

The role of climate model data and long-term data archives in climate change research

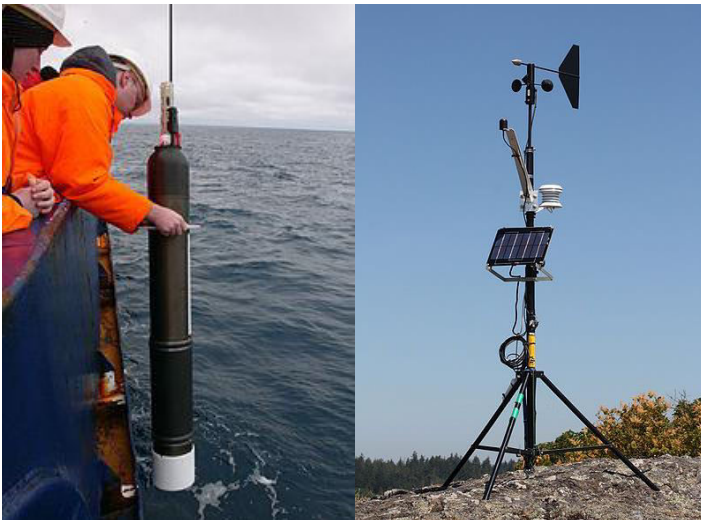
World Conference on Climate Change
October 24-26, 2016 Valencia, Spain

Martina Stockhause, Michael Lautenschlager
Deutsches Klimarechenzentrum (DKRZ)

Observations and Models (1)

Characteristics of Observations

- Diverse: different instruments measuring different parameters
- Discrete: certain spatial-temporal coverage
- Continuously extended
- No future information



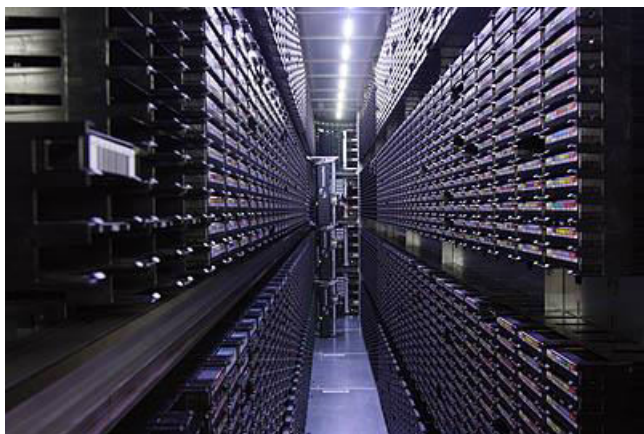
Observation Data Management

- **Big Data:** many different parameters and data formats
- Products for different research purposes, esp. satellite data
→ new versions: reprocessing with improved algorithms
- Description on measurement conditions and provenance required

Observations and Models (2)

Characteristics of Climate Model

- Relatively homogeneous, with standardized formats and naming conventions, e.g. CF
- Static (Once created and many times analyzed)
- 4D data for many parameters as mean values for grid cells over a time interval
- Future projections and scientific questions



Model Data Management

- **High Volume Data:** PBytes of homogeneous data created
- New data versions are rare
→ for post-processed datasets
- Data access of data subsets
- Detailed description of data subsets required

→ Reanalysis Models for observation data assimilation

IPCC Data Distribution Centre (DDC)

IPCC DDC and TGICA

IPCC DDC (Data Distribution Centre) – ipcc-data.org

jointly managed by:

- British Atmospheric Data Centre (BADC):
Climatologies
- World Data Center Climate (WDCC) at DKRZ:
Reference Data Archive for climate model output
- Center for International Earth Science Information Network (CIRESIN) at Columbia University:
social-economic data archive

→ Certified ICSU World Data System (WDS) members



WORLD DATA SYSTEM

IPCC TGICA

(Task Group on Data and Scenario Support for Impact and Climate Analysis)

- Oversees IPCC DDCs
- Enables research and sharing of information across the IPCC Working Groups

→ Mandate and structure of TGICA are currently under review by the IPCC

History of IPCC DDC

- 1995: IPCC SAR climate model data long-term archived
- 1998: IPCC DDC formally established
- 2008: parts of FAR data added to DDC
- 2013/14: IPCC DDC AR5 data long-term archival
- 2016: IPCC Task Force built for transformation of the organization of IPCC data and information to serve the needs of the IPCC during and beyond AR6.
- 2020/21: IPCC DDC AR6 long-term archival

IPCC DDC Reference Data Archive

The IPCC DDC provides data on the long-term for an interdisciplinary user community in support of the IPCC Authors.

