

Klimamodell-Daten: Anforderungen und Nutzung

Von der Vergangenheit in die Zukunft: Neue
Klimasimulationen für Wissenschaft und Gesellschaft

Hamburg, 23./24. Februar 2012

Michael Lautenschlager (DKRZ)



Nutzungsanforderungen CMIP5

- Ergebnisse aus CMIP5 (Coupled Model Intercomparison Project No. 5) für
 - Modellvergleiche zur Verbesserung der Klimamodelle und Vertiefung des Verständnisses des Klimasystems
 - Datenbasis für wissenschaftliche Arbeiten als Grundlage für den IPCC-AR5 (IPCC Assessment Report No. 5)
- **Neu in IPCC-AR5:** alle drei Arbeitsgruppen sollen mit derselben Klimamodelldatenbasis arbeiten
- **Interdisziplinäre Anwendungen** (*IAV – Impact, Adaptation/Mitigation, Vulnerability*) stellen hohe Anforderungen an Qualität und Dokumentation der Daten
- Rückwirkungen auf Datenaufbereitung und Bereitstellung der Klimamodelldaten im IPCC DDC (IPCC Data Distribution Centre) im Vergleich zu AR4
- Durchführung von **Qualitätskontrolle und Datendokumentation** unmittelbar im Anschluss an die Modellrechnungen, um möglichst viele Fehler und Inkonsistenzen vor der (interdisziplinären) Nutzung zu beheben.

