



DKRZ

DEUTSCHES
KLIMARECHENZENTRUM

Dateninfrastrukturen der Klimaforschung

DINI Jahrestagung
Dortmund 21. / 22.09.2011

Michael Lautenschlager

- Einige Worte zum DKRZ
- Einführung in Klimamodellierung
- Dateninfrastrukturen am Beispiel IPCC-AR5/CMIP5
 - Fokus: Organisation
- Herausforderungen

Unser Auftrag

DKRZ (Deutsches Klimarechenzentrum gGmbH) in einem Satz:

- DKRZ - höchste Rechenleistung, ausgereiftes Management größter Datenmengen und kompetenter Service garantieren erstklassige Klimaforschung.

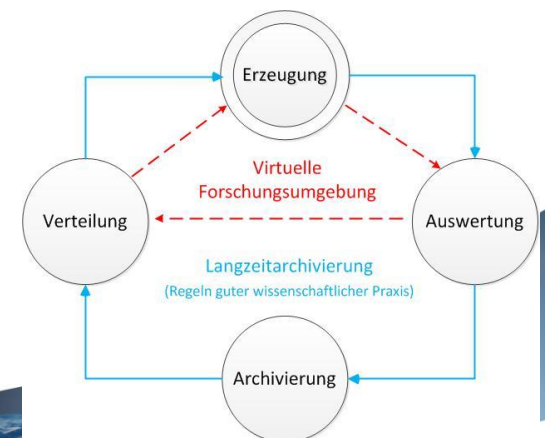
(Thematisch orientiertes Rechen-und Datenzentrum)

Unsere Stärken

DKRZ in fünf Spiegelpunkten:

- Hochleistungsrechner, Speicher- und Visualisierungssysteme ausgerichtet auf die Klimaforschung
- Parallelisierung und Optimierung von Klimamodellen und Arbeitsabläufen
- Effizientes Management größter Datenmengen
- 3D-Visualisierung zur Vermittlung von Forschungsergebnissen
- Unterstützung aktueller Forschungsarbeiten im Bereich der Klimaforschung

Unterstützung Data Life Cycle und virtuelle Forschungsumgebungen



Träger des DKRZ

Gesellschafter

Die Gesellschafter der DKRZ GmbH		Anteil (ca)
 MAX-PLANCK-GESellschaft	Max-Planck-Gesellschaft	55%
 Universität Hamburg DER FORSCHUNG DER LEHRE DER BILDUNG	Freie und Hansestadt Hamburg, vertreten durch die Universität	27%
 AWI	Alfred-Wegener-Institut für Polar- und Meeresforschung	9%
 Helmholtz-Zentrum Geesthacht Zentrum für Material- und Küstenforschung	Helmholtz-Zentrum Geesthacht (HZG)	9%

Das DKRZ ist eine gemeinnützige und nicht kommerzielle GmbH mit vier Gesellschaftern.



Das DKRZ wird finanziell gefördert vom Bundesministerium für Bildung und Forschung

Gebäude



FHH BWF UNI

Deutsches Klimarechenzentrum Bundesstraße 45 Januar 2008

Lehmann + Partner Architekten

Rechnerraum



Rechnerknoten

Festplattensystem

Luftkühlung

Computer

- IBM Power6-System
- 264 Rechnerknoten mit 8448 Rechnerkernen
- Taktrate 4,7 GHz
- Rechnerleistung pro Kern 18,8 GFLOPS
- Spitzenrechnerleistung 159 TFLOPS
(Planung 2014: 3 PFLOPS / x 20)
- Linpack 110 TFLOPS und Platz 72 in der TOP500 im Juni 2011
- Speicherausbau pro Knoten 64 GB / 128 GB
- Festplattenspeicher 7 PB (Planung 2014: 150 PB / x 20)
- Verbindungsnetz 8-fach DDR-Infiniband
- Kühlung 75% Wasser, 25% Luft

Archivsystem

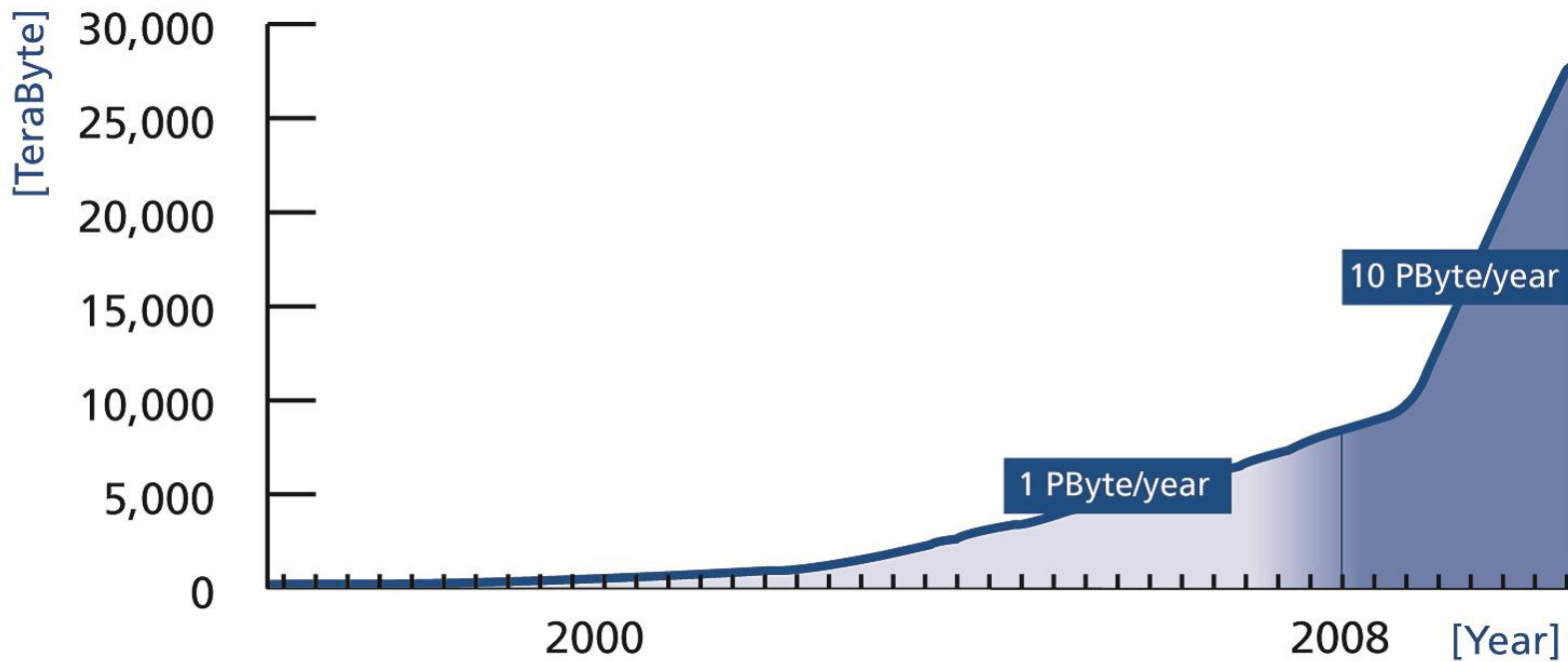


Archivsystem

- ▶ HPSS – High Performance Storage System
- ▶ 7x Sun StorageTek SL8500
- ▶ Insgesamt 67.000 Medienstellplätzen
- ▶ Mehr als 100 PB Speicherkapazität
(Planung 2014: 1 Exabyte / x 10)
- ▶ 90 Bandlaufwerke
 - LTO-5, LTO-4, T10000A/B
 - 9940B, 9840C

Diskussion mit
Klima-Community

Datenspeicherung am DKRZ



World Data Center for Climate

The screenshot shows the homepage of the World Data Center for Climate, Hamburg. At the top left is the M&D logo (Modelle & Daten). The main header reads 'World Data Center for Climate, Hamburg' and includes a globe icon with 'WDC' and a 'CERA' logo. Navigation links include 'Not logged in (Login) | Process List', 'CERA UI Home | WDC Home | Impressum', and 'Browse CERA Experiments'. A search section titled 'Select keyword(s) and/or project' contains two scrollable lists of keywords and projects. Below this is an 'Experiments' section with a list of various climate studies and models.

