

Langzeitarchivierung und DataCite-Datenpublikation

Long-term Archiving and DataCite Data Publication

*Martina Stockhause
and WDC Climate Staff*

SPONSORED BY THE



Federal Ministry
of Education
and Research



Outline

- CMIP5 Quality Control Procedure
- Data Creator's Role:
Participation in DataCite Publication Process
- CMIP5 long-term archive
- Data User's Perspective:
Citation of CMIP5 Data

CMIP5 Quality Control Procedure

Quality Level in CMIP5

Quality Level is assigned to CMIP5 experiments

- **Quality Level 1 – all data:**
separate technical checks of data and metadata
- **Quality Level 2 – output1 data:**
consistency checks – QC tool of MPI-M QC used
- **Quality Level 3 / DataCite DOI**
– **output1 data (+ MPI-M output2 data):**
cross- and double checks of data and metadata
DataCite (datacite.org) DOI data publication process

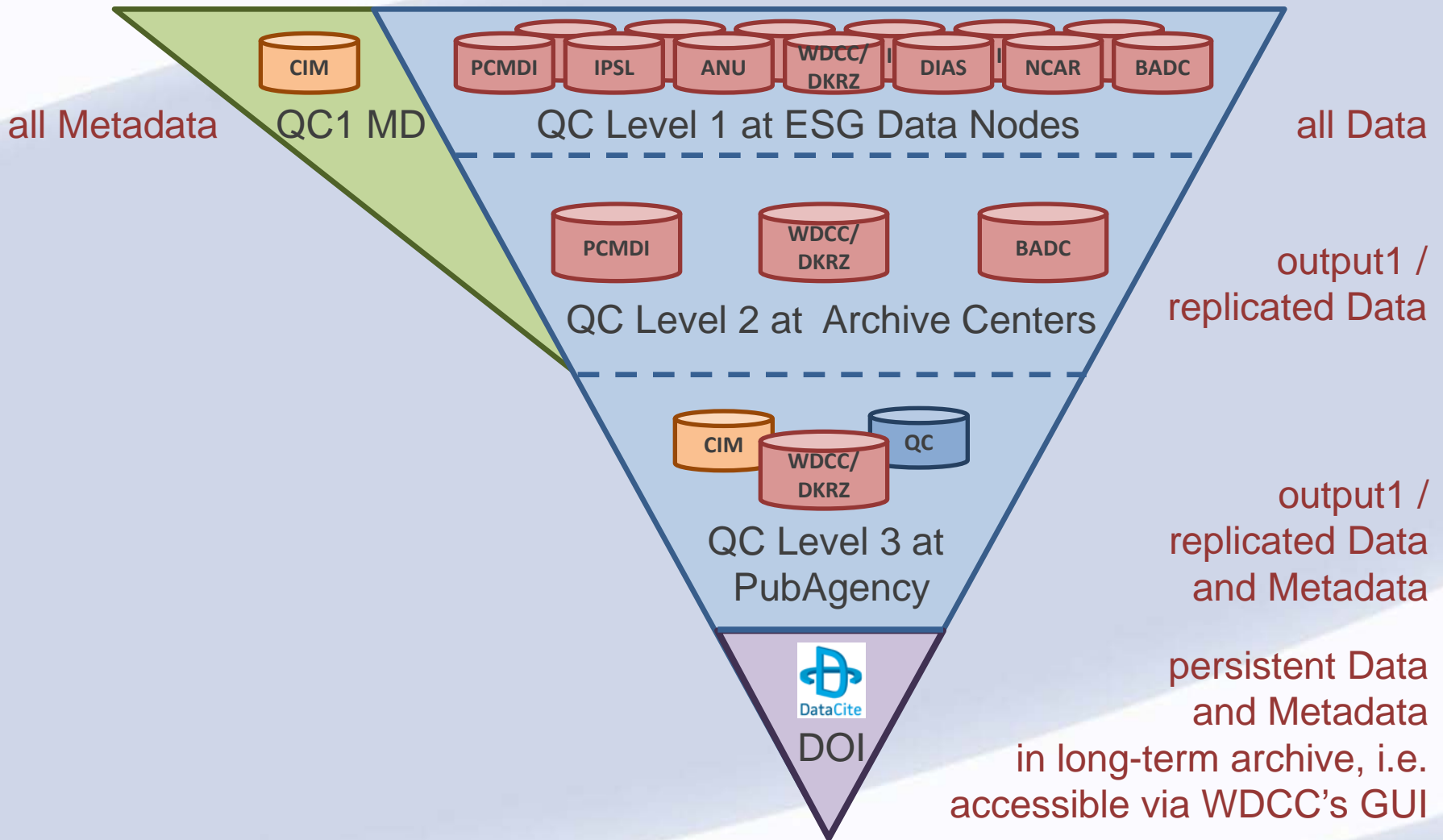
Data Changes
without user notification

Data Changes
with user notification

Data Persistence &
Data Citation by DOI

e.g. [DOI:10.1594/WDCC/CMIP5.NCCNMpc](https://doi.org/10.1594/WDCC/CMIP5.NCCNMpc)
(DOI Resolver: <http://dx.doi.org/10.1594/WDCC/CMIP5.NCCNMpc>)

Federated Quality Procedure



Status of CMIP5 QC

QC Status of CMIP5 Experiments (23.02.2012):

- Quality Level 1: **597** Experiments (on Data Nodes)
- Quality Level 2: **30** Experiments (started: 169)
- Quality Level 3 /
DataCite DOI: **1** Experiment (started: 2)

Status MPI-M data: deadline for data changes set – amip-LR in DOI process

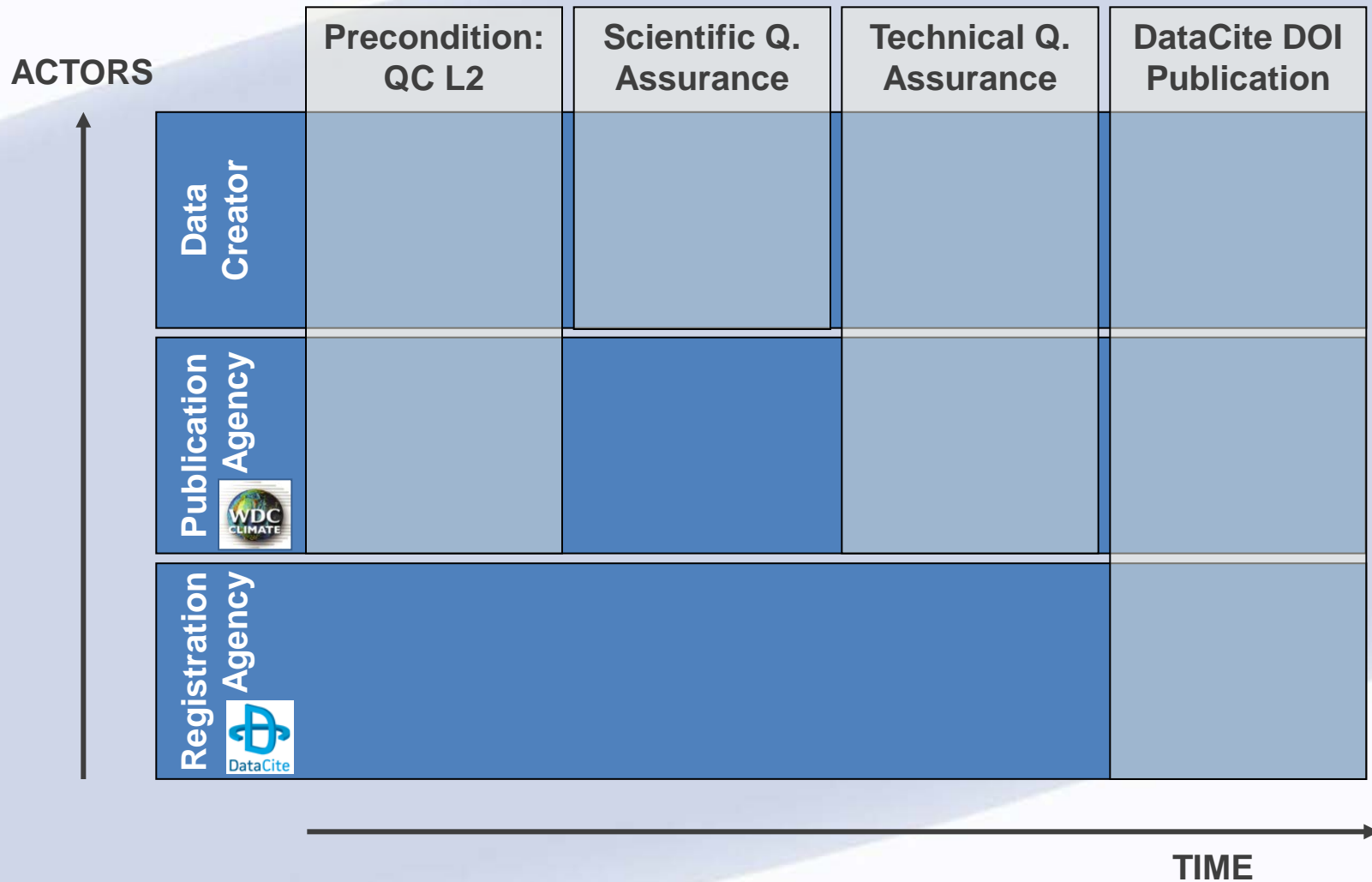
Reasons for the delay of the DOI Data Publications:

