

# Data citation in climate sciences: Improvements in CMIP6 compared to CMIP5

AGU 2017  
New Orleans

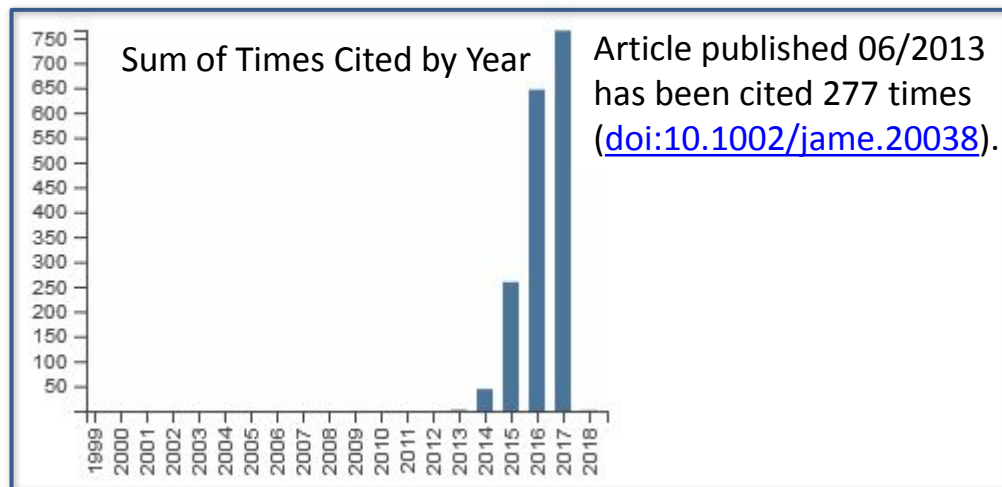
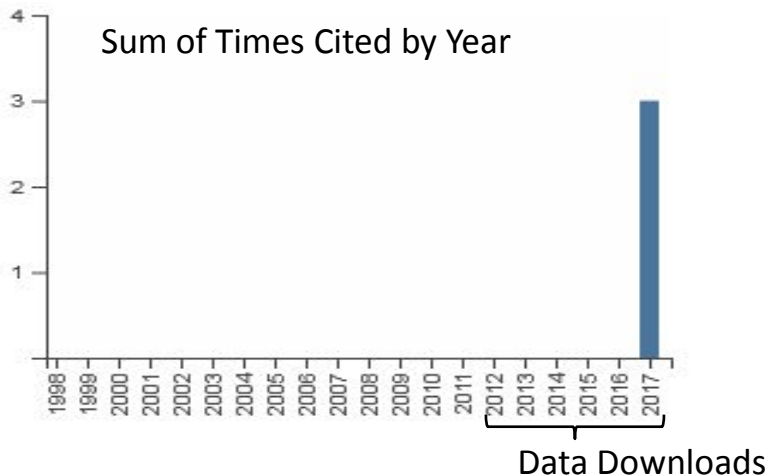
M. Stockhause, M. Lautenschlager  
WDC Climate at DKRZ

# Current Situation (1): Data Citations in Literature

AR5 Data Citation examples from Data Citation Index (Nov. 2017):

