IPCC Data Distribution Centre (IPCC DDC) AR6

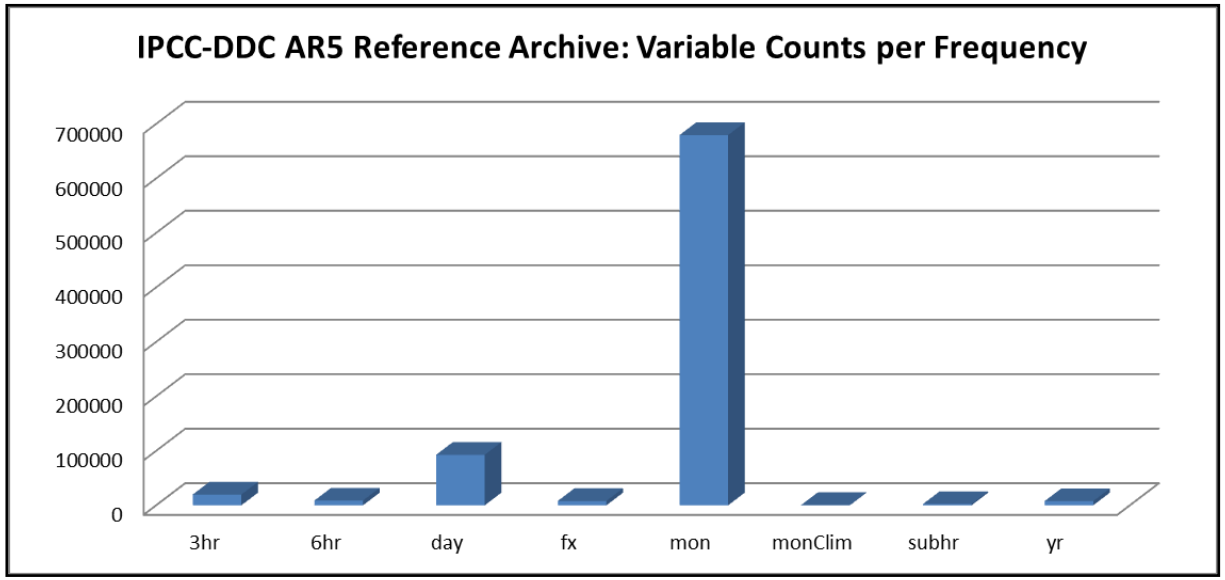
ESGF Conference 2016

M. Stockhause, M. Lautenschlager, S. Kindermann
Deutsches Klimarechenzentrum (DKRZ)

IPCC DDC AR5 (1)

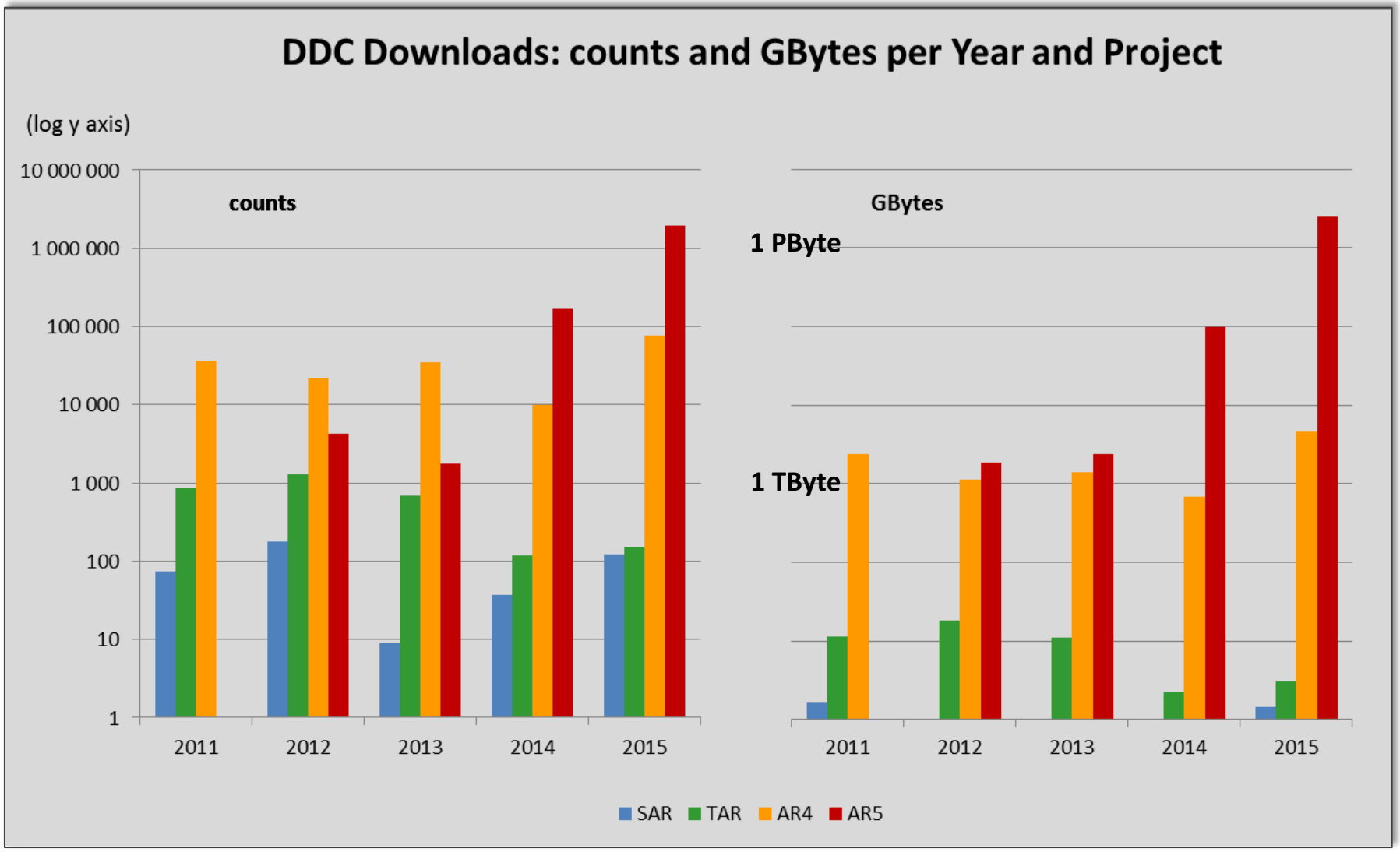
DDC Reference Archive AR5 contents / IPCC WG1 Archive AR5 contents