M&D
Modelle & Daten

**World Data Center
for Climate, Hamburg**

WDC

CERA

Not logged in (Login) | Process List

CERA UI Home | WDC Home | Impressum

Browse CERA Experiments

Select keyword(s) and/or project

aerosol
aerosol: indirect effect
air-sea-fluxes
airborne
aircraft measurement
AMIP
AMIP2
analysis
arctic
atmospheric chemistry

AQUA_AMSRE
AVHRR Pathfinder SST v5
BALTEX
CANIBALT
CARIBIC
CARIBIC-LH
CEOP
CiSAP
CLM_regional_climate_model_runs
CMIP5_test

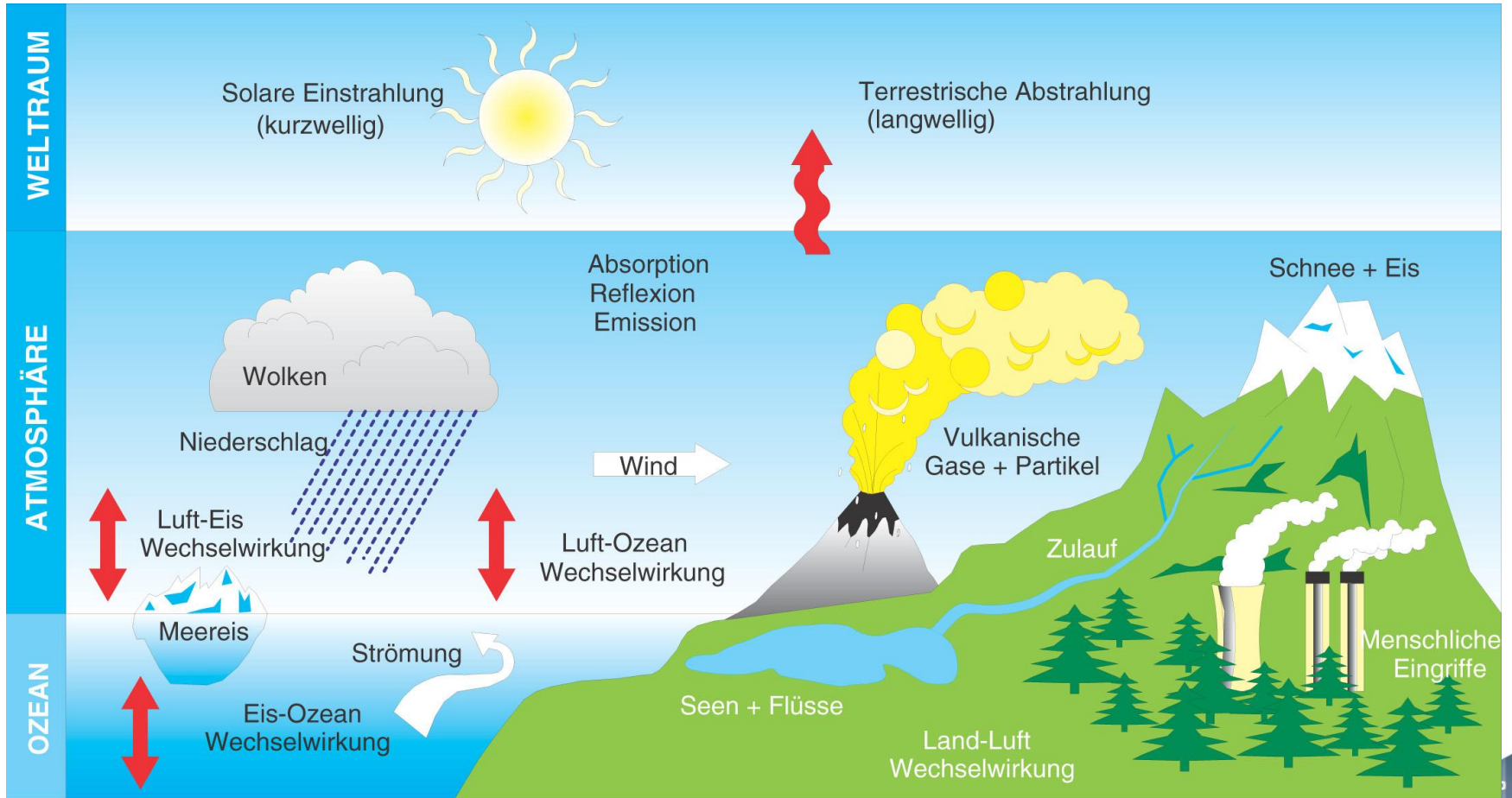
Select keyword(s) Clear selection Select project Clear selection Project information

Experiments

ACSYS 2003 - Arctic Atmospheric Boundary Layer and Sea Ice Interaction Study
Advanced Microwave Scanning Radiometer REMSS Version 5 SST
Advanced Very High Resolution Radiometer: sea surface temperature, monthly mean
amf_black_forest: links to AMF (ARM Mobile Facility) data taken during the COPS Campaign
AVHRR
Baltex Meteorological Data Centre (BMDC) from 31.03.2003
BASIS 1998 - Baltic Air Sea Ice Study 1998
BASIS 2001 - Baltic Air Sea Ice Study 2001
CC01CI01 - CTL: THE CONTROL INTEGRATION
CC01GG01 - GHG: THE GREENHOUSE GAS INTEGRATION
CC01GS01 - GHS: THE SULPHATE AEROSOL AND GREENHOUSE GAS INTEGRATION
CC01GS02 - GHS: THE SULPHATE AEROSOL AND GREENHOUSE GAS INTEGRATION
CC01GS03 - GHS: THE SULPHATE AEROSOL AND GREENHOUSE GAS INTEGRATION
CEOP Inter Transferability Studies: daily mean values of ECPC Regional Spectral Model (2007, JUNE)
CEOP Inter-Continental Transferability Study: daily mean values of Canadian Regional Climate Model (OURAN)
CEOP Inter-Continental Transferability Study: daily mean values of Climate version of the DWD "Lokalmodell"
CEOP Inter-Continental Transferability Study: daily mean values of ECPC Regional Spectral Model (2005, AUC)
CEOP Inter-Continental Transferability Study: daily mean values of Global Environmental Multiscale Limited Ar
CEOP Inter-Continental Transferability Study: Model Output Location Time Series (MOLTS NEW!)
CEOP Inter-Continental Transferability Study: Model Output Location Time Series (MOLTS) data provided by