(1) IPCC AR5 WGI snapshot: no data citations in literature found

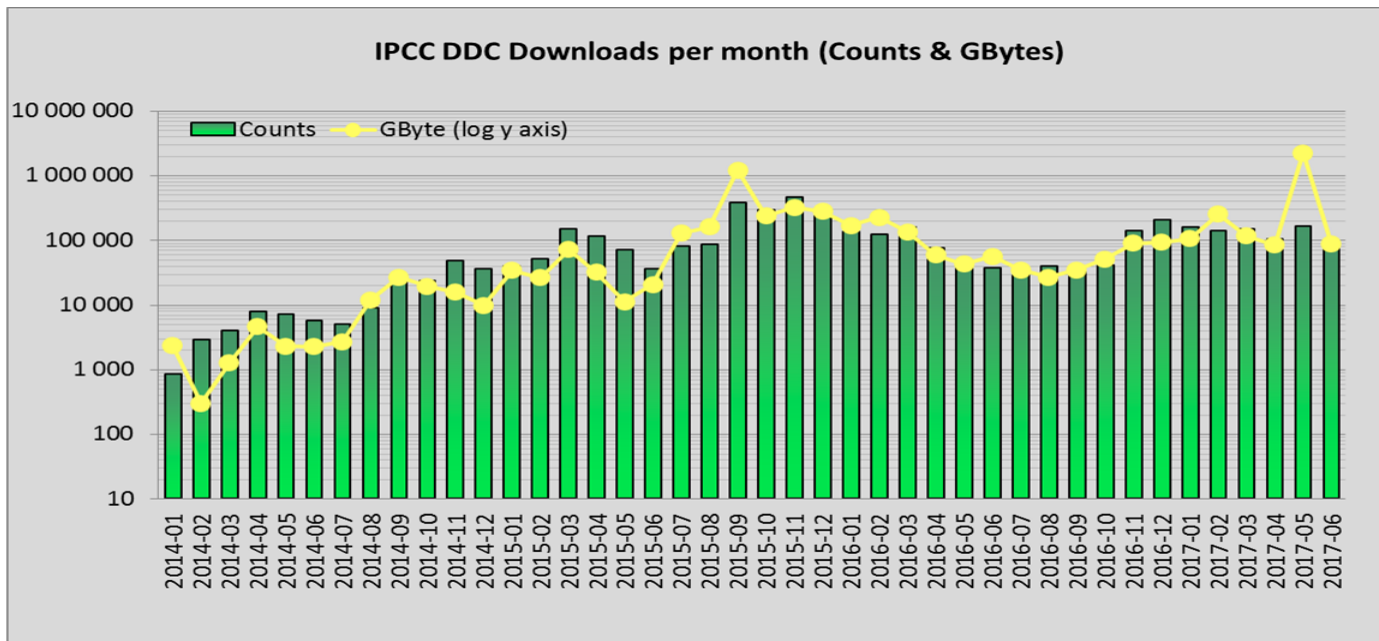
(2) MPI-M contribution to CMIP5: 4 data citations in 3 papers in 2017



→ Search for data reference in literature of an individual data publisher is not possible

# Current Situation (2): Download Rates

## AR5 Data Download Rates:



→ Download rates ~10-100 TBytes/month

→ Why is the data not cited?

## Current Situation (3): Barriers

- **Visibility and documentation how to cite data:**
  - data citation information only visible at IPCC DDC portal
  - documentation of how to cite on project pages is not clear: acknowledgement, citation,...
- **Disconnection between Data and Citation:**
  - data citation information is not documented in the file headers
  - data is downloaded by one person for a whole institution
- **Disconnection between data and paper publications (data comes first!):**
  - few references to papers provided by data creators (papers not yet published)
  - tracing of data usage in literature for data publishers is not supported
- **Citing data is not yet established in the scientific community:**
  - Scientists ask for credit for their data but refer to papers instead of data
  - different data reference models are used:  
cite in reference list, add as supplement, include in acknowledgements,...

# CMIP6 vs. IPCC AR6 (1): Organization

## CMIP6 – Coupled Model Intercomparison Project Phase 6



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

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

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

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

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 Data Distribution Centre (IPCC DDC) and DDC Support