Long-term:

archival with second data copy in an established data center

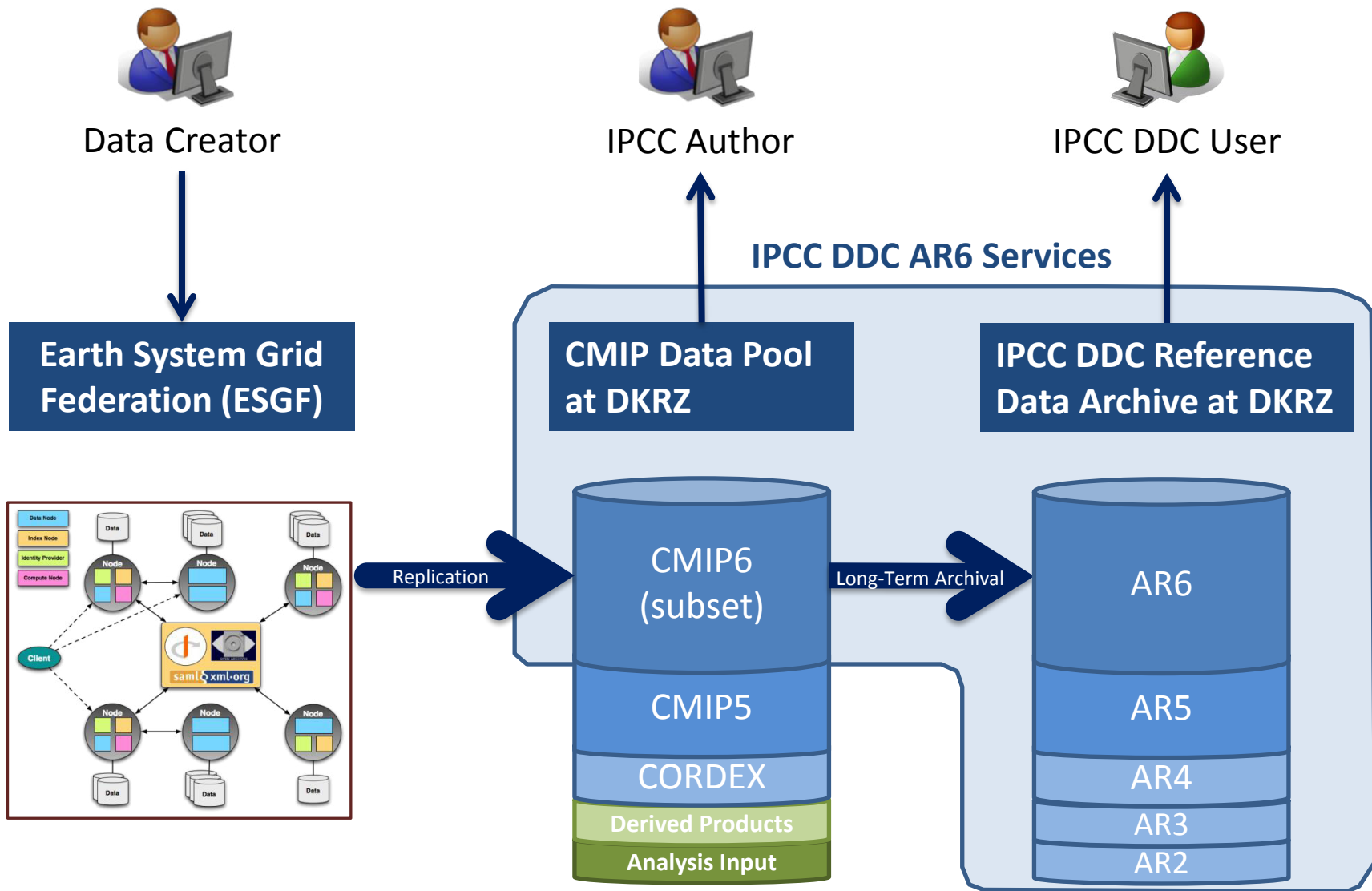
Interdisciplinary Use:

add information to the data for a creator-independent usage

IPCC Author Support:

provide a reliable, up-to-date and easily-accessible CMIP6 data pool

IPCC DDC Services for AR6



Data Management

CMIP Data Pool

Replicate to build CMIP Data Pool

Challenges:

- Timely automated update of evolving data
- Provide easy and script-based access to data
- User support
- Coordination nationally and internationally

Implementation:

- Synda tool used based on ESGF Search API
- User accounts on Linux machine provided
- Request tracker
- Task group built



Principles of Long-Term Archival (1)

Gather what you can...
...as long as it is available.

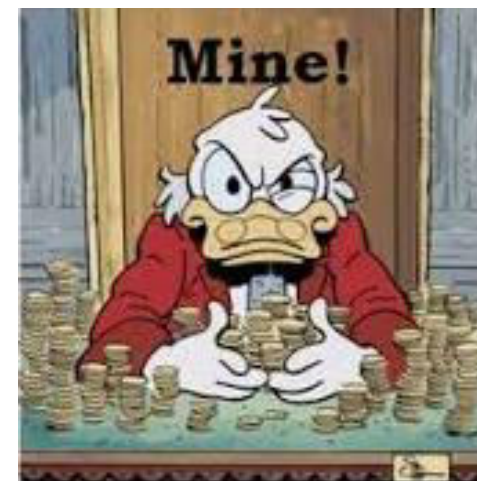
Metadata challenges:

Ancillary metadata is diverse in respect to:

- Granularity,
- Format,
- Access,
- Stability.

Implementation:

Registration of ancillary metadata URL in relation to the data by metadata providers.



Principals of Long-Term Archival (2)

Automate what you can...
...for a timely archival.

Automated access, interpretation and mapping of metadata sorts the different pieces of metadata in the hierarchical structure of the LTA metadata schema.



Principals of Long-Term Archival (3)

Check everything at least twice...
...before archival.

The automated process is interrupted at several stages in order to ensure metadata consistency. After archival such a quality assurance is unfeasible.



Data Usage

IPCC DDC Users

Climate researchers like IPCC WG I Authors:

- familiar with data formats and tools to analyze the data
- skills to interpret and use data without additional services

Climate impact researchers like IPCC WG III Authors:

- need information on data formats and analysis tools
- need more information on how to use the data
- additional at technical services requested like derived climate parameters, data regridding

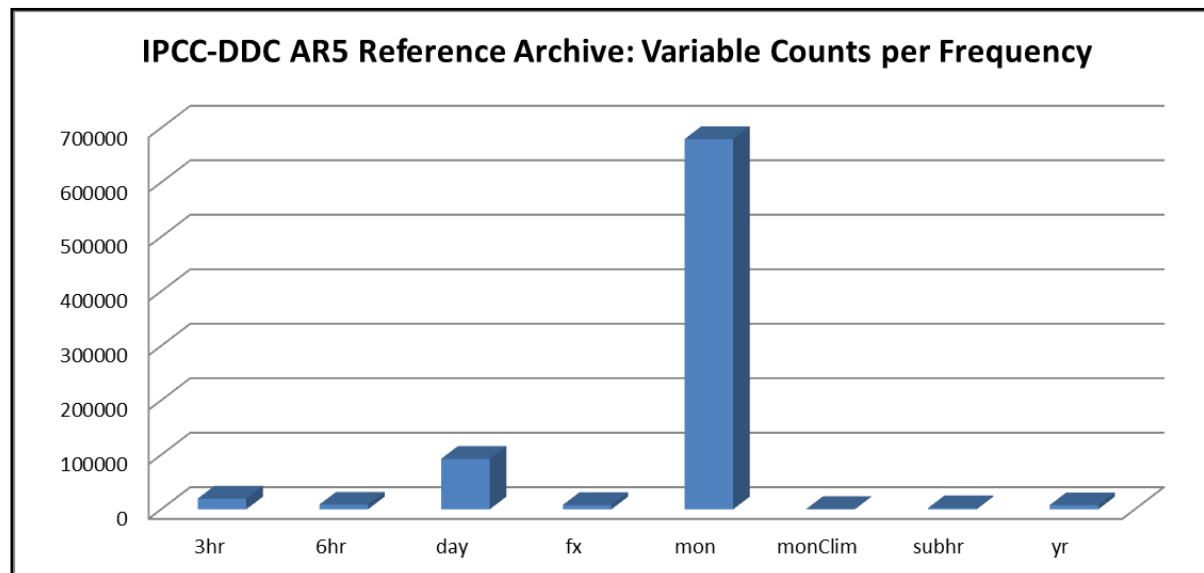
Policy advisors:

- need assistance on data formats, tools and data interpretation
- Personnel as well as technical services required, ideally a climate service center with trained personnel

IPCC DDC: AR5 Reference Archive

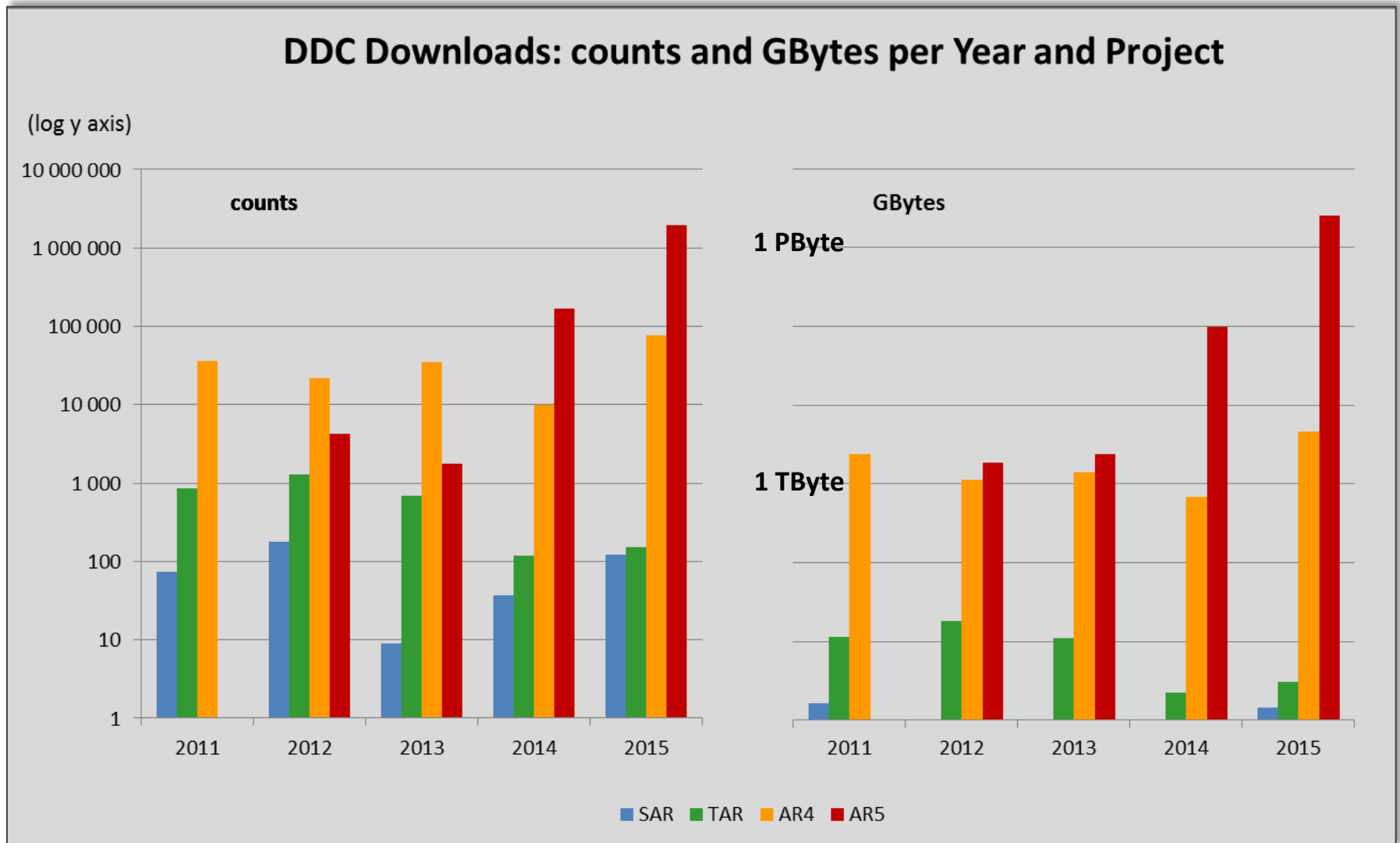
The DDC Reference Archive / The IPCC WG1 Archive

