



Quality Assessment Concept for CMIP5

Martina Stockhause, Michael Lautenschlager



CMIP5 / IPCC-AR5 in Numbers

Coupled Model Intercomparison Project (CMIP5)

- **Participants:**

- ca. 20 participating modelling centres
with ca. 40 model configurations (different resolutions)

- **Experiments:**

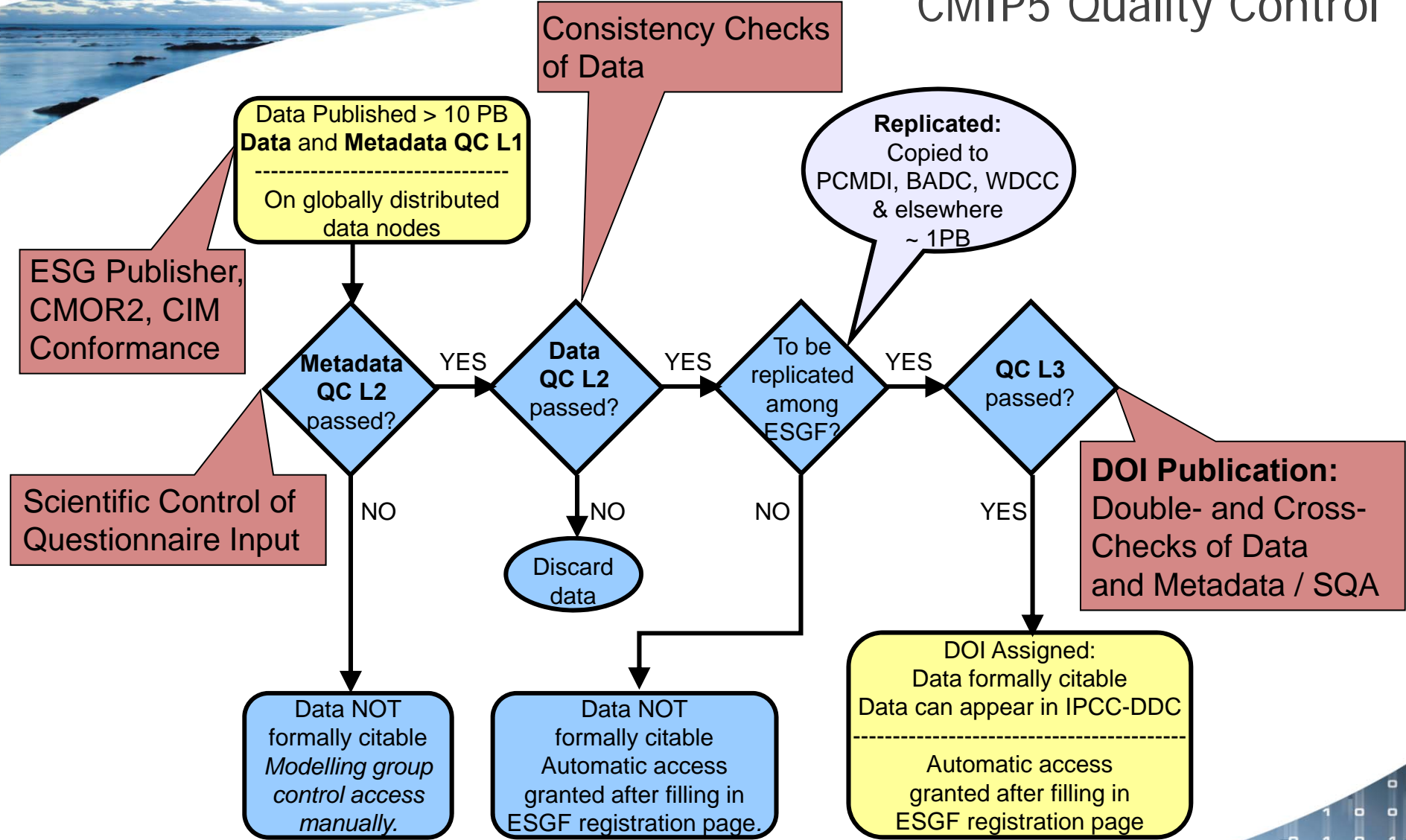
- 60 experiments with partly several realisations (ensemble members)
over 90 000 model years