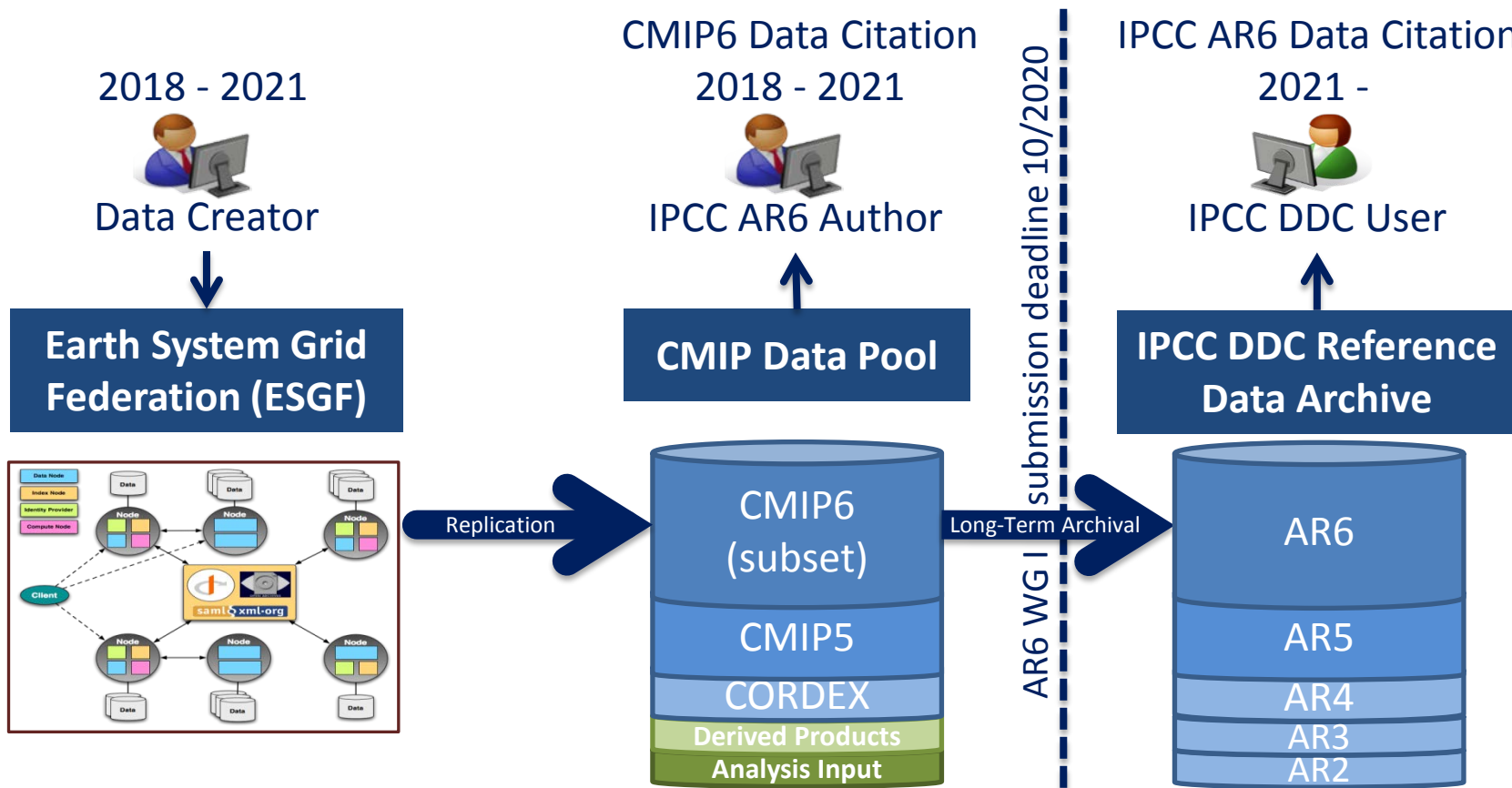
<http://ipcc-data.org>

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

→ 3 Working Groups and 2 Task Groups

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.

# CMIP6 vs. IPCC AR6 (2): Workflow/Timeline



# CMIP6 vs. IPCC AR6 (3): Numbers

## CMIP6 Estimations (2017-11-15):

- 21 MIPs+DECK with ca. 248 experiments are part of CMIP6
- 27 institutions with 79 models are currently registered for CMIP6 → registration process is ongoing
- Data volume: ca. 20 PBytes
- ca. 25 ESGF data nodes world-wide will disseminate data

