

# Climate data not available through ESGF: e.g. IPCC DDC

CP4CDS Workshop, 08.03.2018  
Reading, UK

M. Stockhause  
WDC Climate at DKRZ

# IPCC Data Distribution Centre (IPCC DDC)

IPCC DDC (Data Distribution Centre) – [ipcc-data.org](http://ipcc-data.org)

jointly managed by:

- British Atmospheric Data Centre (BADC) at CEDA:  
Climatology Data, hosting of web site
- World Data Center Climate (WDCC) at DKRZ:  
Reference Data Archive for climate model output
- Center for International Earth Science Information Network (CIESIN)  
at Columbia University:  
Social-economic data archive

→ Certified ICSU World Data System (WDS) members

## 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

→ DDC support group: direct exchange DDC and WG TSUs



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



**Center for International Earth  
Science Information Network**  
EARTH INSTITUTE | COLUMBIA UNIVERSITY



**WORLD DATA SYSTEM**

# IPCC DDC: AR5 Reference Data Archive Content

## The DDC Reference Archive / The IPCC WGI Archive

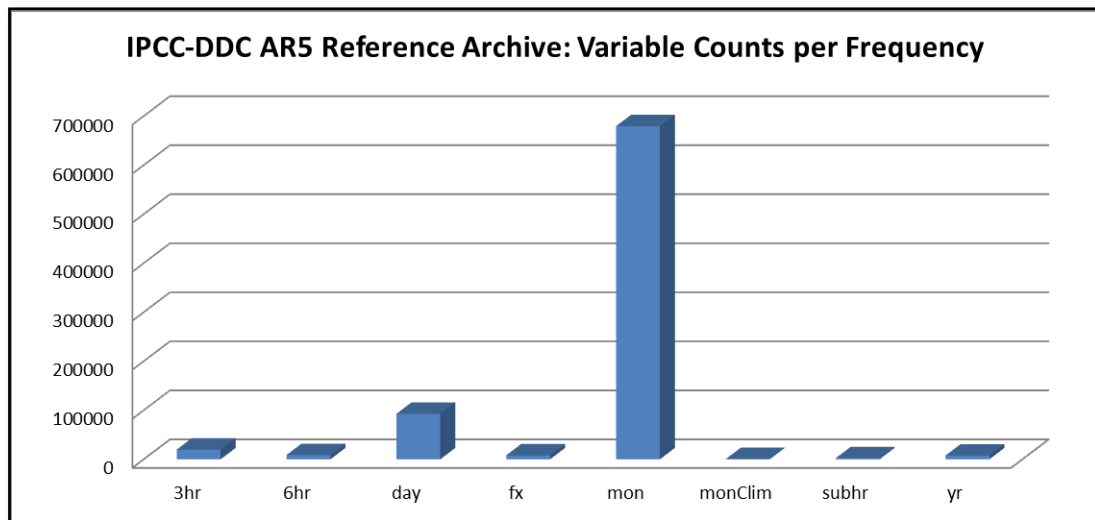
**Experiments** 101 / 78 different experiments /  
: scenarios  
**Variables:** 605 / 123 different variables  
(628 requested variables)

**Size:** 1.6 PByte / (all AR data: 1.7 PByte)  
100 TByte

**Models:** 60 / 58 participating models  
**Institutes:** 27 / 24 participating institutes  
**Simulations:** 1145 / 952 provided simulations  
**Variables:** 818 795 / 93 247 provided variables

## CMIP6 Numbers (2018-03-08):

21 endorsed MIPs + DECK  
248 different experiments (97 tier 1)  
2002 different variables  
82 models registered from 29 institutions  
20 PBytes estimated data volume



# ESGF vs. DDC Data

## Earth System Grid Federation:

- Disseminate data for different projects:  
CMIP5, most CORDEX domains, obs4MIPs, input4MIPs, CMIP6, CMIP3 added, some other...
- ESGF provides no general access to long-term archived data:
  - There is no commitment for long-term data curation and availability among the data nodes:  
Some data from some data nodes might disappear over time or some dataset versions might get unpublished.
  - ESGF Core Data Nodes have agreed for CMIP5/6 on project standards e.g. Quality Assurance (QA)

