



DKRZ

DEUTSCHES
KLIMARECHENZENTRUM

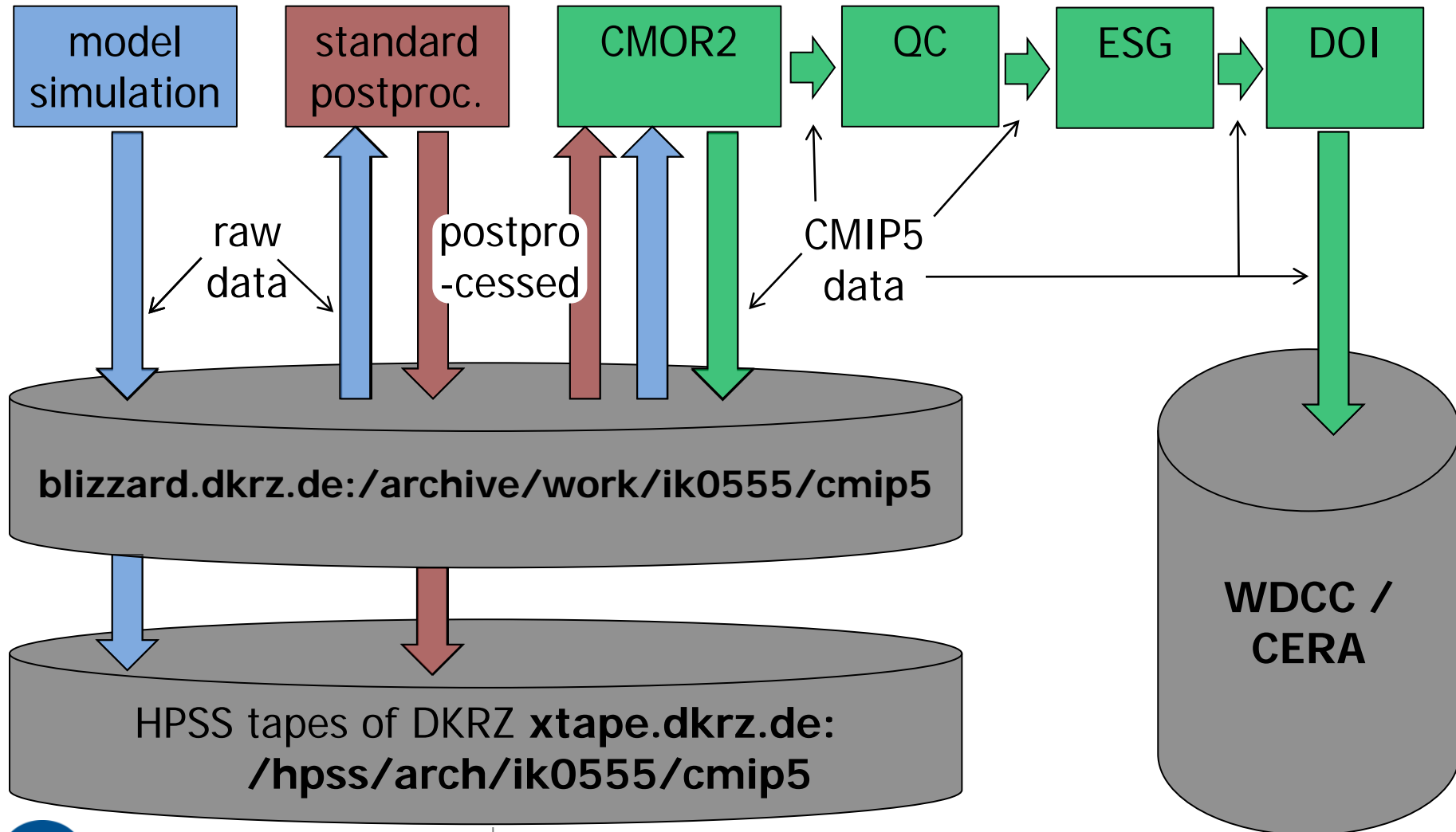
Local Disk and Tape Archives at DKRZ

Dr. Heinz-Dieter Hollweg / Heinrich Widmann

Outline

- Overview
- Data structure
 - DRS paths
 - DRS filenames
- Local storage
- Tape archive
- Quality control (QC)
- Getting Help

