sfcWindmax, wsgsmax, uas, vas* ; (double precision)	"
height attributes	*long_name* = "*height*"	"
	standard_name = "*height*"	"
	units = "*m*"	"
	positive = "*up*"	"
	axis = "*Z*"	"
height value	about 2m for *tas, tasmax, tasmin, huss; about 10m for sfcWind, sfcWindmax,uas, vas, wsgsmax;*	"

computational/native grid coordinates	Tests	Comments Reference to ADD[3] or VR[2]
example: rotated lat/lon		see CF-1.4 convention; i.p. latitude and longitude coordinates have to be provided
rotated_pole:	grid_mapping_name = "rotated_latitude_longitude" ;	ADD 3.2
	grid_north_pole_latitude ;	"
	grid_north_pole_longitude ;	"
rlat	has to be present; (double precision);	"
rlat attributes	long_name = "rotated latitude" ;	"
	standard_name = "grid_latitude" ;	"
	units = "degrees" ;	"
	axis = "Y" ;	"
rlon	has to be present; (double precision)	
rlon attributes	long_name = "rotated longitude" ;	"
	standard_name = "grid_longitude" ;	"
	units = "degrees" ;	"
	axis = "X" ;	
other native grids	attributes as required by the NetCDF/CF-1.4 Convention or later	i.p. latitude and longitude coordinates have to be provided
CORDEX interpolated grid coordinates	Tests	Comments Reference to ADD[3] or VR[2]
lat	has to be present; (double precision)	to be filled in
lat attributes	long_name = " latitude"	"
	standard_name = "latitude"	"
	units = "degrees_north"	"
lat values	values and dimnsion must match those of table 2 in ADD	"; ADD table 2
lon	has to be present; (double precision)	"
lon attributes	long_name = "longitude"	"
	standard_name = "longitude"	"
	units = "degrees_east"	"
lon values	values and dimnsion must match those of table 2 in ADD	"; ADD table 2

cells with fading colors: can not (yet) be checked automatically, but is important to be correct