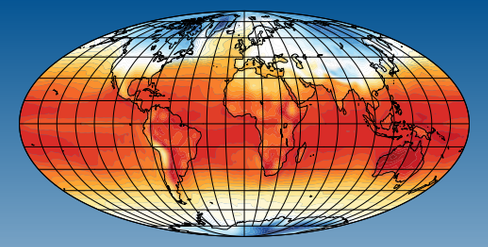
DKRZ
DEUTSCHES
KLIMARECHENZENTRUM

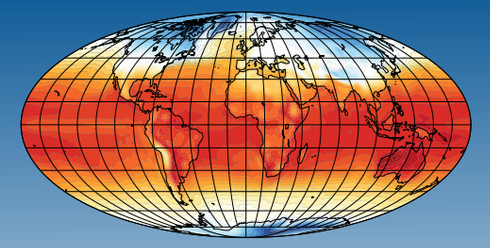


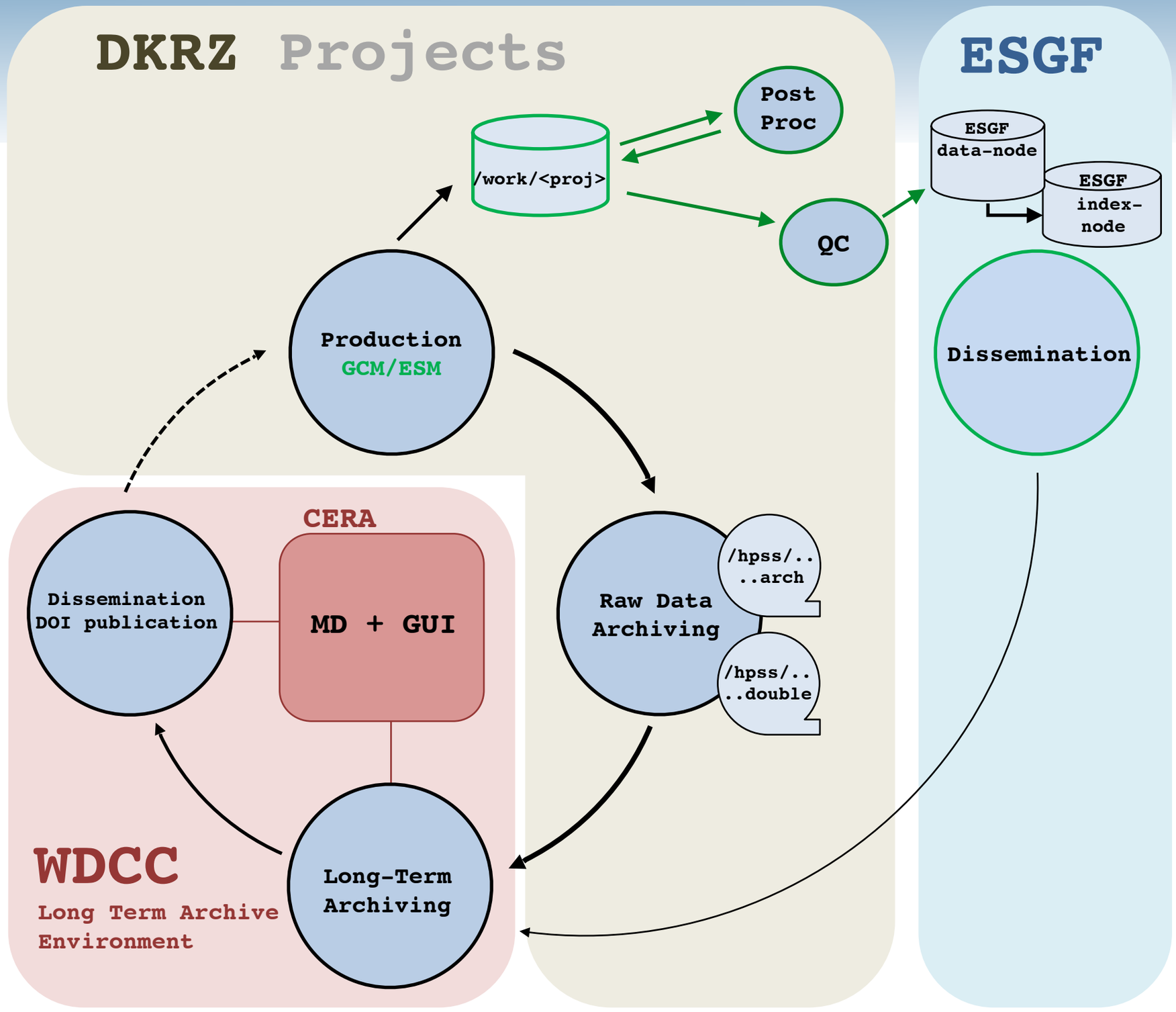
Data Standardization at DKRZ

- **Overview 1: DKRZ data management**
- **ESGF conditions for Data Admission**
- **What has to be specified?**
- **DRS example**
- **DRS elements and NetCDF attributes**
- **ESGF search interface 1**
- **ESGF search interface 2**
- **Overview 2: data standardization**
- **Using CMOR2 Software?**
- **Alternatives**
- **Issues ...**