- **ca. 2 Mio. atomic datasets of ca. 400 experiments**

- **Data Volume:**

- ca. 10 PB output: → **QC L1**
 - thereof ca. 2.5 PB requested *'output1'*, *'output2'* → **QC L2**
 - thereof ca. 1 PB replicated *'output1'*: IPCC-AR5 → **QC L3 / DOI**

CMIP5 Quality Control



(Informal citation still requested where formal citation not available)

Quality Control Level 1

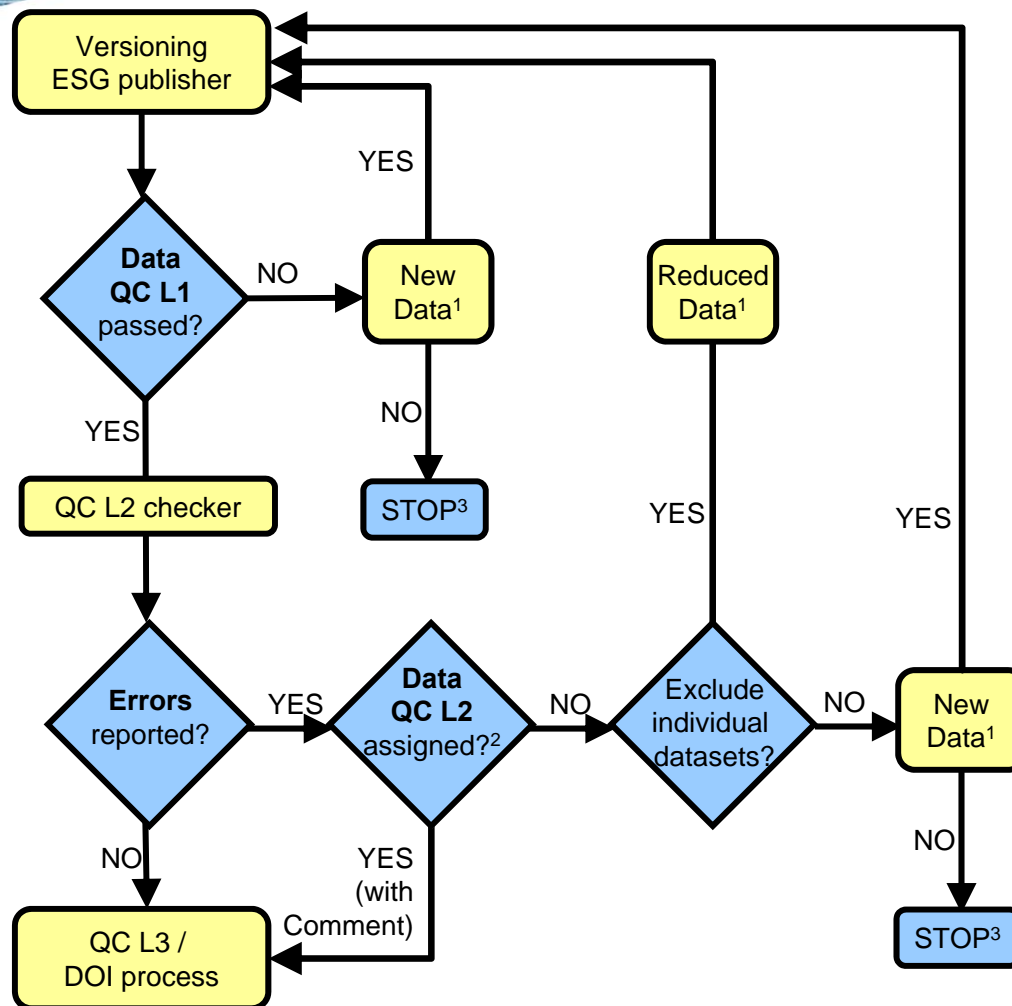
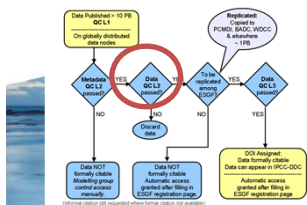


QC Level 1 (CMOR2 and ESG publisher conformance checks): Performed at all ESGF partners during ESG publication

- **Data checks:**
 1. cmor2 compliance checks by the cmor checker `check_CMOR_compliant.py`
 2. esg publisher conformance

- **Metadata checks:** Completeness and technical validation of questionnaire input