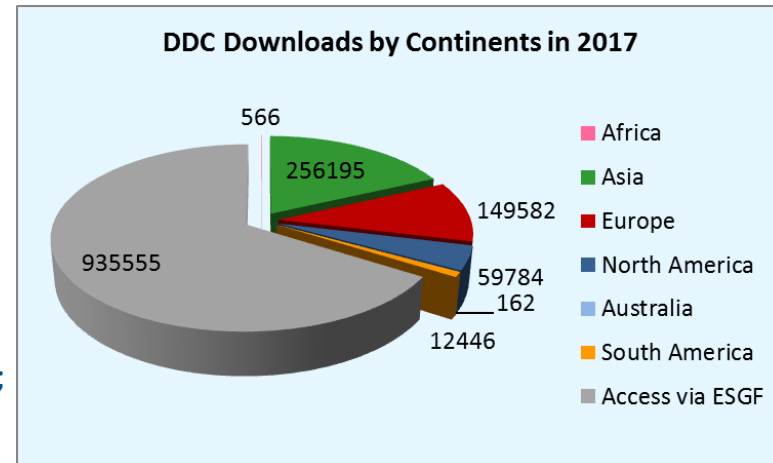
## IPCC Data Distribution Centre:

- Disseminate long-term archived data / user support:  
Long-Term Archive for IPCC data **snapshot** of CMIP data (plus selected MIPs like HAPPI-MIP) for interdisciplinary user community
- DDC applies quality standards:
  - IPCC DDC commits to long-term curation and accessibility of the Reference Data Archive
  - Addition of documentations (ancillary metadata)
  - IPCC DDC data complies to a quality standard (consistency + conformance; for CMIP5 see: <http://cmip5qc.wdc-climate.de>)

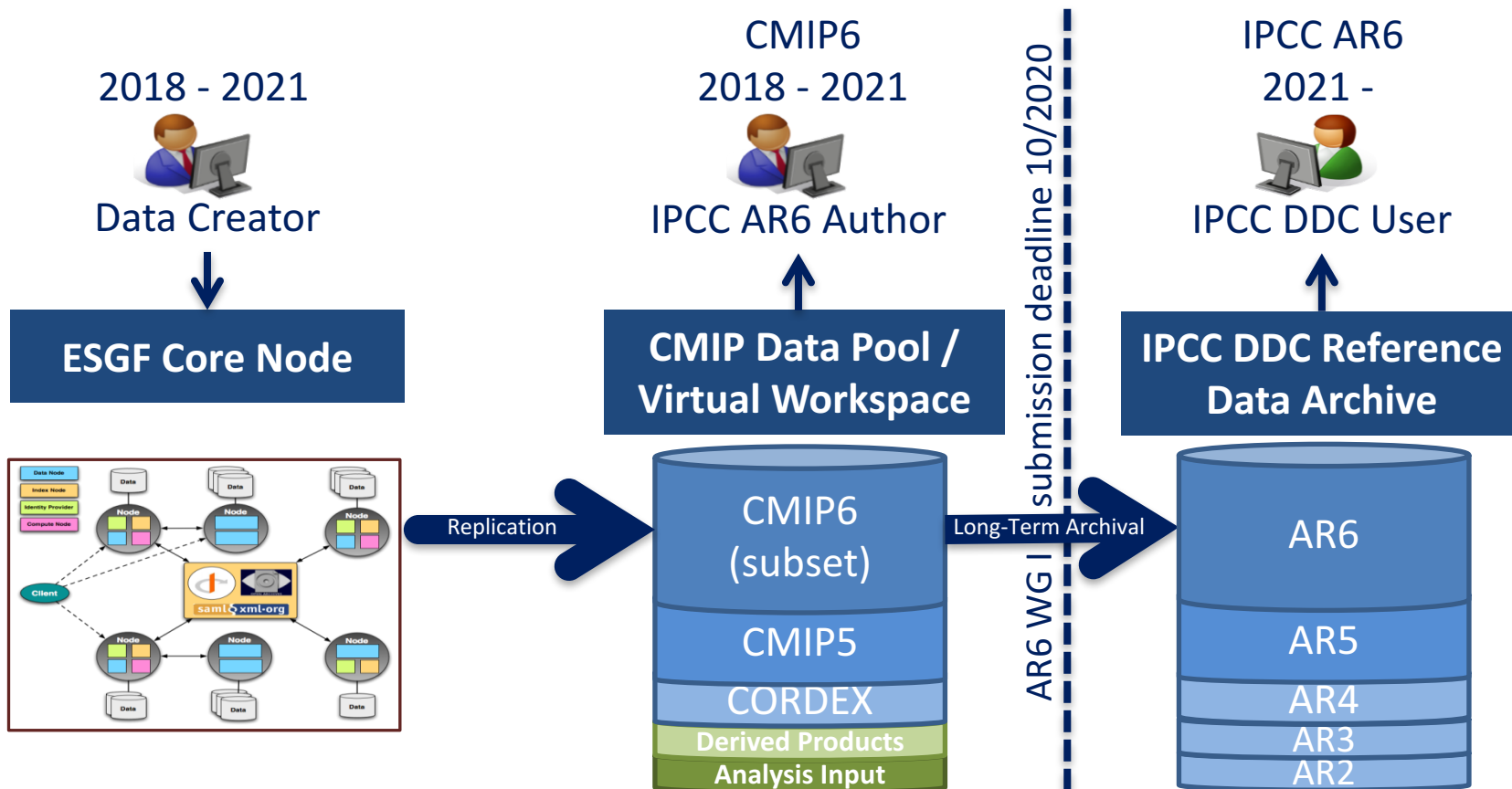
→ IPCC AR5 Reference data partly (DDC snapshot) accessible via ESGF;