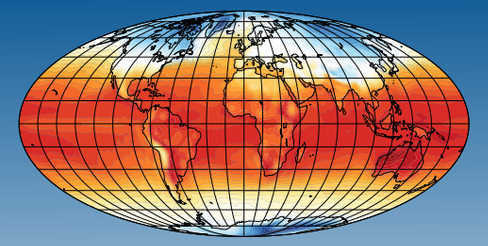
Jörg Wegner, DKRZ, 09.14

DKRZ data management





Data Standardization



metadata (cv):
filename &
directories(DRS);
NetCDF attributes

cmor tables

model output
observation data

projects:

obs4MIPs
CMIP5
GeoMIP
CORDEX
etc.

cdo
nco
etc.

script

cdi

cmor
calls

postprocessing

uuid lib

cmor lib

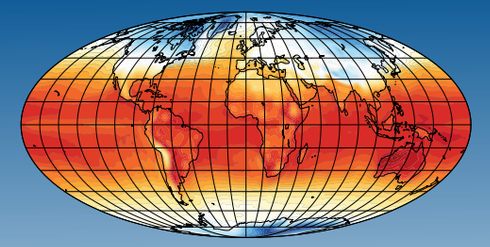
udunits lib

NetCDF4 CF1.4
(CMIP5, ...)

Quality
Check

sharing(ESGF)

DRS



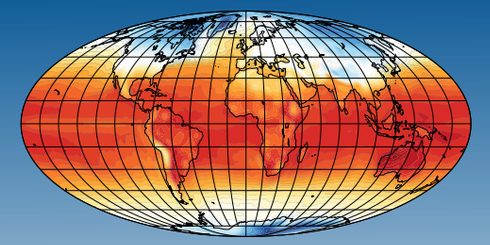
What has to be specified ?

- file format (e.g. NetCDFn, compression, single variable files, CF-1.n convention, …)

- file names implying the content
- directory structure for appropriate file grouping
- meta data (for processing tools, non-expert users)
- controlled vocabulary (CV) for accurate search

DRS(Data Reference Syntax)

DRS examples



project_id = "CORDEX"

product = "output"

CORDEX_domain = "AFR-44"

institut_id = "MPI-CSC"

driving_model_id = "MPI-M-MPI-ESM-LR"

experiment_id = "historical"

driving_model_ensemble_member = "r1i1p1"

model_id = "MPI-CSC-REMO2009"

rcm_version_id = "v1"

frequency = "sem"

float tas

tas_AFR-44_MPI-M-MPI-ESM-LR_historical_r1i1p1_MPI-CSC-REMO2009_v1_sem.nc

NetCDF attribute
=
DRS element !

<activity> = "CMIP5"

<product> = "output"

<institut> = "MPI-M"

<model> = "MPI-ESM-LR"

<experiment> = "historical"

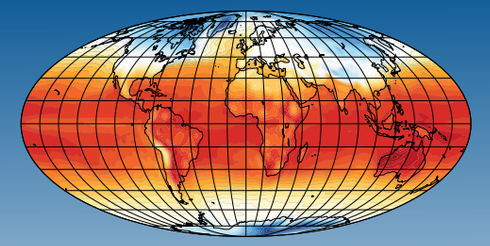
<frequency> = "mon"

<modeling_realm> = "ocean"

<variable_name> = tos

<ensemble_member> = "r2i1p1"

tos_Omon_MPI-ESM-LR_historical_r2i1p1_185001-200512.nc



ESGF Condition for Data Admission

- **ESGF is hosting data for model intercomparison projects;**
- **ESGF provides a GUI where data can be browsed, searched, and downloaded**

  => the data must

  - **be generated for a model intercomparison project (e.g. CMIP5, CORDEX, PMIPn, obs4MIPs, ana4MIPs, …)**
  - **belong to the agreed on set of project variables**
  - **be provided with agreed on common meta data in order to allow for common search criteria**