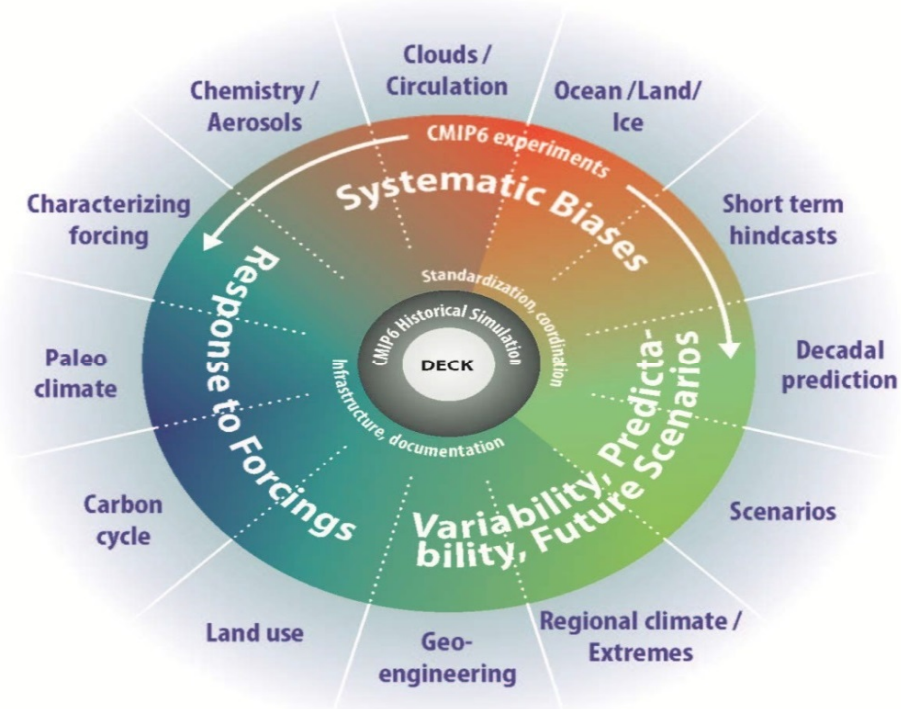
## Infrastructure Components:

- Data dissemination: ESGF
- Ancillary metadata:
  - ES-DOC: documentation of models and simulations

Citation Service  
Errata Service

**New in CMIP6**

- Further Information on web pages of modeling centers



(Source: Eyring, V. et al. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937-1958, [doi:10.5194/gmd-9-1937-2016](https://doi.org/10.5194/gmd-9-1937-2016).)

# 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.

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://gon-gli/fwp31>

**Subjects**

- climate
- CMIP6
- forcing data

**input4MIPs.SOLARIS-HEPPA.solar.CMIP.SOLARIS-HEPPA-3-2 (DKS)**

**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/2gBCuM>). Individuals using the data must abide to the terms of use for CMIP6 data (<https://gom6.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. 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 Report**

DOI ESGF

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>.



# Improvements in CMIP6/AR6 (1)

## CMIP5 / AR5:

- Data Citation:
  - CMIP5 data are not citable
  - IPCC DDC AR5 data are citable  
→ to late for many publications
- Get credit:
  - Authors and contributors without ORCID or other IDs
- Give credit
  - No Data References
  - Paper References
- Get usage statistics
  - Downloads
  - No tracking of data citations in literature

## CMIP6 / AR6:

- Data Citation:
  - **CMIP6 data are citable**
  - IPCC DDC AR6 data are citable  
→ incl. references to CMIP6 data
- Get credit:
  - Authors and contributors with **ORCID** or other IDs
- Give credit
  - **Data References**
  - Paper References
- Get usage statistics
  - Downloads
  - Use of Scholix services planned to provide information on **data usage in literature**
- **Provenance (DDC Support group):**
  - Storage of data underlying figures and tables plus analysis script planned