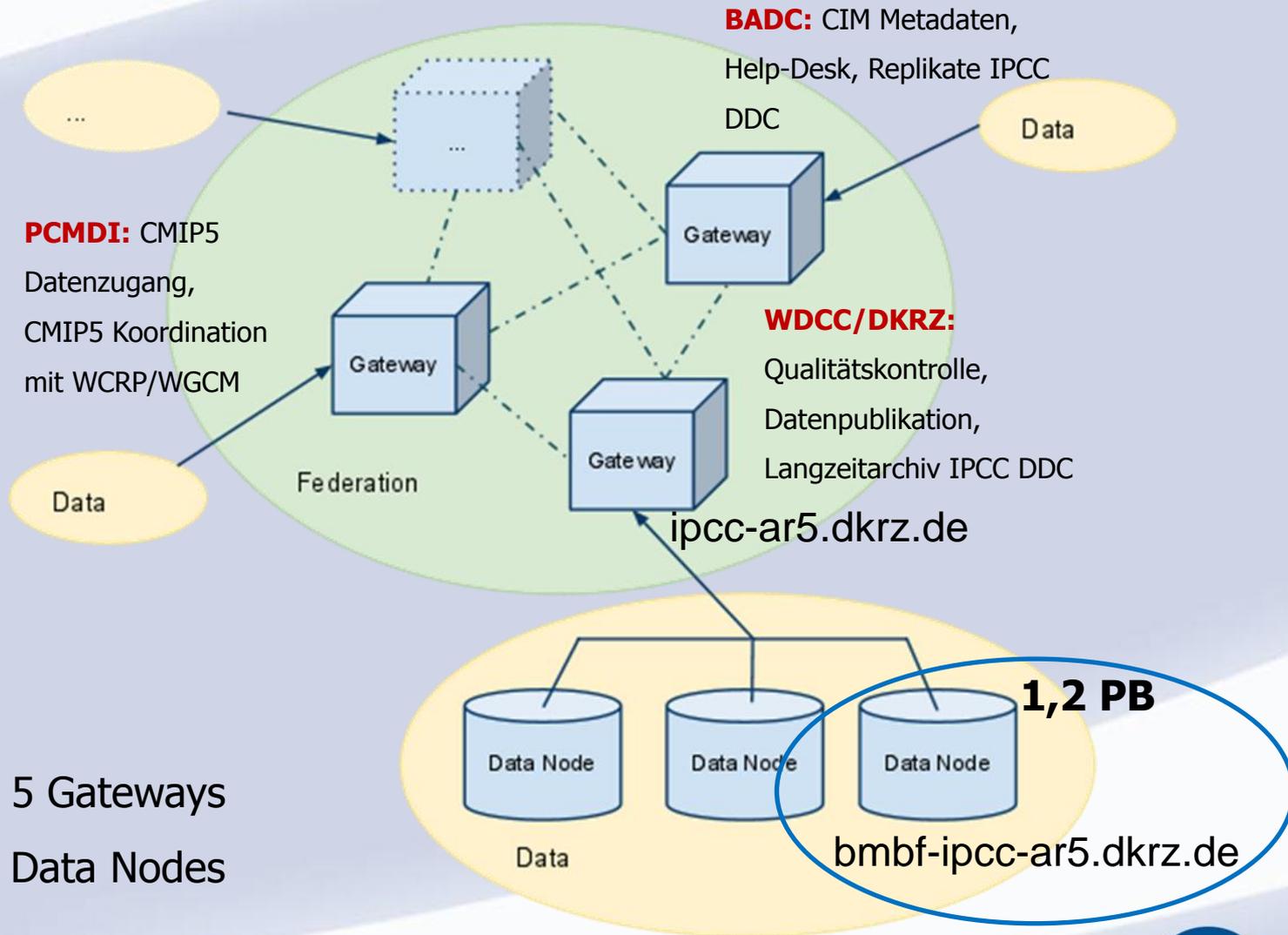
Abstimmung CMIP5 / IPCC DDC

- Enge Abstimmung in der Planung von CMIP5 und IPCC DDC für Datenbereitstellung im AR5 Prozess in den Sitzungen des TGICA (Task Group on Data and Scenario Support for Impact and Climate Analysis):
 - TGICA-15 (Genf, 2008): Diskussion der Variablenliste im DDC
 - TGICA-16 (Boulder, 2010): Verabschiedung der DDC Variablenliste und Anerkennung des CMIP5 Qualitätssicherungsprozesses als Kriterium für Integration der finalen CMIP5-Daten ins IPCC DDC
 - TGICA-17 (Stanford, 2012): Abstimmung der Zeitschiene für den Übergang der finalen CMIP5-Daten ins IPCC DDC (IPCC-AR5 Referenzdatenarchiv)
- Aktueller Status CMIP5: Rechnungen weitestgehend durchgeführt, Prozess der Qualitätskontrolle und Datentransfer ins IPCC DDC in Bearbeitung

Datenmengen

- CMIP3 / IPCC-AR4 (Report 2007)
 - Beteiligung: 17 Zentren mit 25 Modellen
 - Insgesamt 36 TB zentral am PCMDI und ca. ½ TB im IPCC DDC als Referenzdaten
- CMIP5 / IPCC-AR5 (Report 2013/2014)
 - Beteiligung: 25 Zentren mit 40 Modellen
 - Produziertes Datenvolumen: ca. 10 PB davon **MPI-ESM 640 TB**
 - Angefragtes Datenvolumen: ca. 2 PB (in CMIP5 Datenföderation)
 - Datenvolumen für IPCC DDC: ca. 1 PB (voller Qualitätssicherungsprozess) davon **MPI-ESM 60 TB**
- Status CMIP5 Datenarchiv (21.02.12):
 - 940 TB für 38000 Datensätze gespeichert in 2,1 Mio Files auf 16 Datenknoten