ESGF screen shot I

ESGF Earth System Grid Federation

is-enes INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

DKRZ DEUTSCHES KLIMARECHENZENTRUM

WDC CLIMATE

Home | Search | Tools | Account | Logout | Help

Current Selections

No search criteria selected

Search Categories

- Project
- Institute
- Model
- Instrument
- Experiment Family
- Experiment
- Time Frequency
- Product
- Realm
- Variable
- Variable Long Name
- CMIP Table
- CF Standard Name
- Ensemble
- Domain
- Driving Model
- Downscaling realisation
- Data Node

Search

Examples: *temperature*, *"surface temperature"*, *climate AND project:CMIP5 AND variable:hus*.
To download data: add datasets to your Data Cart, then click on *Expand* or *wget*.

Temporal Search
Clear search constraints and datacart
Search Help
Search Controlled Vocabulary

Search All Sites **Show All Replicas** **Show All Versions**

Display 10 datasets per page

Add All Displayed to Datacart Remove All Displayed from Datacart

Results | Data Cart

ESGF screen shot II

ESGF Earth System Grid Federation

is-enes INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

DKRZ DEUTSCHES KLIMARECHENZENTRUM

WDC CLIMATE

Home | Search | Tools | Account | Logout | Help

Current Selections

remove all
(x) project:CORDEX
(x) model:CCLM4-8-17
(x) experiment:evaluation
(x) variable:tas

Search Categories

- Project
- Institute
- Model
- Instrument
- Experiment Family
- Experiment
- Time Frequency
- Product
- Realm
- Variable
- Variable Long Name
- CMIP Table
- CF Standard Name
- Ensemble
- Domain
- Driving Model
- Downscaling realisation
- Data Node

Search

Examples: *temperature*, *"surface temperature"*, *climate AND project:CMIP5 AND variable:hus*.
To download data: add datasets to your Data Cart, then click on *Expand* or *wget*.

Temporal Search
Clear search constraints and datacart
Search Help
Search Controlled Vocabulary

Search All Sites **Show All Replicas** **Show All Versions**

< 1 > displaying 1 to 7 of 7 search results

Display 10 datasets per page

Add All Displayed to Datacart Remove All Displayed from Datacart

Results | Data Cart

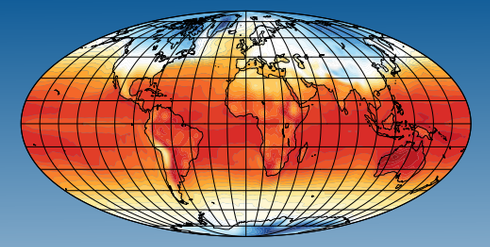
cordex.output.AFR-44.CLMcom.ECMWF-ERAINT.evaluation.r1i1p1.CCLM4-8-17.v1.day.tas
Data Node: carbon.dkrz.de
Version: 20140401
No description available.
Further options: Add To Cart Visualize and Analyze

cordex.output.AFR-44.CLMcom.ECMWF-ERAINT.evaluation.r1i1p1.CCLM4-8-17.v1.mon.tas
Data Node: carbon.dkrz.de
Version: 20140401
No description available.
Further options: Add To Cart Visualize and Analyze

cordex.output.AFR-44.CLMcom.ECMWF-ERAINT.evaluation.r1i1p1.CCLM4-8-17.v1.sem.tas
Data Node: carbon.dkrz.de
Version: 20140401
No description available.
Further options: Add To Cart Visualize and Analyze

cordex.output.EUR-44.CLMcom.ECMWF-ERAINT.evaluation.r1i1p1.CCLM4-8-17.v1.day.tas
Data Node: carbon.dkrz.de
Version: 20140424
No description available.
Further options: Add To Cart Visualize and Analyze

cordex.output.EUR-44.CLMcom.ECMWF-ERAINT.evaluation.r1i1p1.CCLM4-8-17.v1.mon.tas
Data Node: carbon.dkrz.de
Version: 20140424
No description available.
Further options: Add To Cart Visualize and Analyze

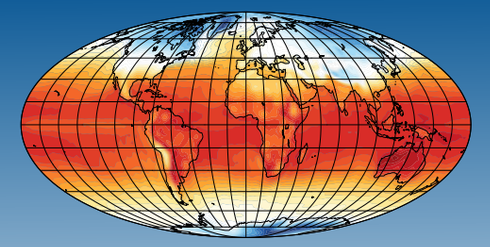


What is CMOR2?