Experiments:	101 / 78	different experiments / scenarios
Variables:	605 / 123	different variables (628 requested variables)
Size:	1.6 PByte / 100 TByte	(all AR data: 1.7 PByte)
Models:	60 / 58	participating models
Institutes:	27 / 24	participating institutes
Simulations:	1145 / 952	provided simulations
Variables:	818795 / 93247	provided variables



IPCC-DDC: Usage (1)

Reference Archive for Climate Model Output Data



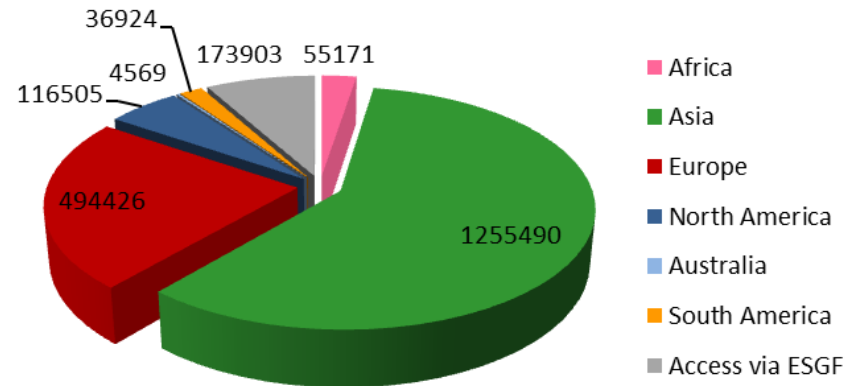
IPCC-DDC Usage (3)

Reference Archive for Climate Model Output Data

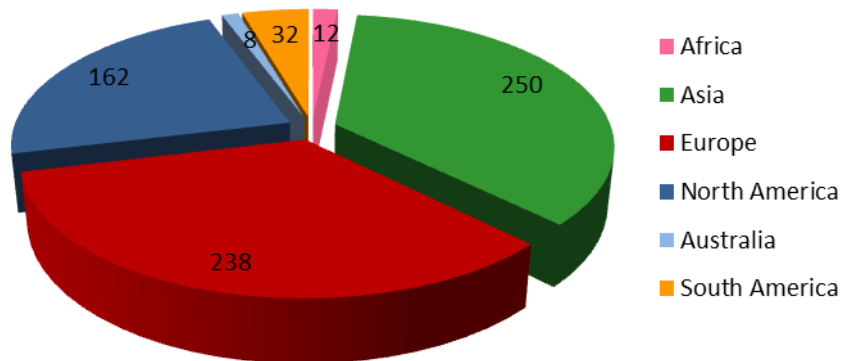
702 Active DDC Users in 2015:

- 42% located in developing and economy-in-transition countries (Africa, Asia, South America)
- 15 users requested regional data on storage media

DDC Download Counts by Continents in 2015



Number of Active DDC Users in 2015 by Continents



Download Counts per Continent in 2015:

- 69 % of downloads from users in developing and economy-in-transition countries (59 % Asia)
- average download count per user = 2 800
- Asian and African users were the most active with average download numbers per user of 5 000 and 4 600.