CMIP5 Datenföderation

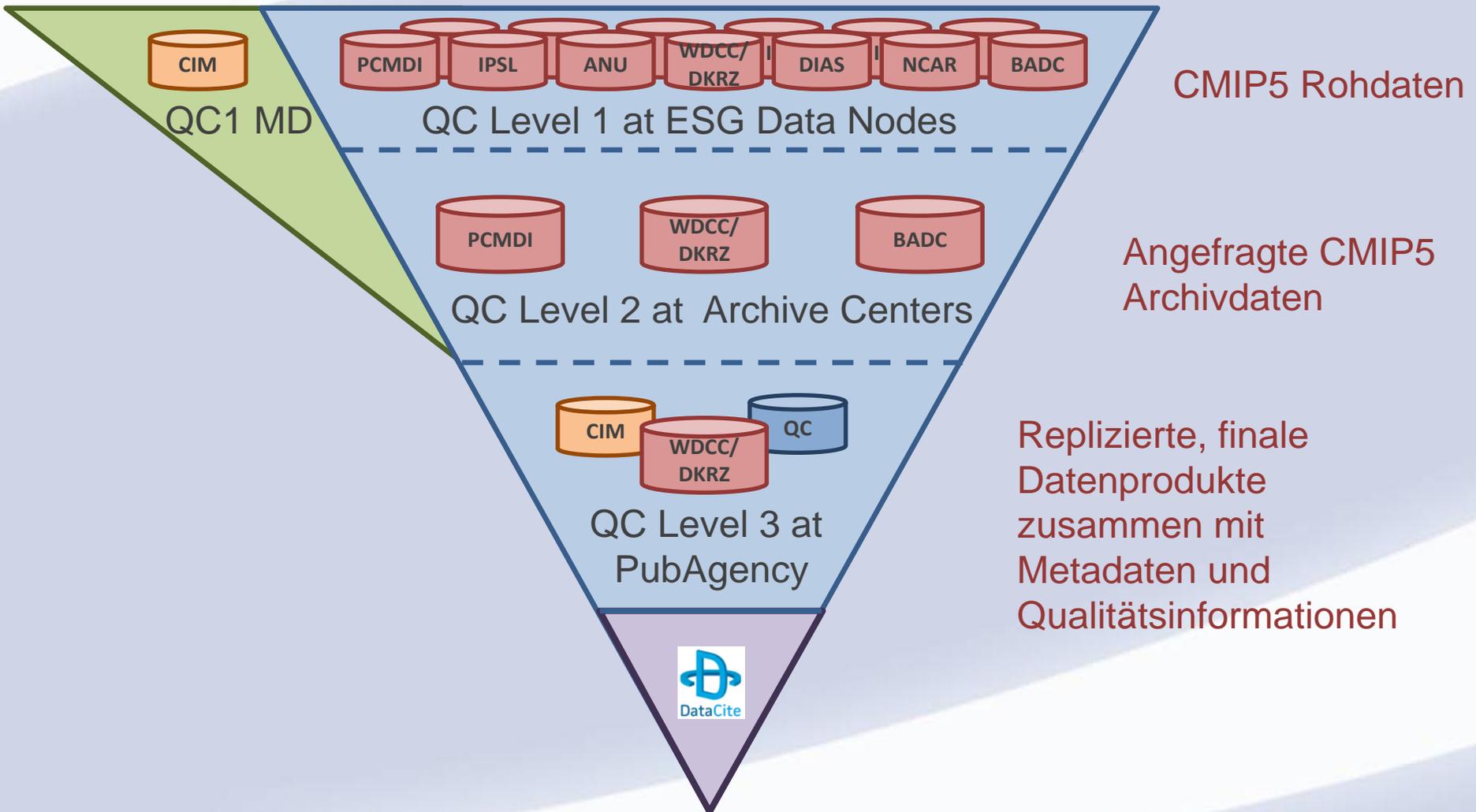


Derzeit 5 Gateways
und 16 Data Nodes

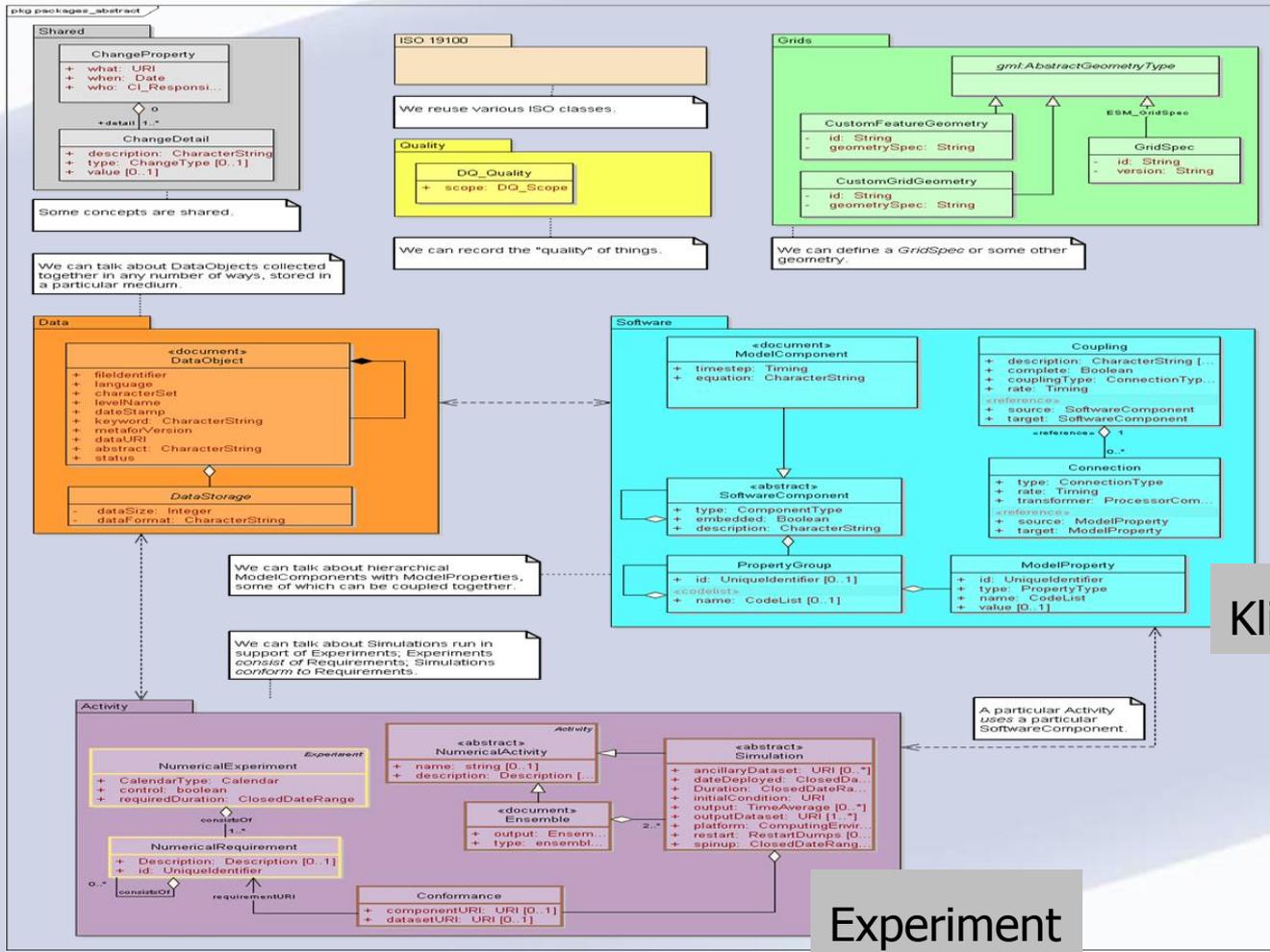
CMIP5 Datenföderation

- 3 Zentrale Managementkomponenten zur Nachnutzung der Daten
 - Hoch strukturierte Datenfiles im selbstbeschreibenden Datenformat NetCDF/CF inkl. Use-Metadaten
 - **Neu:** Durchsuchbare Modell- und Experimentbeschreibungen (CIM Metadaten aus EU-Projekt METAFOR)