AR5 WG I snapshot is only accessible via the IPCC DDC at DKRZ

→ IPCC AR6 Reference Data Archive will be made accessible via ESGF



# DKRZ within CMIP6, ESGF, and IPCC DDC AR6



# CMIP6 and IPCC AR6 DDC

## IPCC DDC Reference Data Archive for AR6:

- Data Pool as AR6 subset:
  - includes CMIP6 core data and requested data from German scientists and IPCC authors (version 2017-06-12: <http://bit.ly/2oNXdNs> )
  - quality-checked either prior to replication or after replication (tool as for QC2 in CMIP5: <http://cmip5qc.wdc-climate.de>) and QC3 (consistency and conformance checks) for the long-term archive
- Ancillary Metadata in the Reference Data Archive for AR6:
  - add ES-DOC information on model, experiment, simulation Reference Data Archive for climate model output
  - add data citation information from [Citation Service](#)
  - add [ESMVal](#) image/result? (low funding)
- Long-Term Metadata Curation:
  - use [Scholix](#) services to add references to articles citing CMIP6 data
  - add [ES-DOC errata](#) information

Trust and Quality

Improve reusability  
by lowering barriers  
for non-modelers

Long-term accessibility  
and curation

# Important Climate Data not in ESGF

## IPCC DDC Reference Data Archive not in ESGF:

1. **IPCC AR6 HAPPI-MIP / SR1.5** incl. forcing data (Special Report on 1.5 Degrees)  
→ high relevance: technical no problem but funding needed for  
ESGF access via DKRZ's LTA node
2. IPCC AR5 WGI Archive
3. Nice to have: IPCC AR4, IPCC TAR, SAR and available part of FAR Reference Data

## Other Data not or only partly in ESGF:

- Regional climate data CORDEX:  
Many domains like Euro-Cordex are in ESGF,  
some like **Med-Cordex** are brought into ESGF  
→ use of ESGF for all CORDEX domains in CMIP6
  - Data for impact user community, e.g. socio-economic data or observations
  - Data for validation, e.g. used by ESMVal  
(observations/reanalyzed observations) – license issues
  - Climate parameters/indices also collected and **defined** within the German  
Project ReKlies (<http://reklies.hlnug.de/>): [relevant climate indices](#) [German]  
ESGF: <https://esgf-data.dkrz.de/search/cordex-dkrz/?project=reklies-index> )  
→ Selection/definitions/names for climate indices to be checked with KNMI
- } Integration into WPS