Data Distribution Centre

ipcc climate change

INTEGOVERNMENTAL PANEL ON climate change

UNEP WMO

Search Advanced search Help Site map IPCC web sites

Location: Home Data Simulations AR5

DDC AR5 Reference snapshot

The DDC AR5 Reference snapshot was collected by the World Data Center Climate (WDCC) at the DKRZ in Germany, the British Atmospheric Data Centre (BADC) in the UK and the Program for Climate Model Diagnosis and Intercomparison (PCMDI) University of California in the US. The DDC AR5 Reference snapshot is based on the status of the CMIP5 data archive as of March 15, 2013.¹ The corresponding PCC WGI AR5 data snapshot can be found [here](#).

CMIP5 data provided through the IPCC DDC has undergone a [quality control procedure](#). To find individual information on data, data creation and data quality for an experiment, please follow the links in the below tables.

The CMIP5 archive is evolving and in some cases the data used in the IPCC 5th Assessment Report may have been superseded. Latest versions for all experiments are available in the Earth System Grid Federation (ESGF) at http://cmip-pcmdi.lln.gov/cmip5/data_portal.html. For corrections of data sets published under later versions please look at the [errata page](#) hosted by PCMDI. Further information on CMIP5 can be found on [this page](#). The data is provided in [netCDF/CF format](#).

¹ Please note that the full CMIP5 data archive continues to be updated with new results, corrections etc. in contrast to the two archives available from the DDC.

CMIP5 License Statement

These data were first published under the license of CMIP5. Terms of use for CMIP5 are applied for DDC-AR5 data. They are provided at <http://cmip-pcmdi.lln.gov/cmip5/terms.html>. Data from some modelling centres are licensed for use in non-commercial research and for educational purposes, other for unrestricted use. Please refer to the [terms of use for the CMIP5 modeling groups](#) for details. DDC-AR5 data should be cited by its DataCite DOI and according to the [citation recommendation of CMIP5](#).

Please note:

A table entry represents a CMIP5 experiment as general citation unit in the literature.

Green experiments have completed the quality control process and are assigned [DataCite DOIs](#) (Digital Object Identifiers). The underlying links resolve to the DOI landing pages, which provide basic information related to the data including access.

For grey experiments the data is securely archived in the WDCC and the quality control process is ongoing. The underlying links resolve to web pages displaying basic information related to the data including access.

How to download data is explained in the [data download guide](#).

Long-term Projections: Centennial and longer Scenarios

Centre(s)	Centre Acronym(s)	Model	pControl	amp	historical	rcp	other	esm	sst	palaeo	aqua	add. data	known issues
Beijing Climate Center China	BCC	BCC-CSM1.1	pControl	amp amp4K amp4xCO2 ampFuture	historical historicalGHG historicalNat	rcp26 rcp45 rcp60 rcp85	abrupt4xCO2 1pctCO2	esmControl esmFdbk1 esmFdbk2 esmFncCm1 esmFncCm2 esmHistorical esmrcp85	sstClim sstClim4xCO2 sstClimAerosol sstClimSulfate	midHolocene past1000	-	-	-
		BCC-CSM1.1(m)	pControl	amp	historical	rcp26 rcp45 rcp60 rcp85	abrupt4xCO2 1pctCO2	esmControl esmHistorical esmrcp85	-	-	-	-	-
Beijing Normal University China	BNU	BNU-ESM	pControl	amp	historical historicalGHG historicalMac historicalNat	rcp26 rcp45 rcp85	abrupt4xCO2 1pctCO2	esmControl esmHistorical esmrcp85	sstClim sstClim4xCO2	-	-	-	CMIP5 page

IPCC DDC:

<http://ipcc-data.org>

DDC at DKRZ:

<http://ipcc.wdc-climate.de>

M. Stockhouse, F. Toussaint, M. Lautenschlager (2015): CMIP6 Data Citation and LTA. WIP white paper. Zenodo. [doi:10.5281/zenodo.35178](https://doi.org/10.5281/zenodo.35178).