Assignment of Quality Control Level 2



- 1 New version assigned to ESG datasets
- 2 Assignment in co-operation with data author according to criteria <http://www.leuchtturm-atlas.de/SCR/qc2list.html>
- 3 Delete possible QC L2 results out of QCDB

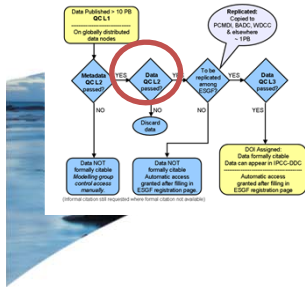
Assignment Criteria

Home Projects Help

CMIP5 » QC

QC2 Exceptions

<http://www.leuchtturm-atlas.de/SCR/qc2list.html>



QC2 Exception Codes (preliminary version)

Fatals: immediate action necessary,
 errors: data unacceptable,
 warnings: data possibly not ok,
 informatory: fine - just for info,
 unclear: title lines or open issues.

Thematic exception groups (see first column):
ACCess errors, **GEN**eral checks, **MET**adata and file name checks,
TABL: inconsistencies in comparison to meta data tables,
TIMEaxis checks, **VARI**ables' checks.
OBSOlete messages (not used for CMIP5 project).

The following flags F<n> refer to general checks,
 mainly on the time axis.

key group	description	comment
F-1 GENE	-- Not checked	
F0 GENE	-- No error found	
F1 TIME	testTimeStep() ^0, ^5 Error: negative time step	
F2 TIME	testTimeStep() ^0, ^5 Error: missing time step	This, of course, is no error if the QC is run over several time slices with intentional gaps in between. You may want to set NON_REGULAR_TIME_STEP in the setup file to check only for positive increments (of perhaps different sizes).
F4 TIME	testTimeStep() ^0, ^5 Error: identical time step	
F8 TIME	testCalendarTimeBounds() Error: negative/zero time bounds range	
F16 TIME	testCalendarTimeBounds() ^0 Error: overlapping time bounds ranges	
F32 TIME	testCalendarTimeBounds() ^0 Warning: gap between time bounds ranges	

DOI for Scientific Data
10.1595/WDCC/TEST_AMP_TR

Title
cmip5 output MPI-M ECHAM6-M

Citation
Lautenschlager, Michael (2011)

Publication Date
2011-02-17

Contact for data entity
[Joerg Wegner](#)

CMIP5 Metadata host:
<http://cera-www.dkrz.de/WDCC/>

Summary
amip is an experiment of the CMIP5 experiments for the next five years.
3.3 amip (3.3 AMIP): AMIP (1976-2010)

Experiment design is described in pcmdi.llnl.gov/cmip5/docs/stan/. The output is stored in netCDF format in the public repository.

Quality
Accuracy: not filled

Consistency: Quality Control Level 1
* Level 0: Spot checks on selected data
* Level 1: CMOR2 and ESG published
* Level 2: Technical checks on time series
* Level 3: Data approved by author

Completeness: not filled

Specification: [CMIP5:QualityLevel1]
[CMIP5:QualityControl2]Common name: T=6hrPlev, var=va, dim=time
metadata not found or not accessible

Link to primary data
[CMIP5Links.jsp?acronym=MXETam](#) (to be replaced by link into PCMDI gateway)

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/CMIP5Tracking.jsp?tracking_id=tracking_id. The tracking_id can be found in the file headers (e.g. using ncdump -h).

Data for CMIP5 experiment MXETam

Please choose your desired data destination. The displayed list of links refer to ESG datasets, i.e. collections of data belonging to a model realm / ensemble member (see DRS syntax definition [http://cmip-pcmdi.llnl.gov/cmip5/docs/cmip5_data_reference_syntax.pdf]). Links resolve to the corresponding Thredds Data Server entries.