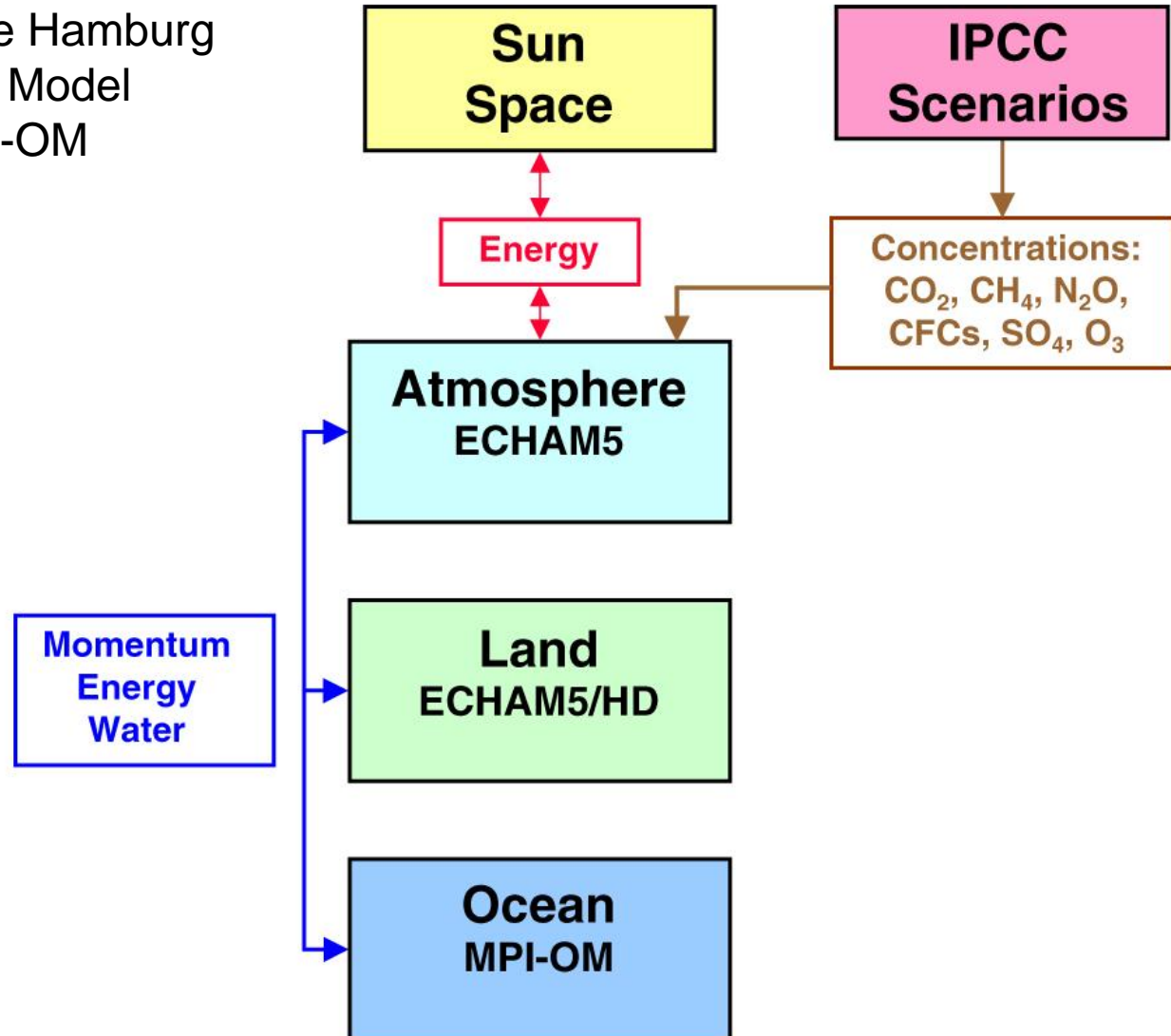
- 2003 Anerkennung durch ICSU
- Langzeitarchivierung
- 500 TB Klimadaten
- Vollständig dokumentiert
- Suchmaschine
- Feld-basierter Zugriff
- Freie Nutzung der Daten

Klimaforschung

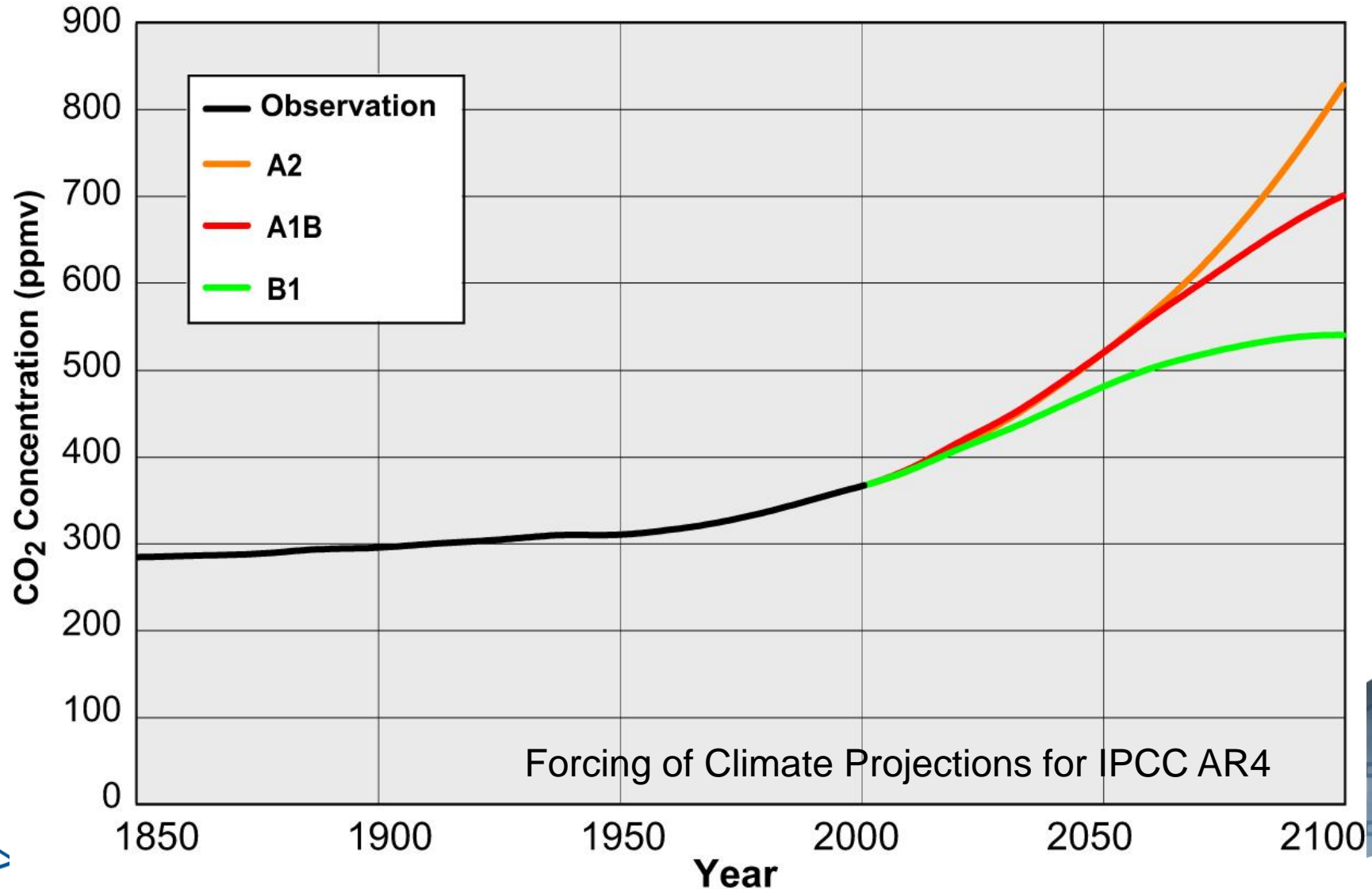


Klimaforschung

Diagram of the Hamburg
IPCC-Climate Model
ECHAM5/MPI-OM



Klimaforschung: Schlüssel-Szenarien IPCC-AR4 (2007)