Overview



Overview

Raw data

Output of the MPI-ESM-[LR | P | MR]
coupled model components:

ECHAM6

JSBACH

MPIOM

HAMOCC

Resolution:

| | | | |
|----|------------------|--------|---------|
| LR | low resolution | T63L47 | GR15L40 |
| P | paleo | T63L47 | GR15L40 |
| MR | mixed resolution | T63L95 | TP04L40 |

Standard post-processed Data

by the **Climate Data Operators (CDO)**
and *the afterburner program*

CMIP5 Data

- **standard_output (pdf, xls, xlsx)** defines requested variables & values in the document: http://cmip-pcmdi.llnl.gov/cmip5/output_req.html
- CMOR2 software generates MPI-M CMIP5 data
- compliant to CF-1.4 convention
- data format: NetCDF (version ≥ 3.6)
- quality controlled

Data Structure: DRS

Defined in:

CMIP5 Data Reference Syntax (DRS) and Controlled Vocabularies,

Taylor et al., 2011,

http://cmip-pcmdi.llnl.gov/cmip5/output_req.html#req_format

Document: cmip5_data_reference_syntax.pdf

Data Structure: DRS

CMIP5 Data Reference Syntax (DRS)

<a>/<p>/<i>/<m>/<x>/<f>/<r>/<v>/<s>

| | DRS Component | Value | | DRS Component | Value |
|-----|-----------------|--------|-----|---------------|------------|
| <a> | activity | CMIP5 | <p> | product | output |
| <i> | institute | MPI-M | <m> | model | MPI-ESM-LR |
| <x> | experiment | amip | <f> | frequency | mon |
| <r> | modeling realm | atmos | <v> | variable name | tas |
| <s> | ensemble member | r1i1p1 | | | |

(read: fixed, green: mutable)

blizzard file-system (example)

/work/ik0555/cmip5/archive/ ➤

CMIP5/output/MPI-M/MPI-ESM-LR/amip/mon/atmos/tas/r1i1p1

Data Structure: DRS

CMIP5 Filename Encoding

<v>_<M>_<m>_<x>_<e>[_t].nc

| | Component | Options http://cmip-pcmdi.llnl.gov/cmip5/output_req.html |
|-----|-----------------|--|
| <v> | variable name | Doc.: standard_output (pdf, xls, xlsx) Column: output variable name |
| <M> | MIP table | fx, Oclim, Oyr, Amon, Omon, Lmon, LImon, Olmon, æerø, day,6hrLev, 6hrPlev, 3hr, cfMon, cfOff, cfDay, cf3hr, cfSites |
| <m> | model | MPI-ESM-LR, MPI-ESM-MR, MPI-ESM-P |
| <x> | experiment | Doc.: cmip5_data_reference_syntax.pdf -> Appendix |
| <e> | ensemble member | r<N>i<M>p<L> with r := realization, i:=initialization indicator, p:= perturbed physics and N, M<2, L<2, M=L. |
| <t> | temporal subset | Format: T = yyyy[mm][dd[HH][MM]] [-clim] Period: T ₁ -T ₂ , always to the end of the period |

CMIP5 Filename Encoding

„... enough (and just enough) of the suffixes should be added to unambiguously resolve the sub-temporal interval...“