 - **Neu:** 3 stufiges Qualitätssicherungskonzept für Daten und Metadaten, das abschließt mit Langzeitarchivierung und Zuweisung einer Zitiervorschrift zum direkten Referenzieren der Klimamodelldaten in wissenschaftlichen Artikeln

Qualitätssicherungskonzept



CIM Metadaten

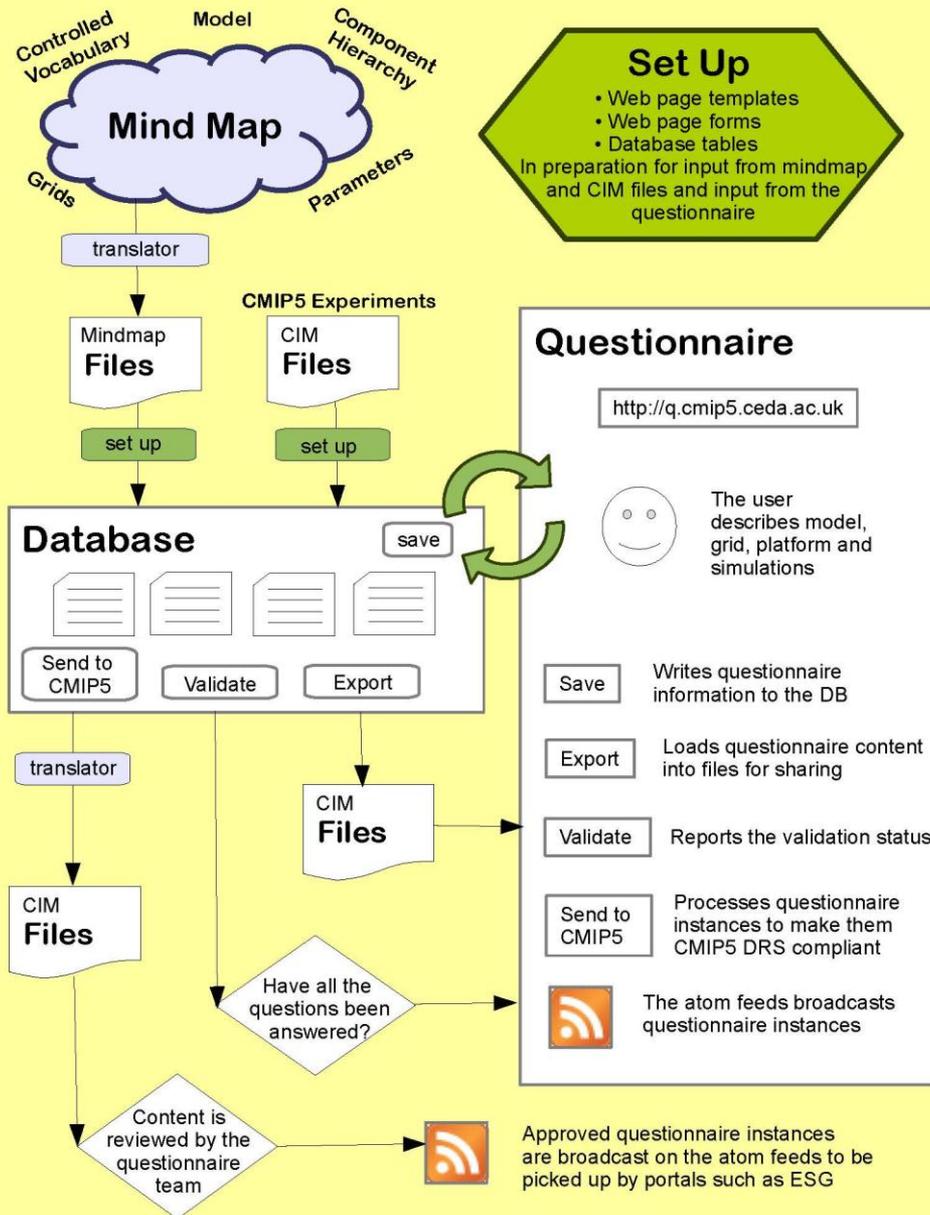


- Entwicklung: FP7 METAFOR
- **Neu:** Dokumentation der Datenerzeugung in enger Verknüpfung mit den Klimamodelldaten
- **Verbesserung:** durchsuchbare Modell- und Experimentbeschreibung