- cmor = climate model output rewriter
- a software library which comprises a set of functions to produce CF1.4 compliant NetCDF file (FORTRAN, C or Python)
- based on NetCDF4 libraries
  - a set of Tables supports cmor2 with project defined settings for dimensions, variables and attributes.
  - the udunits2 library checks the variable units.
  - the uuid library gives every file a unique identifier.
  - is as first quality check

the amount of work only pays for bigger projects

was designed for CMIP5 and some functionality is not changable yet



Alternatives

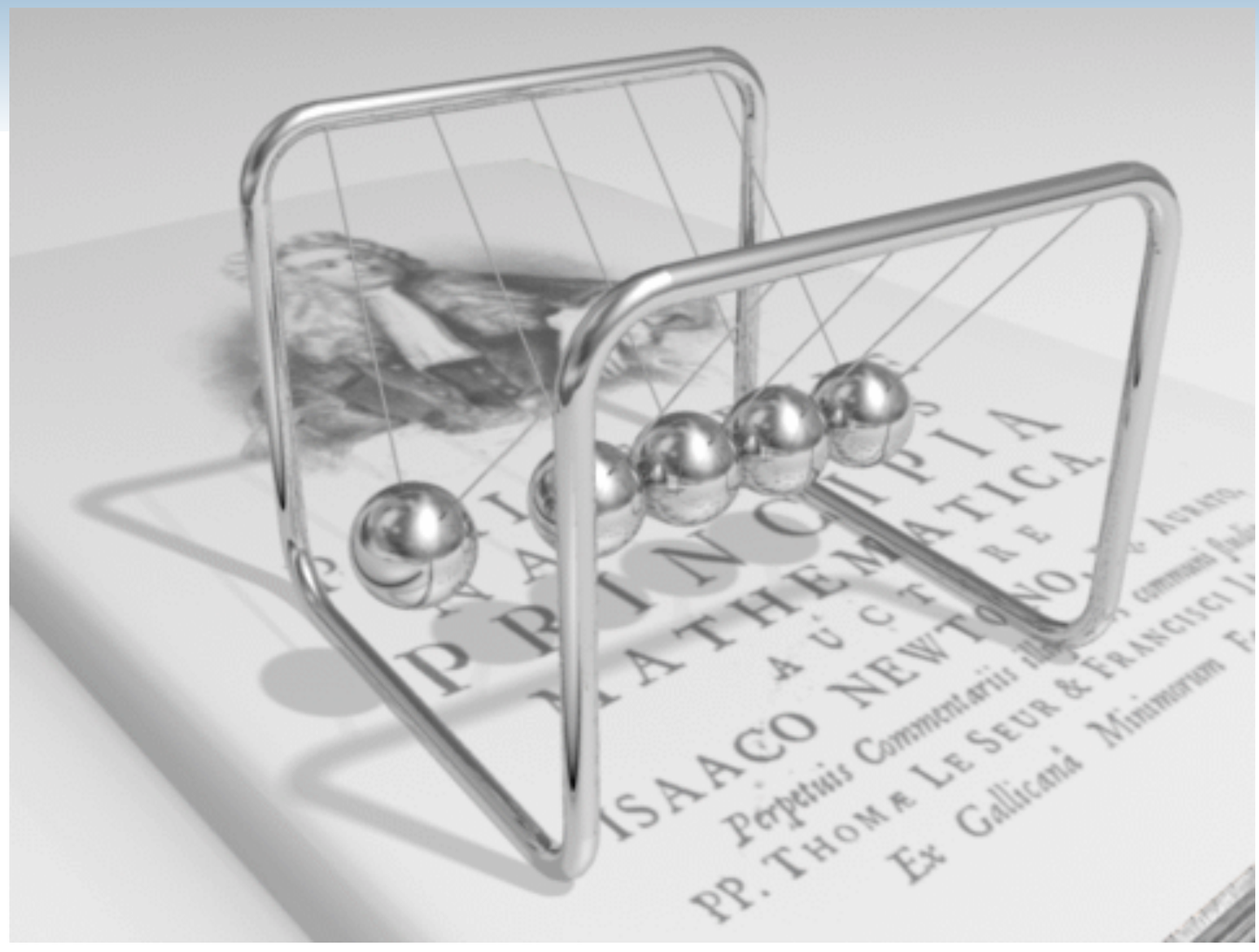
cdms, xconv, ncl ...

Use cdo- or nco- operators in conjunction with ncdump and ncgen to create the netcdf file.

-ncdump & ncgen come with the NetCDF libraries
-cdo: https://code.zmaw.de/projects/cdo/
-nco: http://nco.sourceforge.net

Use NetCDF4 libraries with FORTRAN, C or Python, and write your own program.

Thanks for listening



Questions?

data standardization