Examples

orog_fx_MPI-ESM-LR_piControl_r0i0p0.nc

pp_Oyr_MPI-ESM-LR_piControl_r1i1p1_1850.nc

pp_Oyr_MPI-ESM-LR_piControl_r1i1p1_1850-1899.nc

co2mass_Amon_MPI-ESM-LR_piControl_r1i1p1_185001-284912_clim.nc

cl_cfSites_MPI-ESM-LR_piControl_r1i1p1_198401010000-19881231230.nc

Tape Archive

Access: pftp, sftp, or gridftp

doc: <http://www.dkrz.de/Nutzerportal-en/doku/hpss>

- pftp
 - only within house, i.e. a DKRZ / ZMAW user account is required
 - automatic login via .netrc file
- sftp
 - from outside DKRZ / ZMAW
 - sftp <user>@xtape.dkrz.de
- gridftp
 - better performance than ssh / scp / sftp
 - globus-url-copy gsiftp://xtape.dkrz.de/hpss/arch/ik0555/cmip5/ experiments / *exp-name* / file ↵ file:///home/<mylocaldir>/<mylocalfile>

Tape Archive

/hpss/arch/ik0555/cmip5/experiments ↷

1pctCO2_r1i1p1-MR
1pctCO2_r1i1p1-P
1pctCO2_r1i1p1-LR
abrupt4xCO2_r1i1p1-P
⋮
historical_r1i1p1-LR
⋮
lgm_r1i1p1-P
midHolocene_r1i1p1-P
⋮
rcp26_r1i1p1-LR
⋮
rcp85_r1i1p1-MR
⋮
sstClim_r1i1p1-P

Tape Archive

/hpss/arch/ik0555/cmip5/experiments/historical_r1i1p1-LR ➤

historical_r1i1p1-LR_echam6_ATM_mm_1850-2005.grb
- "- _BOT_mm_1850-2005.grb
- "- _LOG_mm_1850-2005.grb

⋮

- "- _co2_1850-1859.grb

⋮

- "- _co2_2000-2005.grb

- "- _co2_dm_1850-2005.grb

- "- _co2_mm_1850-2005.grb

- "- _echam_1850.grb

⋮

- "- _echam_2005.grb

- "- _echam_trdiag_1850-1859.tar

⋮

- "- _echam_trdiag_2000-2005.tar

(tracer diagnostic, NetCDF format)

Where and How long?

▪ Local Disk

- all CMIP5 data unlimited
</work/ik0555/cmip5/archive/>...CMIP5-DRS...
- Only a few raw data unlimited,
most raw data is going to vanish on short notice.

▪ Tape Archive

- all the raw and forcing data
- **NO** CMIP5 data

Quality Controlled?

▪ CMIP5 data: generally yes

(But perhaps sharing space with still unprocessed data or withdrawn and reproduced CMIP5 files.)

- ### ▪ Raw data is only indirectly checked,
- because: CMIP5 data is derived from raw

Quality Control

Purpose

Technical check of:

- meta data between files and standard_output.xls
- checks on data
- *(scientific evaluation / verification by MPI-M)*