Klimamodell

Experiment

CMIP5 Questionnaire Metadata Pipeline



CIM Qualitätskontrolle

- CIM QC-1 und QC-2 erfolgt parallel zur Qualitätsprüfung der Klimamodelldaten
- Interaktion mit dem Datenproduzenten erforderlich

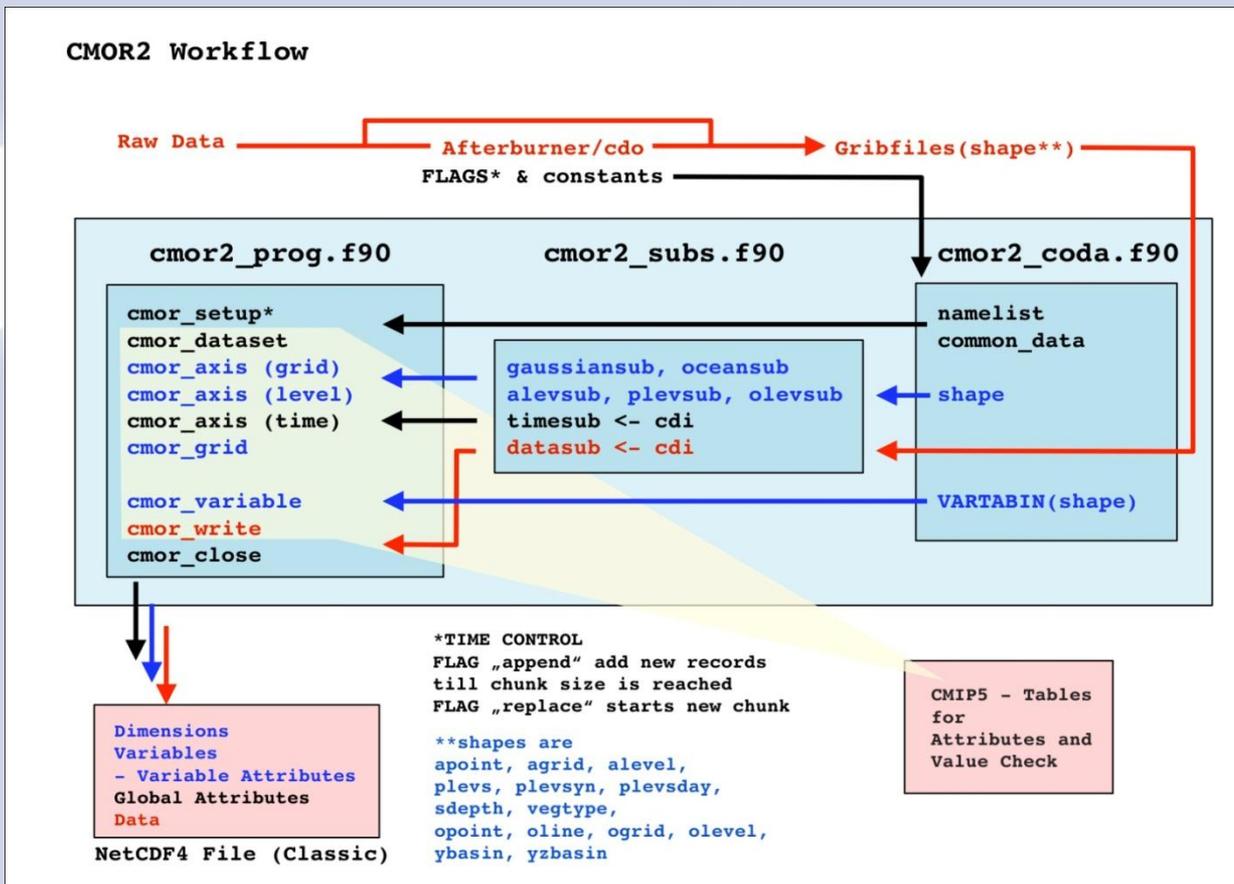


23.02.12



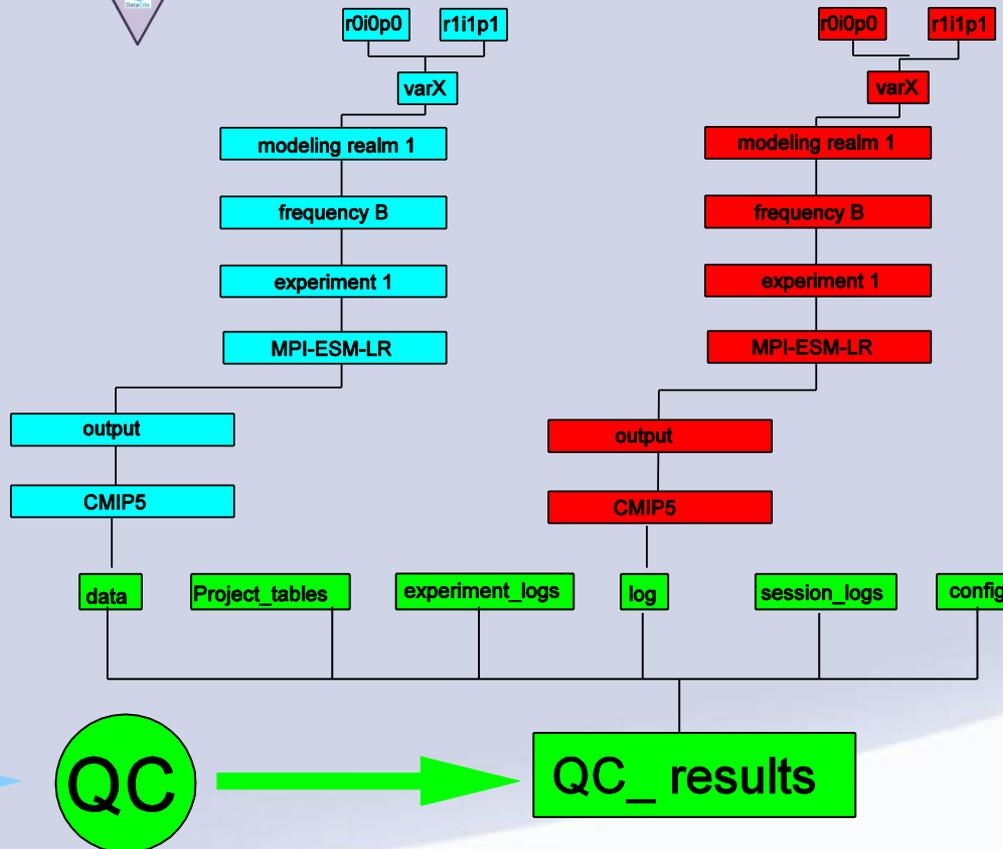
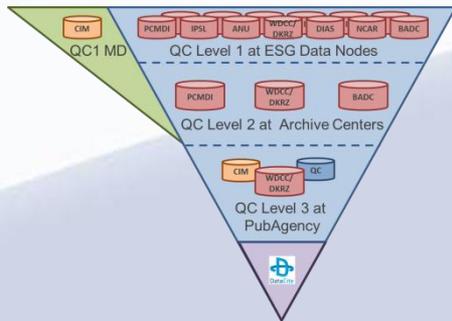
QC-L1: CMOR-2

Eingangsvoraussetzung
für Integration der
CMIP5 Daten in die
Datenföderation



- Einheitliche Datenstruktur
- Standard-Datenformat: NetCDF/CF
- Namenskonventionen in CF-Metadaten für Modell, Zentrum, Experiment, Variable (festgelegt in Werteliste (CVs))
- **Vorteil:** Alle CMIP5 Daten lassen sich direkt mit denselben Werkzeugen bearbeiten.

QC-L2: Automatisierte Kontrolle der CMIP5 Daten



- Inhalt der Datenfiles wird überprüft:
 - Zeitliche Kontinuität
 - Plausibilität der Feldwerte
 - Konventionen für File-Namen (z.B. Zeitstempel)
 - CF Metadaten gegen CMIP5 Vorgaben

QC-L2 Meldungen

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.

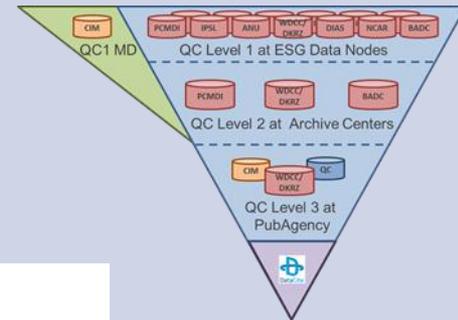
Thematic exception groups (see first column):
ACCESs errors, GENERAL checks, METAdata and file name checks,
TABL: inconsistencies in comparison to meta data tables,
TIMEaxis checks, VARIABLEs' checks.
OBSOlete messages (not used for CMIP5 project).