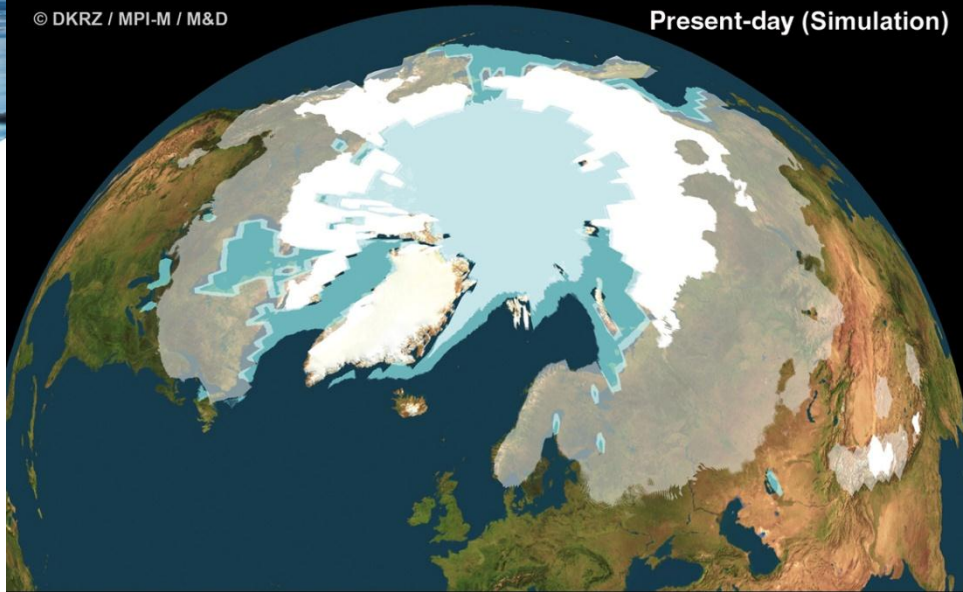
Forcing of Climate Projections for IPCC AR4



IPCC Climate Simulations with ECHAM5 / MPI-OM

© DKRZ / MPI-M / M&D

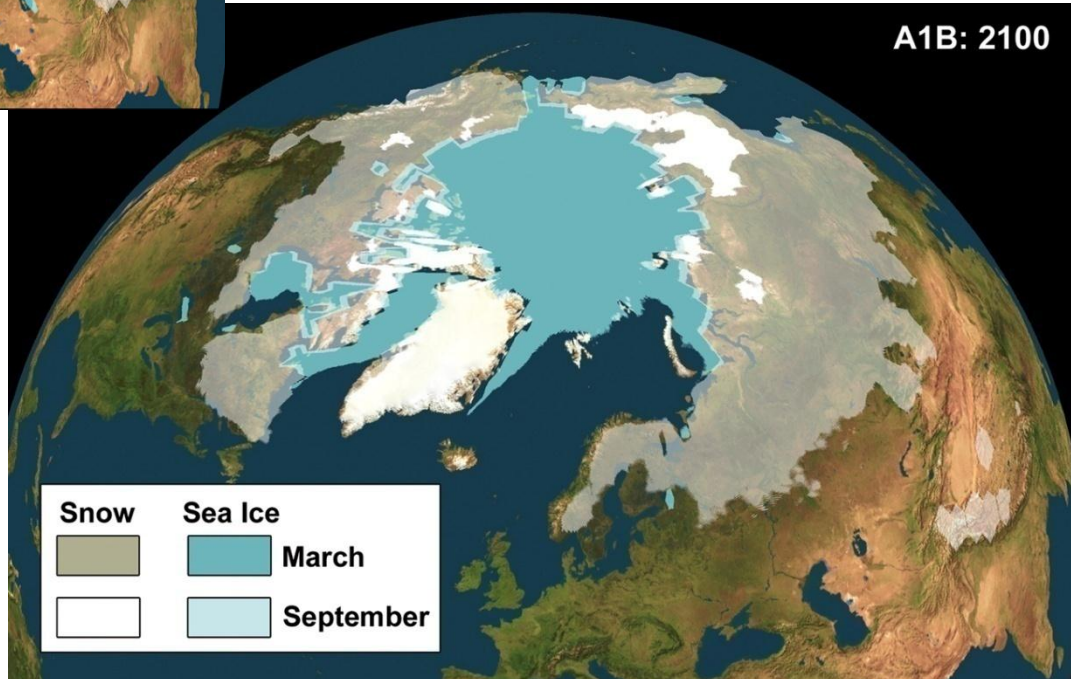
Present-day (Simulation)



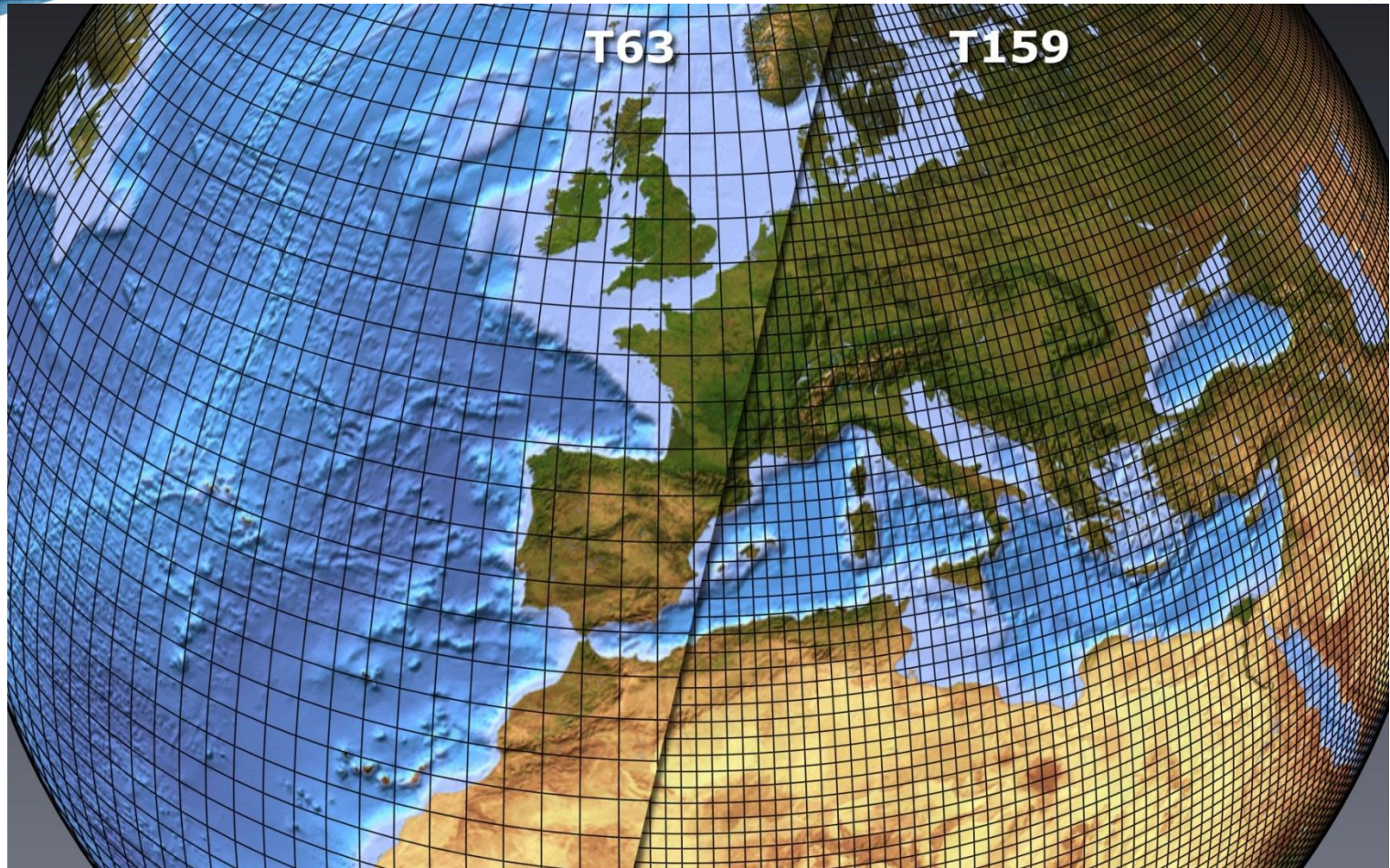
Klimaforschung

Comparison of the present-day sea ice cover
In March and September
(oben) with the climate
projection for the scenario A1B
(unten) in 2100.
Additionally the snow over land
can be obtained.