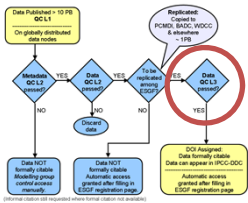
Location

Dataset
cmip5_output MPI-M ECHAM6-MPIOM-TR amip mon atmos Amon r2i1p1 v20100928
cmip5_output MPI-M ECHAM6-MPIOM-TR amip fx atmos fx r0i0p0 v20100928
cmip5_output MPI-M ECHAM6-MPIOM-TR amip mon atmos Amon r1i1p1 v20100928
cmip5_output MPI-M ECHAM6-MPIOM-TR amip mon landlce L1mon r1i1p1 v20100928
cmip5_output MPI-M ECHAM6-MPIOM-TR amip mon land Lmon r1i1p1 v20100928
cmip5_output MPI-M ECHAM6-MPIOM-TR amip 6hr atmos 6hrPlev r2i1p1 v20100928
cmip5_output MPI-M ECHAM6-MPIOM-TR amip day atmos day r1i1p1 v20100928
cmip5_output MPI-M ECHAM6-MPIOM-TR amip day land day r1i1p1 v20100928
cmip5_output MPI-M ECHAM6-MPIOM-TR amip day atmos day r2i1p1 v20100928
cmip5_output MPI-M ECHAM6-MPIOM-TR amip mon land Lmon r2i1p1 v20100928
cmip5_output MPI-M ECHAM6-MPIOM-TR amip 6hr atmos 6hrPlev r1i1p1 v20100928
cmip5_output MPI-M ECHAM6-MPIOM-TR amip fx land fx r0i0p0 v20100928
cmip5_output MPI-M ECHAM6-MPIOM-TR amip mon landlce L1mon r2i1p1 v20100928

Back to [the metadata page](#).

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

GUI for Scientific Quality Assurance



Process Overview

General

- Authors
- DOI Contact
- Contributors
- Relations
- Coverage
- Quality
- Summary

Atarrabi:

- GUI for final author check of selected metadata (general, contact, authors, contributors, citations,...)
- Scientific quality information should be added
- Publication of environ. data:

umwelt.wikidora.com

cera-www.dkrz.de/atarrabi/

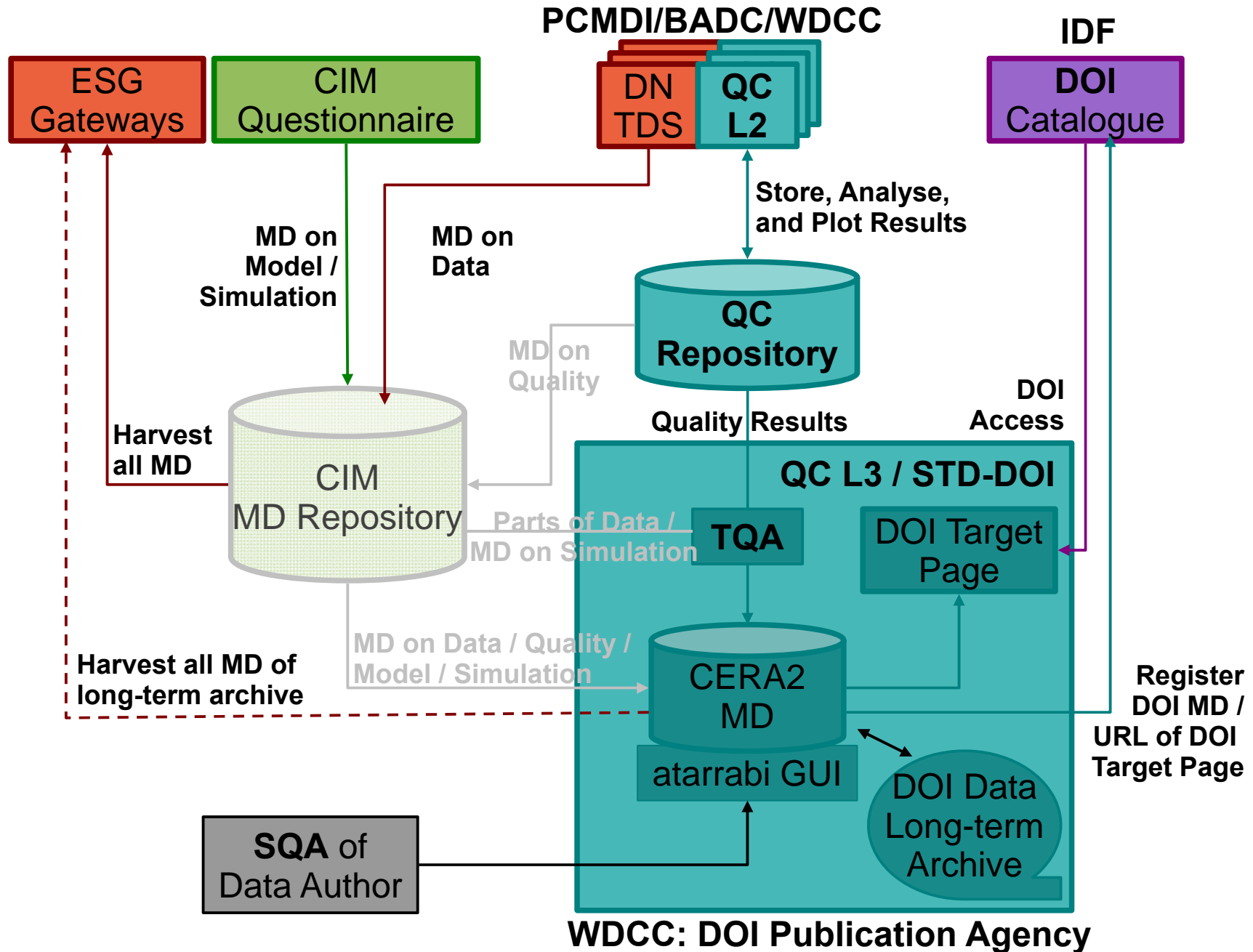


General information for MXETam

These general data information are shown on the [DOI/URN](#) target page and library catalogues like [GetInfo](#). The core metadata properties are chosen for accurate and consistent data identification in citations and for data retrieval, along with recommended use instructions. They belong to the metadata kernel of the DOI (STD-DOI).

Entry acronym	<input type="text" value="MXETam"/>
Entry name	<input type="text" value="cmip5 output MPI-M ECHAM6-MPIOM-TR amip"/>
Entry title *	<input type="text" value="cmip5 output MPI-M ECHAM6-MPIOM-TR amip"/>
Entry description summary	<p>amip 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. 3.3 amip (3.3 AMP): AMP (1979 - at least 2008). Impose SSTs and sea ice from observations but with other conditions as in experiment 3.2 historical. Experiment design is described in detail in http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf and the list of output variables and their temporal resolutions are given in http://cmip-pcmdi.llnl.gov/cmip5/docs/standard_output.pdf. The output is stored in netCDF format as time series per variable in model grid spatial resolution. For more information on the Earth System model and the simulation please refer to the CIM repository.</p>
Additional entry description	<input type="text"/>
Creation date	<input type="text" value="11/08/2010"/>
Language	<input type="text" value="English"/>
Project name	<input type="text" value="IPCC/CMIP5_test"/>
Project description summary	<input type="text" value="this is for testing purposes ony"/>
DOI/URN target page with actual citation information	<input type="text" value="http://cera-www.dkrz.de/WDCC/aii/CMIP5Compact.jsp?acronym=MXETam"/>
Rules for citation *	<p>At least one author is required as contact independent of your choice:</p> <p><input checked="" type="radio"/> Cite by persons: [author(s)][(PublicationDate)];[Title].[Publisher].[doi:DOI].[http://dx.doi.org/DOI]</p> <p><input type="radio"/> Cite by institutes: [Contributor(s)][(PublicationDate)];[Title].[Publisher].[doi:DOI].[http://dx.doi.org/DOI]</p>
CMIP5 questionnaire *	<input checked="" type="checkbox"/> I approve that the metadata I filled in the CMIP5 questionnaire and sent to CMIP5 is correct and up-to-date.

CMIP5 Distributed Quality Control Status



Data Publication and Quality Control Procedure for CMIP5 / IPCC-AR5 Data

Martina Stockhause, Michael Lautenschlager, Heinke Hoeck, and Frank Toussaint

EGU2011-2859

CMIP5 Organization & Infrastructure Components

For distribution of data connected to the next IPCC report, the Earth System Grid Federation (ESGF) was founded. Its members have different responsibilities within the data infrastructure:

- PCMDI / LLNL: data and security infrastructure (ESG)
- BADC (British Atmospheric Data Centre): metadata infrastructure (METAFOR / CIM)
- WDCC (World Data Center Climate) / DKRZ: quality control, data publication (DataCite DOI)

CMIP5 Quality Control (QC)

For CMIP5 ca. 3 PB of officially requested data are expected to be archived. About 1 PB of that data will likely be of especially high interest and will be replicated by the three ESGF partners. Because of the high data volume the QC checks up to level 2 are performed distributed among the ESGF. The final QC level 3 checks for the DOI assignment are carried out by WDCC. Afterwards CMIP5 data is formally citable and remains persistent.