# Excursus: Connections and Interlinking in CMIP6/AR6

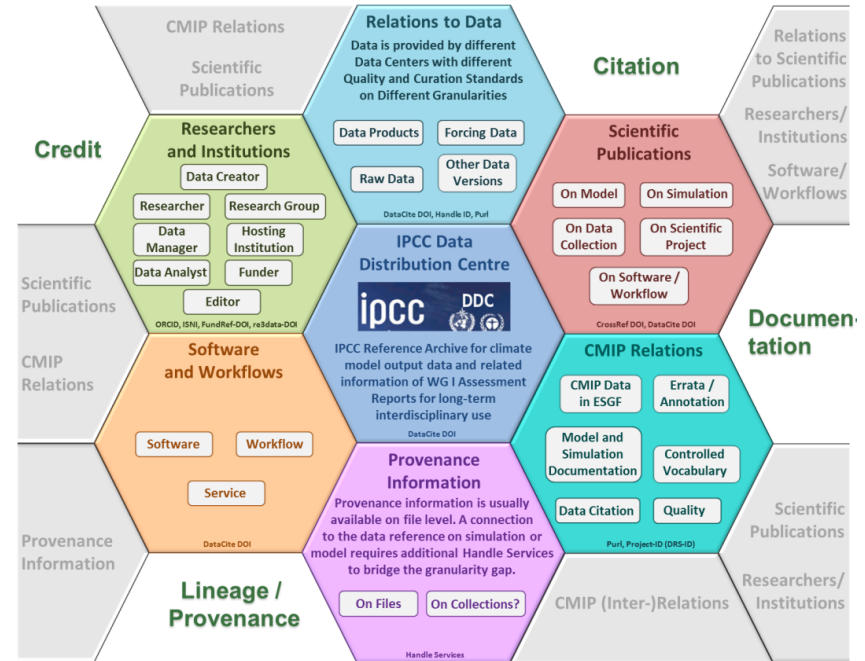
## Interlinking for CMIP6/AR6

Collection from data providers within citation service

- Recommendation to provide ORCID, paper references and references to forcing data DOI (input4MIPs) and obs4MIPs data DOI
- Possibility to provide 'Crossref Funder ID' for funders and data and software references

### Addition of object IDs by WDCC

- WDCC add data-data references
  - for 2 different CMIP6 data citation granularities
  - for IPCC DDC data citation on same granularity
- WDCC add subject reference to DRS\_id (CMIP naming convention)
- Reference to ES-DOC document in discussion: persistence unclear






# Summary of issues for CDS

## Summary of issues for CDS (similar to DDC issues):

- Long-term accessibility of ESGF data and provided data services incl. user support
- Server-side processing to reduce transfer volume and to provide additional derived datasets like climate parameters (OGC WPS)
- Documentation (e.g. ES-DOC, data citation incl. interlinking), quality of ESGF data:
  - Ancillary metadata URLs in ESGF index for machine access
- Curation: Report/add information on errata (ES-DOC errata service)
- Curation: Tracking of data usage in literature (Scholix) and interlinking to researchers via ORCID
- Data provenance of data underlying figures/tables (PID service)

stockhause@dkrz.de

 [orcid.org/0000-0001-6636-4972](https://orcid.org/0000-0001-6636-4972)

# CMIP6 vs. IPCC AR6 (4): CMIP6 citation service

## Evolving CMIP6 data gets citable (requested by WGCM-CMIP6)

### CMIP6 data citation concept:

- WDCC/DKRZ hosts the repository and the landing page
- ESGF provides the data access
- CMIP6 data citations use DataCite DOIs issued by WDCC on behalf of the ESGF; publisher is the ESGF.
- Stable citation reference with version information to identify data status and ESGF support filtering for version ranges: Authors (YYYY): Title. **Version YYYYMMDD**. Earth System Grid Federation. DOI.

Matthes, Katja; Funke, Bernd; Kruschke, Tim; Wahl, Sebastian (2017). input4MIPs.SOLARIS-HEPPA.solar.CMIP.SOLARIS-HEPPA-3-2. **Version 20170103**. Earth System Grid Federation. <https://doi.org/10.22033/ESGF/input4MIPs.1122>.

WCRP World Climate Research Programme

PCMDI

British Atmospheric Data Centre

WDC CLIMATE

DOI for Scientific and Technical Data 'input4MIPs.SOLARIS-HEPPA.solar.CMIP.SOLARIS-HEPPA-3-2'  
doi:10.22033/ESGF/input4MIPs.1122

General Information Creators Funders Relations

General Information

**Name** input4MIPs.SOLARIS-HEPPA.solar.CMIP.SOLARIS-HEPPA-3-2

**Abstract** CMIP6 Forcing Datasets (input4MIPs): The forcing datasets (and boundary conditions) needed for CMIP6 experiments are being prepared by a number of different experts. Initially many of these datasets may only be available from those experts, but over time as part of the 'input4MIPs' activity most of them will be archived by PCMDI and served by the Earth System Grid Federation (<https://esgf-node.llnl.gov/search/input4mips/>). More information is available in the living document: <http://goo.gl/Rup31>