- data changes and additions
- slow data replication due to narrow band widths
- CIM metadata on simulations unavailable
- changes in the data and metadata infrastructure
- increased complexity of cross-checks due to a high heterogeneity in the data nodes

Status: <http://cera-www.dkrz.de/WDCC/CMIP5/QCStatus.jsp>

Data Creator's Role: DataCite Publication Process

QC L3: DataCite DOI Publication Workflow



Scientific Quality Assurance

QC Result Details for *cmip5/output1/MPI-M/MPI-ESM-LR/piControl*

The listed CMIP5 [<http://cmip-pcmdi.llnl.gov/cmip5/>] experiments are in the quality control process [<http://cmip5qc.wdc-climate.de/>]. Exception codes are explained at <http://cera-www.dkrz.de/CMIP5/QC/2/qc2list.html>, where F exception codes are available as 2D fields in the netCDF result files as sum of all occurred F exceptions codes in a grid cell. All other (W/E) exception codes are documented in the ASCII files: *_experiment.log and the variable exception messages for a specific atomic dataset (text file for a *.nc entry). If available, *.pdf visualizes the QC results of all *.nc files.

| Short name | Version | Status | Timestamp | Text | Binary |
|---|-----------|---------------|---------------------|------|--------|
| qc_areacello_fx_MPI-ESM-LR_piControl_r0i0p0.nc | v20111006 | QCL2_assigned | 2011-10-14T03:11:56 | | |
| qc_baresoilFrac_Lmon_MPI-ESM-LR_piControl_r1i1p1.nc | v20111006 | QCL2_assigned | 2011-10-14T13:37:03 | | |
| qc_baresoilFrac_Lmon_MPI-ESM-LR_piControl_r1i1p1.nc | v20111119 | QCL2_assigned | 2011-11-21T17:58:39 | | |
| qc_bfe_Omon_MPI-ESM-LR_piControl_r1i1p1.nc | v20111006 | QCL2_assigned | 2011-10-15T07:52:10 | | |
| qc_bsi_Omon_MPI-ESM-LR_piControl_r1i1p1.nc | v20111006 | QCL2_assigned | 2011-10-15T11:39:25 | | |
| qc_burntArea_Lmon_MPI-ESM-LR_piControl_r1i1p1.nc | v20111006 | QCL2_assigned | 2011-10-14T12:49:08 | | |
| qc_calc_Omon_MPI-ESM-LR_piControl_r1i1p1.nc | v20111006 | QCL2_assigned | 2011-10-15T07:57:11 | | |
| qc_calc_Oyr_MPI-ESM-LR_piControl_r1i1p1.nc | v20111006 | QCL2_assigned | 2011-10-15T13:18:34 | | |
| qc_cct_Amon_MPI-ESM-LR_piControl_r1i1p1.nc | v20111006 | QCL2_assigned | 2011-10-14T03:15:16 | | |