Future Perspective

The current DOI publication procedure is comparable to the publication of grey literature in scientific print media. For the integration of a peer review process quality procedures accepted and agreed on by the earth system modeling community are necessary. The distributed quality control approach could be reduced in complexity by the integration of the QC Repository into the CIM Metadata Repository.

CMIP5 Quality Control Workflow

Overall CMIP5 QC Workflow

Three Quality Control (QC) Levels are defined for CMIP5 data:

- **QC Level 1:** Metadata: Technical checks on METAFOR questionnaire input data
Data: CMOR2 and ESGF publisher performance checks
- **QC Level 2:** Metadata: METAFOR questionnaire metadata checked by scientist
Data: Technical checks e.g. on the reliability of variable ranges and the consistency checks between data and data requirements
- **QC Level 3 / DOI:** Data approved by author and published as DOI
QC checks for data and metadata are performed, separately, for levels 1 and 2. During the cross-checks of QC L3 checks their results are reviewed.
Data assigned a DOI is formally citable and is granted persistent access.

Quality Control Workflow CMIP5

The different QC levels are connected with different access rights for registered users:

- **Restricted Access (QC Level 1 Data):** After ESGF publication the access is restricted under control of the modelling centre.
- **Scientific Access (QC Level 2):** The scientific community is granted access of data of QC L2.
- **QC Level 3 / DOI:** With the DOI assignment the data archive is opened for access by non-scientific users.

Granularity of Quality Control

QC is accomplished on DRS Atomic Dataset level. The QC results are aggregated on DRS experiment level.
In the gateways data discovery is supported down to the level of Ensemble versions (ESG dataset).

Granularity of QC in CMIP5 context

CMIP5 Process: DRS Name / Hierarchy Level

DOI: QC

More Information: <http://purl.org/org/cmip5/qc>

Distributed Quality Control Approach

Workflow of Distributed QC in CMIP5

ESGF: Earth System Grid, MC: Metadata, DR: Data Node, TDS: Thredds Data Server, TQA: Technical Quality Assurance, SQA: Scientific Quality Assurance

Distributed QC Procedure in CMIP5

CMIP5 data is delivered to one of the three ESGF partners, where it is ESGF published and thus QC L1 Data checked. Afterwards QC L2 Data consistency checks are performed, before a data subset is replicated among the ESGF. QC L2 results are stored in a central QC Repository.

During QC L3 / DOI checks the QC results are accessed by the DOI Publication Agency WDCC. Other sources for cross- and double-checks are the CIM Metadata Repository, the Thredds Data Server (TDS), and the metadata stored in the long-term archive at WDCC.

Thus, the effort of the QC L2 Data checks is shared among the ESGF. But the QC L3 / DOI checks are performed at one site making use of the QC L2 results stored in a central QC Repository.

Our distributed QC approach consists of different software components:

- Multiple sites performing QC checks
- Central Repository
- DOI Publication Agency

Organisation of CMIP5 Data Federation

More Information: <http://purl.org/org/cmip5/qc>

Data Publication Procedure

Actors in DOI Publication Process

	Permission: QC L2	Scientific Q. Assurance	Technical Q. Assurance	DOI: Publication
Scientific				
Publication Agency				
Registration Agency				

DOI Construction Rule for CMIP5:
doi: 10.1594/WDCC/CMIP5_<opaque bit>

DOI Publication GUI startarrabi

QC L3 / DOI Process in CMIP5

The cross- and double checks of the Technical Quality Assurance make use of the QC result of the preceding levels. Data as well as metadata is reviewed and data accessibility checked.

The final author approval step is supported by the GUI startarrabi. Authors check basic metadata and add information about their own quality assurance (Scientific QA). A DOI is assigned and registered at the International DOI Foundation (IDF: dx.doi.org) via the Registration Agency DataCite.

More Information: <http://cera-www.dkrz.de/startarrabi>



martina.stockhause@zmvw.de
WDC Climate / DKRZ: www.wdc-climate.de; www.dkrz.de
CMIP5: cmip-pcmdi.llnl.gov/cmip5; **CMIP5 Quality Control:** purl.org/org/cmip5/qc



<http://cmip-pcmdi.llnl.gov/cmip5/>
<http://cmip5qc.wdc-climate.de>



10.05.2011
Stockhause and Lautenschlager, GO-ESSP 2011