A1B: 2100



Klimaforschung



Halbierung der räumlichen Auflösung führt zu Mehraufwand von Faktor 16 ($= 2^{**}4$) und Faktor 8 in der Datenproduktion.

CMIP5: Organisation

At a **September 2008 meeting involving 20 climate modeling groups** from around the world (i.e., most of the major groups performing climate change research today), the **WCRP's Working Group on Coupled Modelling (WGCM)**, with input from IGBP's AIMES, agreed on a new set of coordinated climate model experiments, to be known as phase five of the Coupled Model Intercomparison Project (CMIP5).

*CMIP5 is meant to provide a framework for **coordinated climate change experiments** for about the next five years and thus includes **simulations for assessment in the AR5 as well as others that extend beyond the AR5**. CMIP5 is not, however, meant to be comprehensive; it cannot possibly include all the different model intercomparison activities that might be of value, and it is expected that various groups and interested parties will develop additional experiments that might build on and augment the experiments described here. In the IPCC assessment context, it is expected that **CMIP5 will provide information of value to all three IPCC Working Groups**.*

Taylor, Stouffer, Meehl, Dec. 2009

Klimaforschung: CMIP5 / IPCC-AR5 (2013)

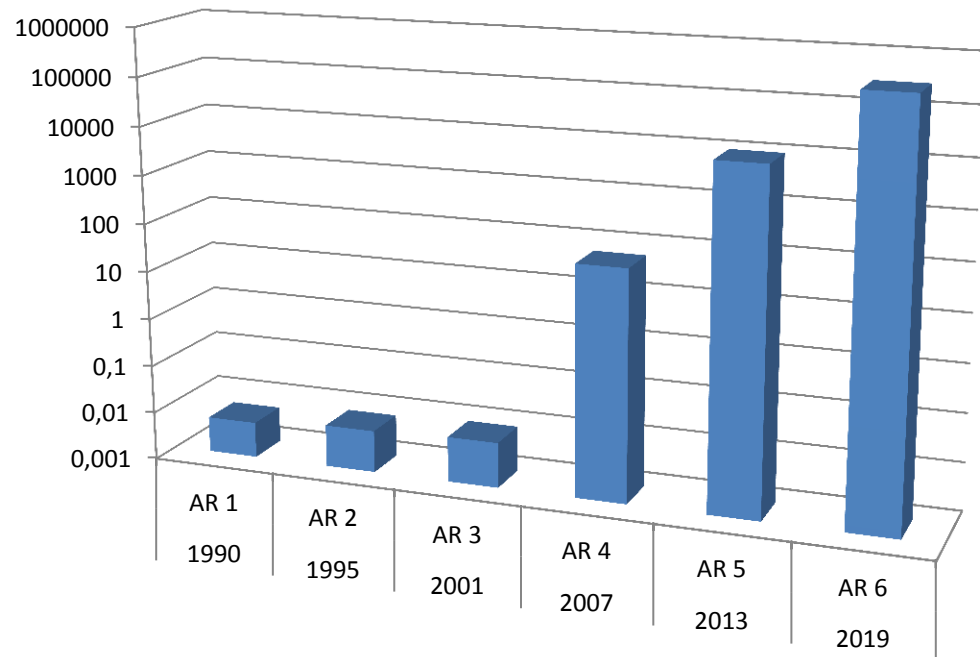
ECHAM6/MPI-OM

(Deutscher Beitrag zum
IPCC-AR5)

Rechenaufwand:
20,000,000 CPU-h

Klimamodelldaten: 0,6 PB

Personalaufwand: 10 PJ



Klimamodelldaten zu den Statusberichten
des IPCC (International Panel on Climate
Change)

Forschungsdateninfrastrukturen

Datenprovider

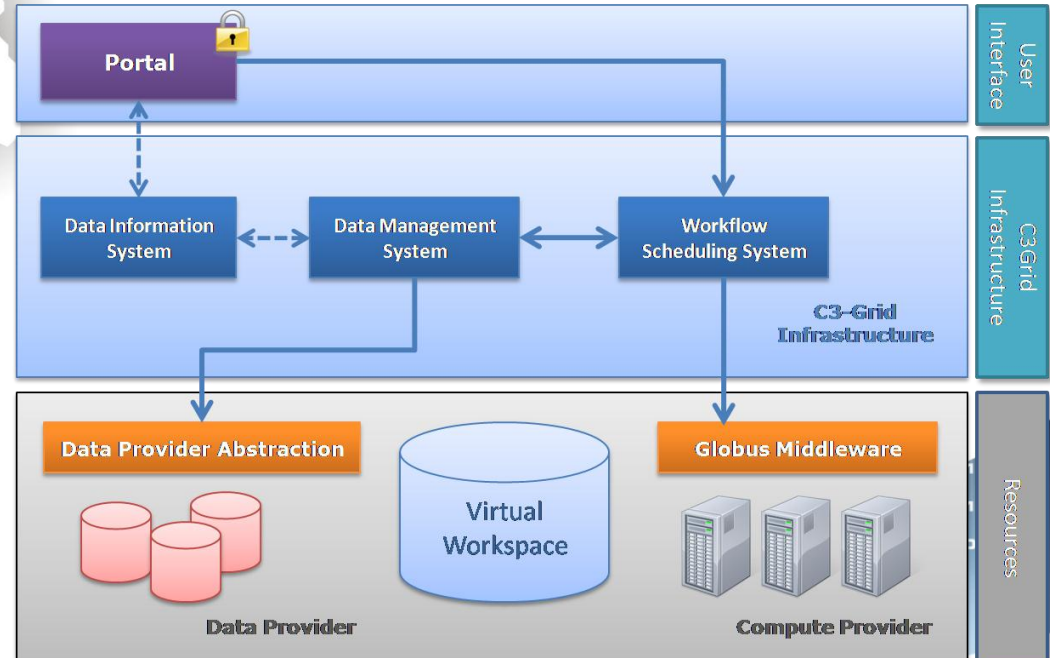
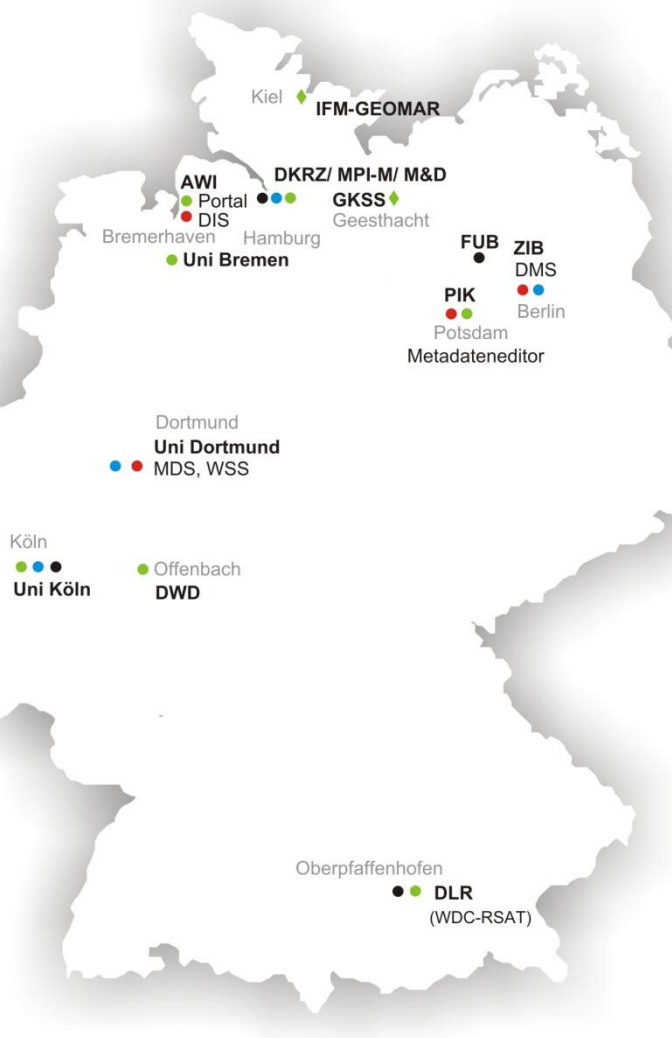
- Zugriff auf eigenes Archiv
- ◆ DKRZ Archiv
- Zentrale C³ Middleware
- Compute-provider
- Workflow-provider