| qc_cct_Amon_MPI-ESM-LR_piControl_r1i1p1.nc | v20111119 | QCL2_assigned | 2011-11-21T08:49:33 | | |
| qc_chl_Omon_MPI-ESM-LR_piControl_r1i1p1.nc | v20111006 | QCL2_assigned | 2011-10-15T08:02:16 | | |
| qc_chl_Oyr_MPI-ESM-LR_piControl_r1i1p1.nc | v20111006 | QCL2_assigned | 2011-10-15T13:29:49 | | |
| qc_cl_Amon_MPI-ESM-LR_piControl_r1i1p1.nc | v20111006 | QCL2_assigned | 2011-10-14T10:23:45 | | |
| qc_cl_Amon_MPI-ESM-LR_piControl_r1i1p1.nc | v20111119 | QCL2_assigned | 2011-11-21T09:44:13 | | |

Additional information download

QC_add_info_2283264: [QCL2_results_cmip5_output1_MPI-M_MPI-ESM-LR_piControl.tar](#) (647085k)
 QC_add_info_2283265: [QCL2_logfile_cmip5_output1_MPI-M_MPI-ESM-LR_piControl.txt](#) (2162k)
 QC_add_info_2283266: [QCL2_exceptions_cmip5_output1_MPI-M_MPI-ESM-LR_piControl.txt](#) (19669k)
 QC_add_info_2283409: [CIM_cmip5_output1_MPI-M_MPI-ESM-LR_piControl.xml](#) (442k)
 QC_add_info_2290283: [QCL2_results_cmip5_output1_MPI-M_MPI-ESM-LR_piControl.tar\(2\)](#) (323550k)
 QC_add_info_2290284: [QCL2_logfile_cmip5_output1_MPI-M_MPI-ESM-LR_piControl.txt\(2\)](#) (3362k)
 QC_add_info_2290285: [QCL2_exceptions_cmip5_output1_MPI-M_MPI-ESM-LR_piControl.txt\(2\)](#) (19669k)

[Back to summary page...](#)

This page is hosted at [WDCC](#), please send technical inquiries to data@dkrz.de. [Impressum / Legal notice](#)

Data Creator's QC Workflow

1. Confirmation of QC L2 assignment
 - 1a. Publication Agency starts QC L3 Process*
2. Scientific Quality Assurance via WDCC's Data Publication System atarrabi
 - 2a. Publication Agency performs Technical Quality Assurance and publishes DOI via DataCite*
3. Notification of DOI by Publication Agent

<http://cera-www.dkrz.de/atarrabi2>

CMIP5 Long-term Archive

CMIP5 Long-term Archive at WDCC

The screenshot displays the CERA Web Interface for file details. The browser address bar shows the URL: `http://127.0.0.1:7101/Cera_WebInterf...`. The page title is "Internal / CERA Data Access (Detail)". The header includes the logos for "M&D Modelle & Daten", "World Data Center for Climate, Hamburg", and "WDC". Navigation links include "CERA UI Home", "WDCC Home", and "Impressum".