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 gap in between. You may want to set NON_REGULAR_TIME_STEP in the setup file 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	This warning refers to the bounds CF defin for every time step within a file. Righthound(n) should be = Lefthound(n+1)

- Klassifizierung der Meldungen
 - Fehler: Daten werden nicht akzeptiert
 - Warnungen: Rücksprache mit Datenproduzent, Korrektur oder Vermerk im Prüfprotokoll
 - Informationen: Vermerk im Prüfprotokoll

QC-L3: Abschluss der Qualitätssicherung



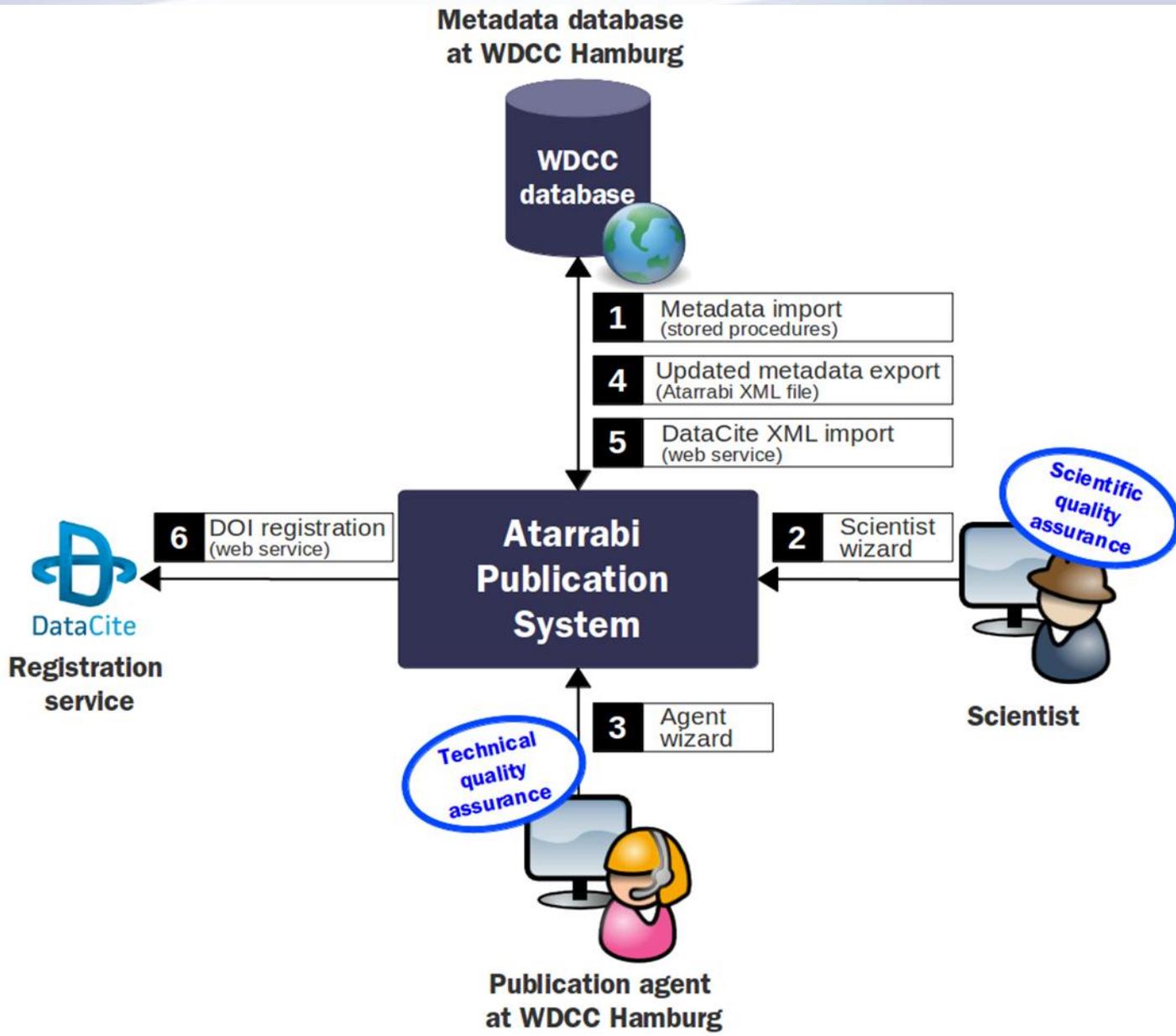
QC-L2		vorgestellte Ergebnisse		
		Workflow Processes		
	Permission	SQA	TQA	Publication
Scientist				
WDCC				
TIB				

Scientific Quality Assurance – SQA

Technical Quality Assurance – TQA

TIME

DataCite Datenpublikation



Entwicklung:



Förderung:



Abschluss der Qualitätsprüfung

- Nach Endkontrolle von Daten und Metadaten (CIM und CF)
 - Qualitätsstatus: „**approved by author**“
 - Daten werden als unveränderbar (**irrevocable**) gekennzeichnet.
 - Langzeitarchivierung im **WDC Climate** des DKRZ
- Durchführung der DataCite Datenveröffentlichung und Integration in Bibliothekskataloge
 - Datenentität (hier ein Modellexperiment) erhält eine **Zitiervorschrift** zur direkten Verwendung in wissenschaftlichen Publikationen und für den transparenten Datenzugriff über **DOI** (Digital Object Identifier)
 - Zitiervorschrift enthält Datenautor und Titel sowie WDC Climate als Herausgeber und den DOI
 - Auflösung des DOI führt auf „**Landing Page**“, deren Adresse in der zentralen Datenbank des DOI Handle Servers gespeichert ist.