*Collaborative Climate Community
Data and Processing Grid –
C3Grid™*

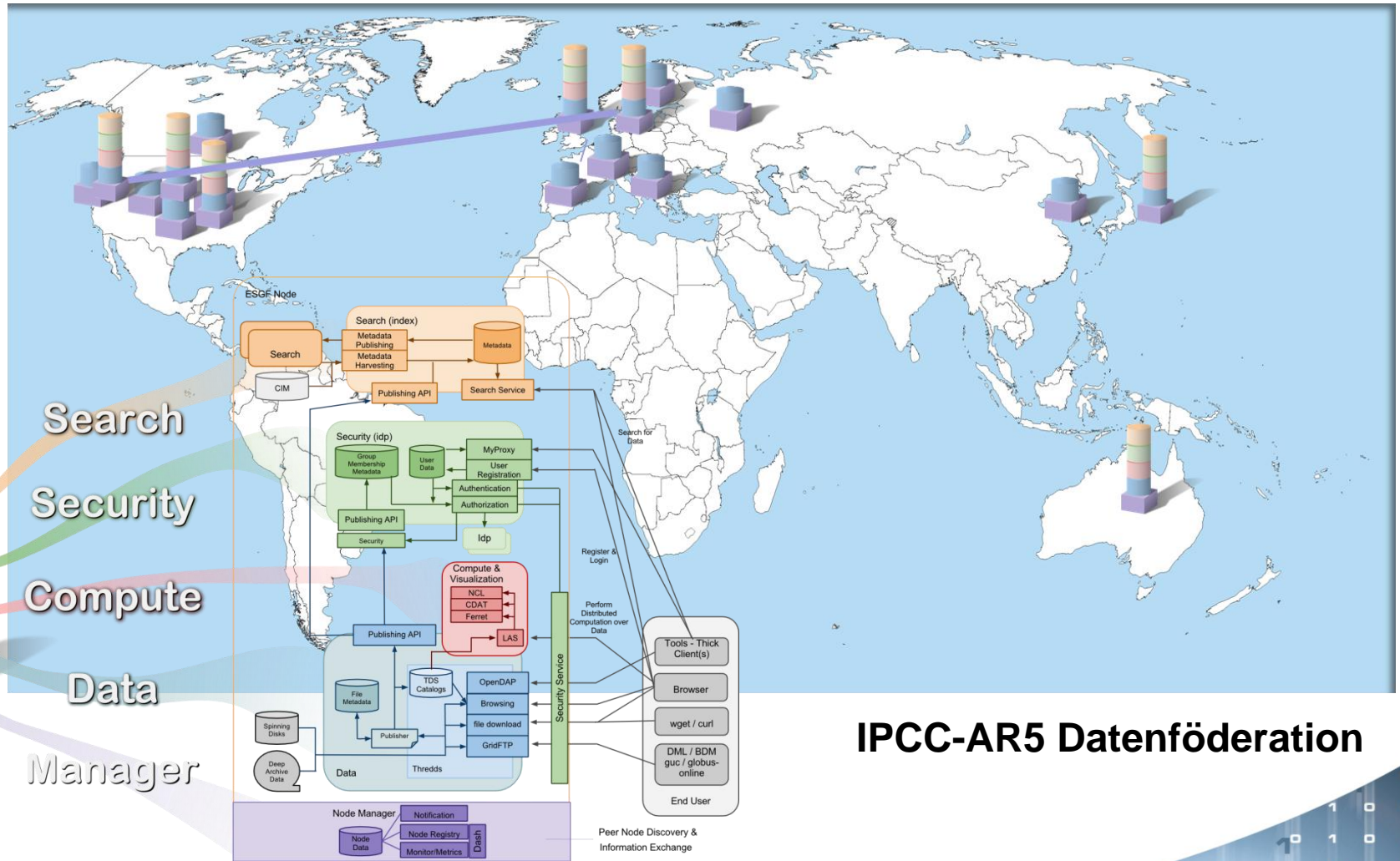
BMBF Förderung:

I) 01.09.2005 – 28.02.2009

II) 01.10.2010 – 30.09.2013

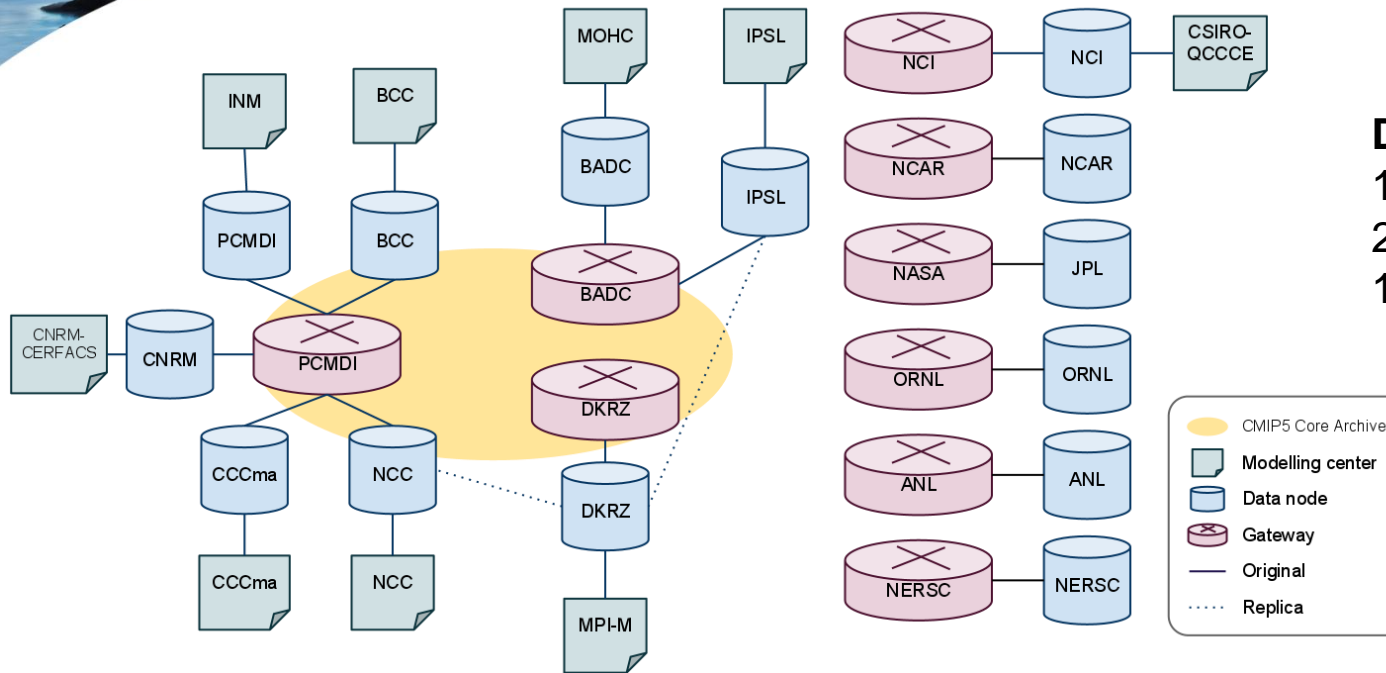


Forschungsdateninfrastrukturen



IPCC-AR5 Datenföderation

CMIP5 Data Federation



Data estimates 2010:
 10 PB in total
 2.5 PB WCRP requested
 1 PB IPCC-AR5 core

Summary

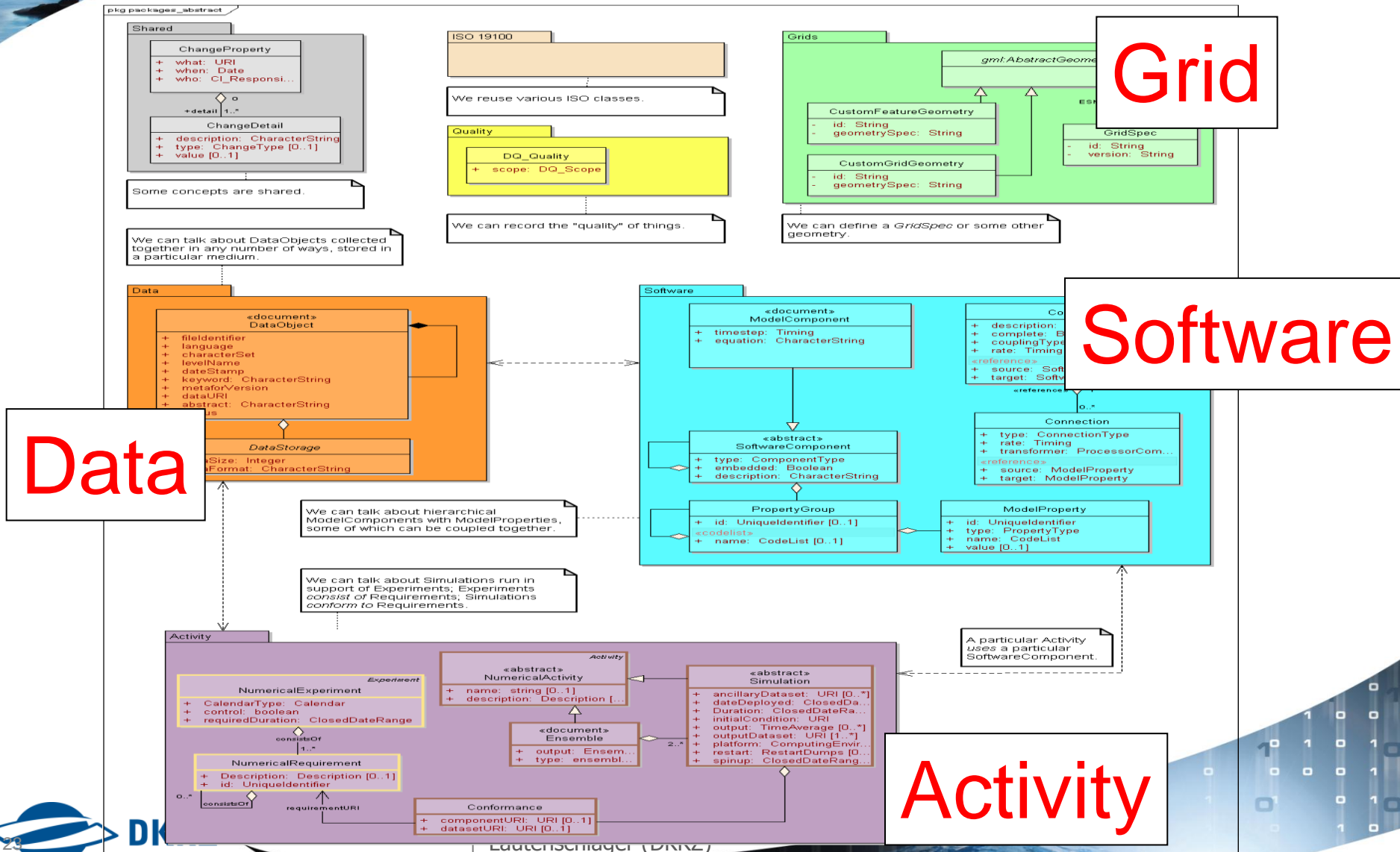
<i>Modeling centers</i>	13
<i>Models</i>	17
<i>Data nodes</i>	13
<i>Gateways</i>	5
<i>Datasets</i>	15950
<i>Size</i>	177.25 TB
<i>Files</i>	407792

ESG infrastructure for CMIP5
 provided by
 NCAR (ESG Portal)
 PCMDI (ESG Data Node)

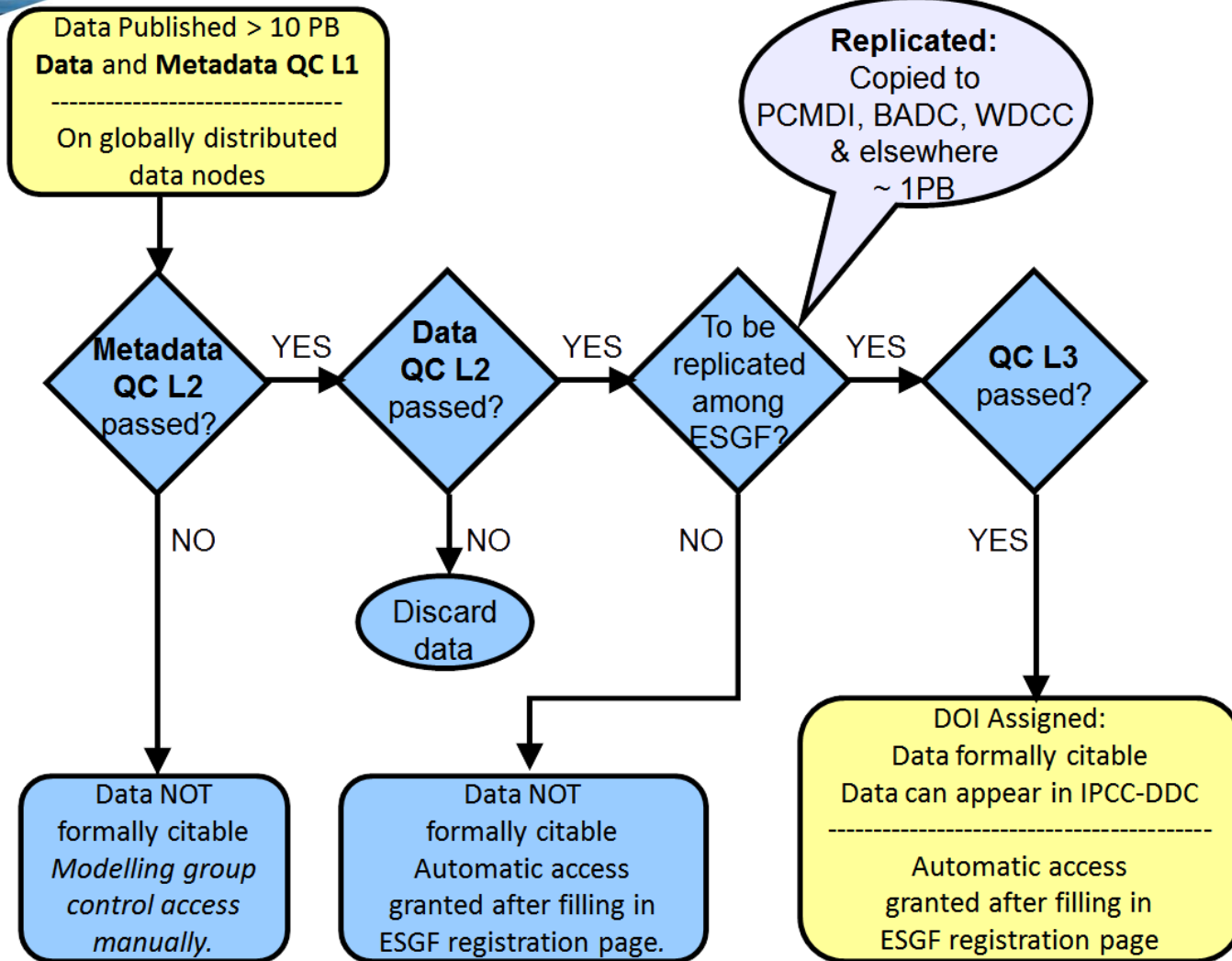
CMIP5 Archive Status

Monday, 19. September 2011

Common Information Model (CIM)