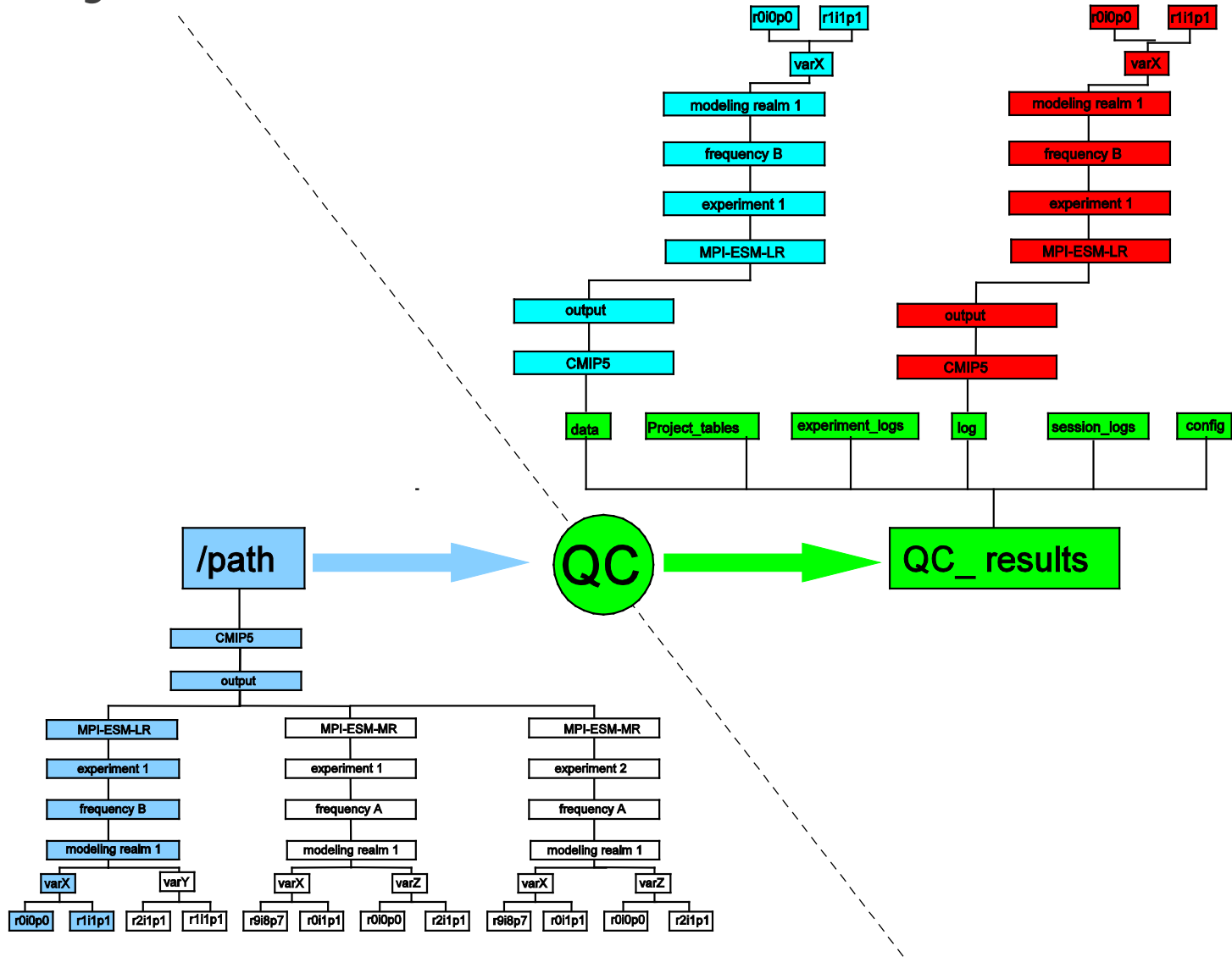
Input

NetCDF files

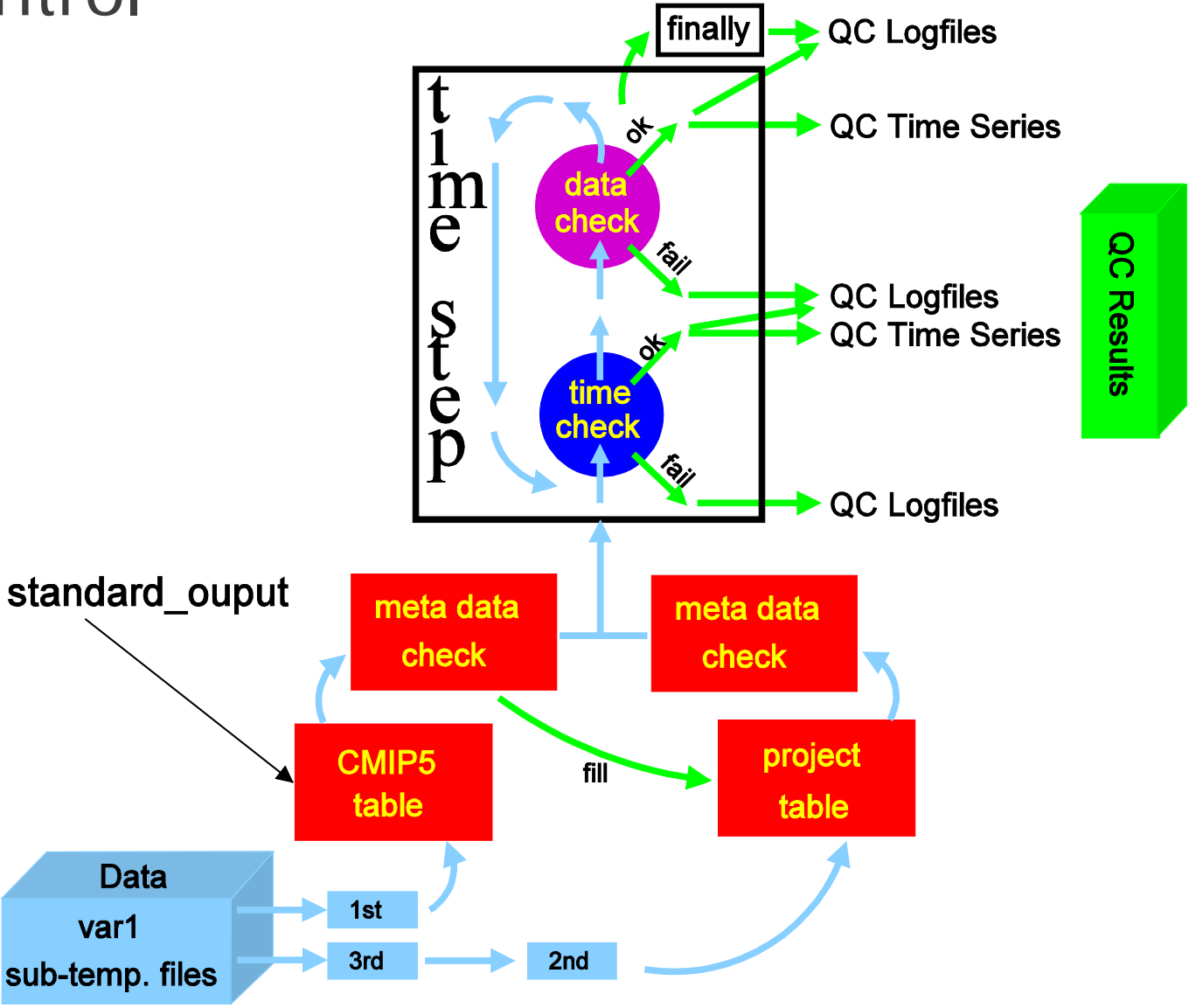
Output

QC results are stored akin to the DRS directory structure

Quality Control



Quality Control



Meta Data Check

- **Dimensional Properties**
 - Standard name, long name, output name
 - Units
 - Requested values (e.g. level pressure)
 - Requested variables with boundaries and requested ranges
- **Properties of Variables**
 - Standard name, long name, output name
 - Dimensions
 - Units
- **Consistency across sub-temporal files** (required)
- **Consistency between successive experiments** (optional)
- **Check of filename encoding**

Time Data Check

- Time Values and Steps
 - Negative time steps
 - Identical time values
 - Gaps
- Time Bounds
 - Negative/zero range
 - Overlapping ranges
 - Gaps
- Check of the time-stamp in the filename
- Checks also across sub-temporal files / experiments
- Duration of experiments

Data of Variables

■ Checks / Detection

- time values **entirely** of **constant** or **_Fill_Value**.
- **infinity** or **not-a-number** (Inf / NaN)
- time values with **repeated data**
- **extreme outliers**
- consistency of **auxiliary data** (e.g. lon/lat values) across files
- sub-temporal files **entirely** of **constant** or **_FillValue** data

■ Additional information

- **global averages** and **standard deviations** (optionally area weighted)
- **global minimum/maximum** per time step and of entire experiment

Each and every piece of **meta data**, **time value** and **variable data**

is touched

Getting Help

CMIP5

cmip5-mpi-esm@dkrz.de

WDCC/CERA

data@dkrz.de