# Improvements in CMIP6/AR6 (2)

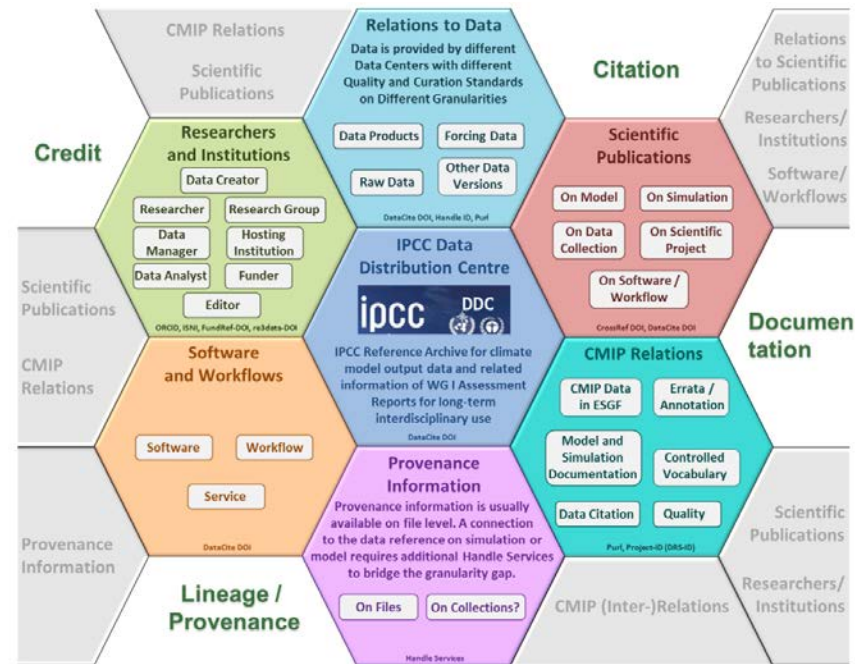
## Interlinking for CMIP6/AR6

### Collection from data providers

- 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 (1)

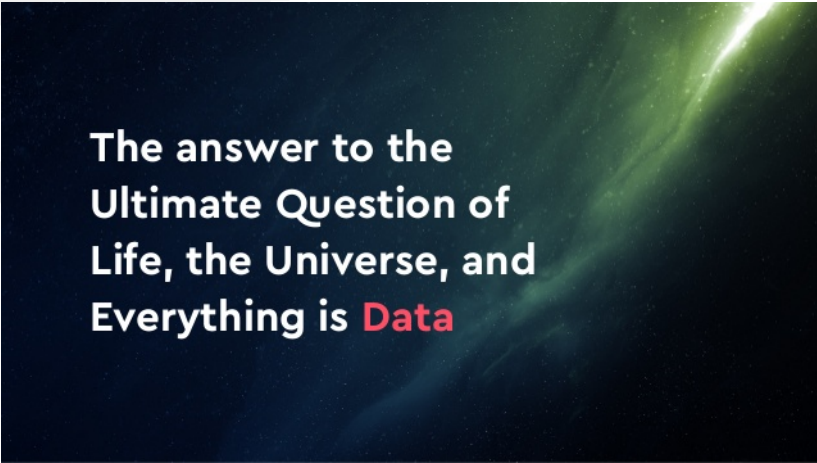
## 1. CMIP6 / AR6:

- Evolving CMIP6 data gets citable
  - credit for creators and contributors, incl. funders
- Connection between Files and Citation on data collections
  - via further\_info\_URL page hosted by ES-DOC
- Provenance for data underlying figures/tables in AR6 (WGI)
  - DDC Support group
  - storage of CMIP6 source data, analysis/results planned
- Statistics on data usage in literature
  - Scholix link exchange services

## Summary (2)

### 2. CMIP7 / AR7:


- ESGF as DataCite member
  - getting from a service at DKRZ for international projects to a federated structure of client data DOI publishers
- Implementation of RDA WGDC recommendation
  - referenceable **user-defined data subsets underlying a paper**



The answer to the  
Ultimate Question of  
Life, the Universe, and  
Everything is **Data**

[cmip6cite.wdc-climate.de](http://cmip6cite.wdc-climate.de)

stockhause@dkrz.de

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

Stockhause, M., and Lautenschlager, M., (2017). CMIP6 Data Citation of Evolving Data. Data Science Journal. 16, p.30. doi:10.5334/dsj-2017-030. <http://doi.org/10.5334/dsj-2017-030>.

Stockhause, M., and Lautenschlager, M. (2017). CMIP6 Data Citation and IPCC Data Distribution Centre Services. PICO at EGU 2017, Vienna. Zenodo. <http://doi.org/10.5281/zenodo.569646>.