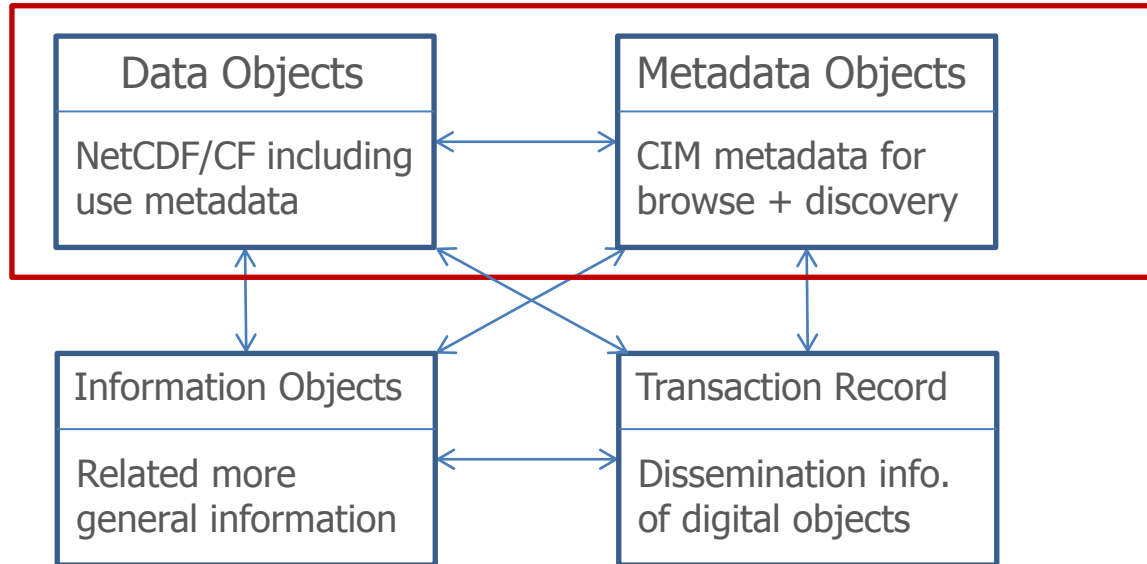


Qualitätskontrolle



(Informal citation still requested where formal citation not available)

Digital Object Architecture of Climate Model Data



DataCite scientific data publication entity:

- DOI has been assigned
- Digital objects are frozen and approved by author
- Citation reference is assigned for direct use in scientific literature
- Realized with QC-L3 in the CMIP5 data quality assessment

Future Development: Identification of distinct data objects in data federations with PID and handle system (Cooperation with European Persistent Identifier Consortium (EPIC), <http://pidconsortium.eu/>).



Workflow Processes

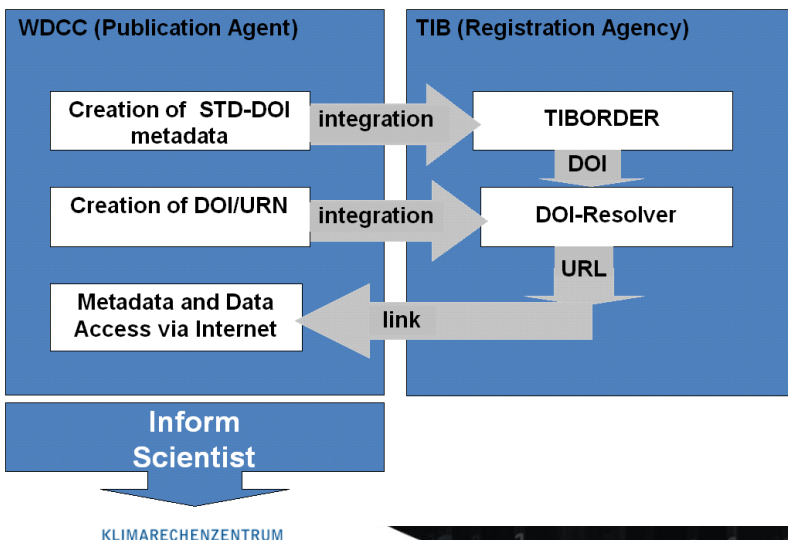
	Permission	SQA	TQA	Publication
Scientist				
WDCC				
TIB				

Scientific Quality Assurance – SQA

Technical Quality Assurance – TQA

TIME →

Publication



Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences



DKRZ

universität**bonn**

World Data Center for Climate, Hamburg



Atarrabi Publication
System (V1.5 Beta)

You are not logged in. | [About](#) | [Impressum](#)

Login

Welcome to the Atarrabi Publication System!

Please login using your existing CERA account credentials to publish data or to view work in progress.

Username

Password

Create a new CERA account:
 [Register new CERA user](#)

Login

About Atarrabi

The ATARRABI interface provides a key part of the STD-DOI publication process. Please choose your data domain:

CMIP5

Environmental Data

The ATARRABI interface provides a key part of the STD-DOI publication process of [CMIP5](#) core data by supporting the authors final quality control of data and metadata. After publication the data remains unchanged. A copy of the data is stored on tapes in the long-term archive of the World Data Center for Climate ([WDCC](#)). Two persistent identifiers ([DOI](#), [URN](#)) are assigned, which can be used for data citation for the credit of the data originator as well as to provide provenance information for derived data product. This step finalizes the overall [CMIP5 quality process](#).
The [WDCC](#) as publication agency is responsible for the STD-DOI process within CMIP5, which follows the [DataCite](#) regulations. The [German National Library of Science and Technology \(TIB\)](#) functions as STD-DOI registration agency, which is currently the managing agency of the DataCite consortium.

Preconditions to start the CMIP5 STD-DOI publication process are:

- a personal WDCC account and a special permission to start the STD-DOI publication process for the related primary data: Click [here](#) to apply for a WDCC account.
- long term availability of primary data in [ESG](#)
- long term availability of metadata in [CIM](#)
- open access to primary data and metadata

This program is web-based and therefore doesn't require any extra client-side software installation or special firewall configuration. Contact: data@dkrz.de

About this software



This software is funded by the [German Research Foundation](#)



Learn more about the research project [Publication of Environmental Data](#)

Künftige Klimadatenföderationen

Wachstum der Archivgrößen von PetaByte nach ExaByte erfordert:

- **Ausreichende Information**, um Daten zu suchen und zu finden und um sie interdisziplinär nutzen zu können.
- **Ausreichende Standardisierung** zur automatisierten Bearbeitung großer Datenmengen
- Anerkannte **Flags für Datenqualität** zur Stärkung des Vertrauens in die Daten der Föderation
- Neue Methoden, um **Informationen in Datenföderationen** zu erkennen und zu extrahieren
- Unterstützung des **gesamten Lebenszyklus von Daten** (Erzeugung – Bearbeitung – Langzeitarchivierung – Verteilung), um große/riesige Datenarchive verwalten zu können