The main content area displays details for the file: `NCCNMpcDADclt111v110901_clt_day_NorESM1-M_piControl_r1i1p1_08000101-08491231.nc`. The details table is as follows:

| | |
|---------------------|--|
| Filename | cmip5/output1/NCC/NorESM1-M/piControl/day/atmos/day/r1i1p1/v20110901/clt/clt_day_NorESM1-M_piControl_r1i1p1_08000101-08491231.nc |
| Start date | 0800/01/01 00:00:00 |
| Number of timesteps | - |
| Size (bytes) | 1009603500 |
| Comment | [tracking-id: 41042191-bc7e-405b-a37f-8bdbb59697c1] [checksum: b39bd3ea3a62521882332e0a35963c15] |

Below the table is a "Download" button. A "M&D Data Section" checkbox is also visible.

On the left sidebar, the "Download from" button is circled in orange. Other sidebar elements include "General Info", "External CMIP", "CERA files for", "M&D Data Section", "Creation date", "Download from", "Plot metadata", "Parameters", and "Topic".

M. Kurtz

CMIP5 Long-term Archive

CMIP5 experiments with DOI are long-term archived

- DOI catalog information
- DOI points to the DOI landing page
- DOI landing page provides basic information and links to additional information and data download
- WDCC long-term archive data access via CERA

Work in progress:

- Improve CMIP5 data access via CERA GUI
- Change data access link on DOI landing page to point into ESG gateway for next ESG gateway generation

User's Perspective: Data Citation

Data

Quality
CMIP

DOI Landing Page

Firefox
Citation
cera-ww

WCRP
World Climate R

CMIP5 Citation Info

Use this service to receive the tracking_id out

Note:
As long as no DOI is assigned refer to <http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=NCCNmpc>

Tracking ID:

CMIP5 Citation

Use this service to receive the tracking_id out

Note:
As long as no DOI is assigned refer to <http://cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=NCCNmpc>

Tracking ID: 91f59

This page is hosted at

Firefox
Citation
CERA - entry information for NCCNmpc
cera-www.dkrz.de/WDCC/CMIP5/Compact.jsp?acronym=NCCNmpc

WCRP World Climate Research Programme

PCMDI

British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL

WDC CLIMATE

DOI for Scientific and Technical Data
10.1594/WDCC/CMIP5.NCCNmpc

Title
cmip5 output1 NCC NorESM1-M piControl, served by ESGF

Citation
Bentsen, Mats; Bethke, Ingo; Debernard, Jens; Drange, Helge; Heinze, Christoph; Iversen, Trond; Kirkevåg, Alf; Seland, Øyvind; Tjiputra, Jerry (2011): cmip5 output1 NCC NorESM1-M piControl, served by ESGF. World Data Center for Climate. DOI:10.1594/WDCC/CMIP5.NCCNmpc. <http://dx.doi.org/10.1594/WDCC/CMIP5.NCCNmpc> [Creator (Publication Year). Title. Publisher. Identifier]

Publication date
2011-10-10

Contact for data entity
[Trond Iversen](mailto:Trond.Iversen@met.no)

Link to external metadata
<http://cera-www.dkrz.de/WDCC/ui/Entry.jsp?acronym=NCCNmpc>

Summary
piControl is an experiment of the CMIP5 - Coupled Model Intercomparison Project Phase 5 (<http://cmip-pcmdi.llnl.gov/cmip5/>). CMIP5 is meant to provide a framework for coordinated climate change experiments for the next five years and thus includes simulations for assessment in the AR5 as well as others that extend beyond the AR5.
3.1 piControl (3.1 Pre-Industrial Control) - Version 1: Pre-Industrial coupled atmosphere/ocean control run. Imposes non-evolving pre-industrial conditions.
Experiment design: http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf
List of output variables: http://cmip-pcmdi.llnl.gov/cmip5/docs/standard_output.pdf
Output: time series per variable in model grid spatial resolution in netCDF format
Earth System model and the simulation information: CIM repository

Entry name/title of data are specified according to the Data Reference Syntax (http://cmip-pcmdi.llnl.gov/cmip5/docs/cmip5_data_reference_syntax.pdf) as activity/product/institute/model/experiment/frequency/modeling realm/MIP table/ensemble member/version number/variable name/CMOR filename.nc.