**Subjects**

- climate
- CMIP6
- forcing data

**Rights** Creative Commons Attribution 4.0 International License (CC BY-SA 4.0)

**License** input4MIPs forcing data for CMIP6 is evolving in the sense that altered datasets might be added as new versions. The author list and the title are not final, either. Cite this data collection including the latest dataset version according to the Data Citation Guidelines (<http://bit.ly/2gBCuqM>). Individuals using the data must abide to the terms of use for CMIP6 data (<https://pcmdi.llnl.gov/CMIP6/TermsOfUse>). Details on any license restrictions are recorded as global attributes in the files.

**Contacts**

- Matthes, Katja
- Funke, Bernd
- Kruschke, Tim
- Wahl, Sebastian

**Funders**

- GEOMAR Helmholtz Centre for Ocean Research Kiel (GEOMAR)
- Instituto de Astrofísica de Andalucía (IAA)

Cite this data

**Citation** Matthes, Katja; Funke, Bernd; Kruschke, Tim; Wahl, Sebastian (2017). input4MIPs.SOLARIS-HEPPA.solar.CMIP.SOLARIS-HEPPA-3-2. Version YYYYMMDD<sup>[1]</sup>. Earth System Grid Federation. <http://doi.org/10.22033/ESGF/input4MIPs.1122>

[1] Please use the latest dataset version or if not available the latest data download date as version in your data citation.

**Data Access**

[https://esgf-data.dkrz.de/search/esgf-dkrz/?project=input4MIPs&institution=SOLARIS-HEPPA&dataset\\_category=solar&target\\_mip=CMIP6&source=SOLARIS-HEPPA-3-2](https://esgf-data.dkrz.de/search/esgf-dkrz/?project=input4MIPs&institution=SOLARIS-HEPPA&dataset_category=solar&target_mip=CMIP6&source=SOLARIS-HEPPA-3-2)

[https://esgf-node.llnl.gov/search/input4mips/?project=input4MIPs&institution=SOLARIS-HEPPA&dataset\\_category=solar&target\\_mip=CMIP6&source=SOLARIS-HEPPA-3-2](https://esgf-node.llnl.gov/search/input4mips/?project=input4MIPs&institution=SOLARIS-HEPPA&dataset_category=solar&target_mip=CMIP6&source=SOLARIS-HEPPA-3-2)

Metadata Export

XML JSON

# CMIP6 vs. IPCC AR6 (1): Organization

## CMIP6 – Coupled Model Intercomparison Project Phase 6



<https://www.wcrp-climate.org/wgcm-cmip/wgcm-cmip6>

WCRP CMIP6 coordinates the research project and has endorsed 21 MIPs (Model Intercomparison Projects).  
→ 21 groups coordinating individual MIPs

## WGCM (Working Group on Climate Models) Infrastructure Panel (WIP)

<https://www.earthsystemcog.org/projects/wip/>

WIP was established in 2014 for CMIP6 to coordinate the technical requirements for CMIP6.

## IPCC AR6 – Intergovernmental Panel on Climate Change 6<sup>th</sup> Assessment Report



<http://ipcc.ch>

IPCC assesses the status of climate research ca. every 6-7 years  
**based on CMIP results.**

→ 3 Working Groups and 2 Task Groups

## IPCC Data Distribution Centre (IPCC DDC) and DDC Support

<http://ipcc-data.org>

IPCC DDC is jointly managed by BADC (UK), CIESIN (US) and WDCC/DKRZ (Germany). It hosts data underlying the IPCC ARs on the long-term and supports the work of the IPCC WGs.

## Additional: Evolving CMIP6 Data Citation (2)

### Evolving nature of CMIP6 data = Adding versioned datasets:

- Phase 1 – Data collection built-up:  
In the first phase after initial ESGF publication, ensemble members will be added.
  - Phase 2 – Dataset corrections due to community review – CMIP6 data citation:  
Later few corrected variables are published as new dataset versions and the collection content stabilizes.
- The frequency of dataset additions as new versions decreases with time and differs largely among modeling centers and ESGF data nodes.