DOI Landing Page

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; Tziputra, 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](#)

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 nview
-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_L3#Criteria-for-QC-L3DOI-publication] [CMIP5:SQA=checked by author] [CMIP5:DateOfQualityControl3=2011-12-01]

Link to primary data
<Links.jsp?acronym=NCCNMpc>

Zitiervorschrift

Ansprechpartner

Metadaten

Zusammenfassung

Informationen

Zu Qualität

Direkter Zugriff

Auf Klimadaten

DataCite DOI Schnittstelle im WDCC

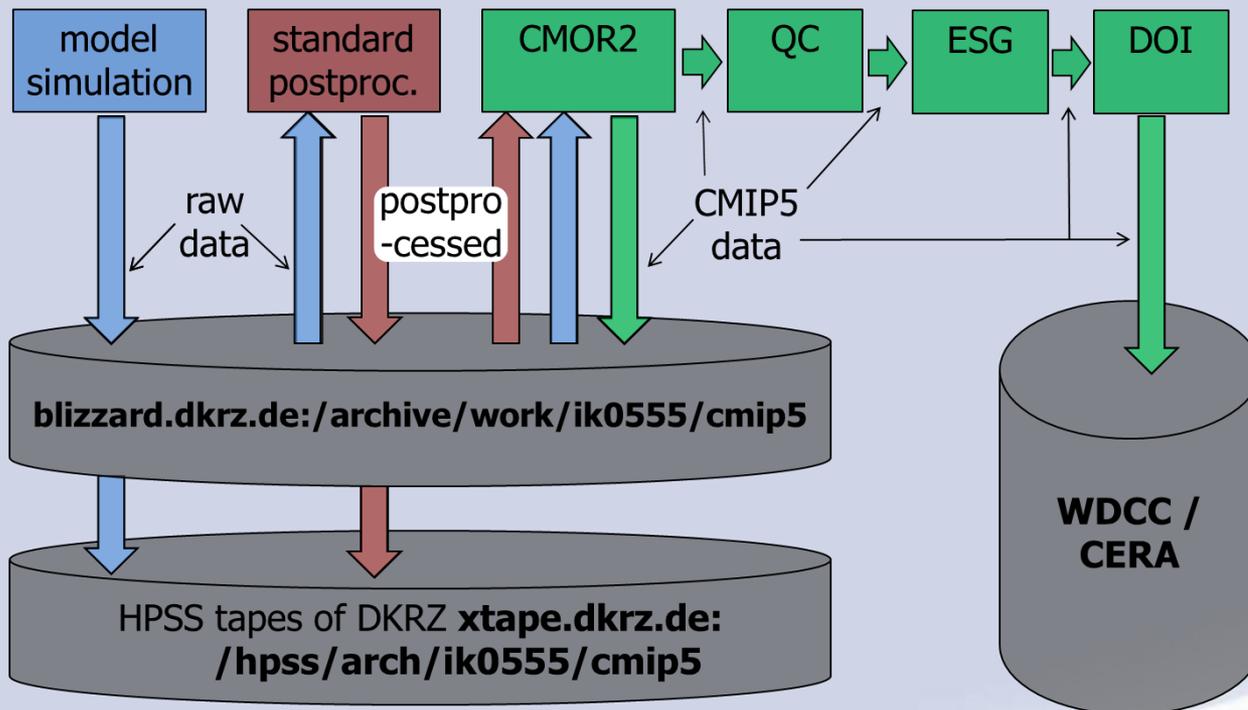
- Zeige mir alle Datenentitäten im WDCC, die einen DataCite DOI besitzen
 - URL: <http://cera-www.dkrz.de/WDCC/ui/FindDoiPublications.jsp>
- Zeige mir für einen CMIP5 NetCDF/CF Datensatz, ob er zu einer publizierten Datenentität gehört
 - URL: <http://cera-www.dkrz.de/WDCC/CMIP5/Citation.jsp>

Zusammenfassung

Neue Aspekte im CMIP5 Datenmanagement:

- Qualitätsgeprüfte, dokumentierte und unveränderliche Datenprodukte werden für allgemeine Nutzung zur Verfügung gestellt.
- DataCite Datenveröffentlichung fördert Anerkennung der Datenautoren (Zitiervorschrift) und transparenten Datenzugriff (DOI).
- Es wurde eine internationale Föderation verteilter Datenarchive ins Leben gerufen, um auch zukünftig die große Datenvolumina der Klimaforschung organisieren zu können.

CMIP5 Daten MPI-ESM



- Ausblick auf WS für Datennutzer (Fr. 24.03.):

Wo sind die MPI-ESM Daten in der DKRZ Infrastruktur gespeichert?

Wie wird die Klimadatenpublikation durchgeführt?

Welche Zugriffsmöglichkeiten stehen zur Verfügung für MPI-ESM und CMIP5?

Wie kann man die MPI-ESM Daten bearbeiten?