Quality
Accuracy: In addition to CMORs automated quality checks we performed manually checks on a data sample for each output variable. These comprise:
-visual inspection using ncview
-check whether min/max values are in physical range and consistent with variable units
-direction of flux variables and positive attribute
-consistency of data with vertical axis definition (i.e. whether data is flipped)
-land/ocean/sea ice masking
-rotation of vector data ;SQA - Scientific Quality Assurance 11/10/2011 10:28:03

Consistency: Quality Control Levels in CMIP5:
* Level 0: Spot checks on selected data
* Level 1: CMOR2 and ESG publisher conformance checks
* Level 2: Data consistency checks
* Level 3: Double- and cross-checks of data and metadata and data publication as DataCite DOI
QC Result Access: <http://cera-www.dkrz.de/WDCC/CMIP5/QCResult.jsp?experiment=cmip5/output1/NCC/NorESM1-M/piControl>

Completeness: [CMIP5:AuthorComment=We are still in the process of filling out the METAFOR CMIP5 questionnaire. Therefore its contents might be subject to change. Publications based on NorESM1-M are under preparation.]

Specification: [CMIP5:QualityLevel=3] [CMIP5:QualityFlag=approved by author] according to CMIP5 quality control rules <http://cmip5qc.wdc-climate.de> [CMIP5:DateOfQualityControl2=2011-10-10] [CMIP5:QualityControl2Comment=QC Level2 assigned at 2011-10-11 11:17:51 UTC: assignment confirmed by NCC (Ingo Bethke)] [CMIP5:CimMetadata=CMIP5 questionnaire data not found or not accessible] [CMIP5:TQA=done according to criteria for QCL3 defined at https://redmine.dkrz.de/collaboration/projects/cmip5-qc/wiki/Qc_13#Criteria-for-QC-L3DOI-publication] [CMIP5:SQA=checked by author] [CMIP5:DateOfQualityControl3=2011-12-01]

Link to primary data
http://cera-www.dkrz.de/WDCC/CMIP5/Tracking.jsp?tracking_id=tracking_id

Please note
WDC Climate as DOI publishing agency grants the validity of data accessed at WDC Climate. It recommends to check all downloaded data files by their tracking_id with the CMIP5 Data Validation Service at http://cera-www.dkrz.de/WDCC/CMIP5/Tracking.jsp?tracking_id=tracking_id The tracking_id can be found in the file headers (e.g. using ncdump -h).

This page is hosted at [WDCC](http://wdcc.dkrz.de), please send technical inquiries to data@dkrz.de.

[Impressum / Legal notice](#)

Data Citation Service (2) – pre DOI

QC Status Summary for CMIP5 Experiments

The listed CMIP5 [<http://cmip-pcmdi.lln.gov/cmip5/>] experiments are in the quality control process [<http://cmip5qc.wdc-climate.de/>] They appear in the list as soon as the QC Level 2 checks have started.

QC Results are available at: <http://cera-www.dkrz.de/WDCC/CMIP5/QCResult.jsp>

Status Explanation:
 QCL2_started: QC Level 2 checks are ongoing.
 QCL2_assigned: QC Level 2 is assigned to the experiment.
 QCL3_started: QC Level 3 checks are ongoing.
 QCL3_assigned: QC Level 3 is assigned to the experiment and the data is DOI published according to the DataCite regulations [<http://datacite.org/>]. This data, its basic metadata and its DOI are persistent.
 QCL1: The data of the experiment is withdrawn and waiting for a QC Level 2 check restart.

| Experiment | Status / (Comment, if available) | Last Change | Revision of QC L2 |
|--|----------------------------------|---------------------|-------------------|
| cmip5/output1/CNRM-CERFACS/CNRM-CM5/decadal1990 | QCL2_started | 2011-11-07T11:32:57 | 2 |
| cmip5/output1/CNRM-CERFACS/CNRM-CM5/decadal2005 | QCL2_started | 2011-12-06T01:38:27 | 1 |
| cmip5/output1/CNRM-CERFACS/CNRM-CM5/historical | QCL2_started | 2011-10-31T12:33:36 | 1 |
| cmip5/output1/CNRM-CERFACS/CNRM-CM5/piControl | QCL2_started | 2011-10-31T13:01:46 | 1 |
| cmip5/output1/CNRM-CERFACS/CNRM-CM5/rcp45 | QCL2_started | 2011-12-08T19:31:14 | 1 |
| cmip5/output1/CNRM-CERFACS/CNRM-CM5/rcp85 | QCL2_started | 2011-10-28T21:31:33 | 1 |
| cmip5/output1/INM/inmcm4/amip | QCL2_started | 2011-11-16T07:21:47 | 1 |
| QC Level set to QCL2_started (2011-11-16 07:21:47 UTC); New data received from INM; QC Level2 assigned at 2011-11-12 01:13:18 UTC; no exceptions | | | |
| cmip5/output1/IPSL/IPSL-CM5A-LR/amip4K | QCL2_started | 2011-10-27T15:06:02 | 2 |
| cmip5/output1/IPSL/IPSL-CM5A-LR/aquaControl | QCL2_started | 2011-10-27T11:51:38 | 4 |
| cmip5/output1/IPSL/IPSL-CM5A-LR/aqua4K | QCL2_started | 2011-10-26T16:09:09 | 3 |
| QC restart (2011-08-19 22:32:43 UTC); QC Level set to QCL2_started (2011-08-18 12:18:42 UTC); Additional data to be checked: mon; QC Level2 assigned at 2011-07-19 13:51:31 UTC; Info and Warning messages raised by QC; considered acceptable by data providers | | | |
| cmip5/output1/IPSL/IPSL-CM5A-LR/historical | QCL2_started | 2011-11-01T18:39:45 | 1 |
| cmip5/output1/IPSL/IPSL-CM5A-LR/piControl | QCL2_started | 2011-10-26T21:33:50 | 1 |
| cmip5/output1/IPSL/IPSL-CM5A-LR/rcp45 | QCL2_started | 2011-10-26T21:14:09 | 1 |

Find Data Citation (3)

1. Find citation information for data file:

Currently, DOI and quality information are not displayed in the ESG gateways !

<http://cera-www.dkrz.de/WDCC/CMIP5/Citation.jsp>

2a. QC L3 data with DOI:

Find the final data citation information and the DOI

2b. Data without DOI:

Find the preliminary citation information or citation construction rule:
author:<modeling center>; title:<DRS name>, served by ESGF;
data identifier:<download data node or gateway destination>

E.g.: MPI-M (2011): cmip5 output1 MPI-M MPI-ESM-LR amip, served by ESGF.

Gateway "DKRZ" (24.02.2012): <http://ipcc-ar5.dkrz.de>

3. Recheck for assigned DOIs after finishing your paper (cf. 1.)

Questions / Comments ?

More Information:

CMIP5 QC: <http://cmip5qc.wdc-climate.de>

Citation Service: <http://cera-www.dkrz.de/WDCC/CMIP5/citation.jsp>

All QC Services: <http://cera-www.dkrz.de/WDCC/CMIP5>

WDCC Data Publication System: <http://cera-www.dkrz.de/atarrabi2>

CMIP5 Project: <http://cmip-pcmdi.llnl.gov/cmip5/>

stockhause@dkrz.de / data@dkrz.de