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 AR5 (2)

Reference Archive for Climate Model Output Data – Download Statistics



IPCC DDC at WDCC / DKRZ

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

jointly managed by:

- British Atmospheric Data Centre (BADC)
- World Data Center Climate (WDCC) at DKRZ:
- Center for International Earth Science Information Network (CIESIN) at Columbia University:

→ Certified ICSU World Data System (WDS) members

→ Overseen by IPCC TGICA

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

IPCC DDC at WDCC / DKRZ

Reference Archive for Climate Model Output Data

- 1995: LTA for IPCC climate model data since SAR
- 2008: parts of FAR added to DDC
- 2013/14: IPCC DDC AR5 long-term archival
- 2020/21: IPCC DDC AR6 long-term archival

IPCC AR6: Timeframe

Updated timeframe for IPCC AR6 (version: 1 December 2016) (http://ipcc.ch/activities/pdf/ar6_schedule.pdf):

- 05/2017: AR6 Scoping Meeting
- 09/2017: IPCC approval of AR6 outline
- **02/2018: Decision on selection of WG I authors**
- 05/2019: WG I AR6 first-order draft expert review
- **03/2020: WG I AR6 second-order draft expert review**
- 04/2021: WG I AR6 IPCC acceptance/adoption/approval at IPCC-53

IPCC DDC: Reference Data Archive

The IPCC DDC provides data on the long-term for an interdisciplinary user community and 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

IPCC DDC: Improvements for AR6 (1)

IPCC Author Support – CMIP Data Pool:

IPCC WGs request a reliable, up-to-date and easily-accessible CMIP6 data pool which supports script-based analyses

→ IPCC DDC plans to open its CMIP data pool for IPCC authors:

- Idea presented at "IPCC Expert Meeting on the future of TGICA" (Geneva, 01/2016 - IPCC-XLIII/Doc. 10, Corr.1)
- Information and Draft Concept available at https://redmine.dkrz.de/projects/dkrz_cdp/wiki
- Coordination between IPCC WG and IPCC DDC on Data Pool to be organized

→ CMIP6 subset built based on user (IPCC author) requests

→ IPCC DDC AR6 Reference Data Archive will consist of a single data collection based on the Data Pool

A better integration of IPCC DDC / TGICA into IPCC processes is needed:

- Ongoing IPCC Review

IPCC DDC: Improvements for AR6 (2)

Data Management Aspects:

- Reduce number of interfaces to handle
- Stable and reliable technical interfaces providing the same information over time
- No mapping of DRS components
- CMIP6 infrastructure provides all necessary information for LTA / IPCC DDC
- Shift responsibility for relations to ancillary metadata from LTA to providers

→ **Compared to AR5:**
simpler and automatable LTA
with reallocated responsibilities

IPCC DDC Services for AR6

Data Creator

IPCC Author

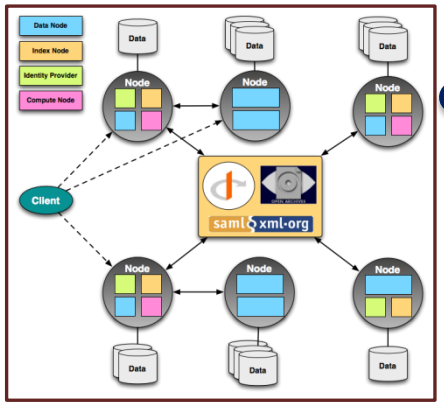
IPCC DDC User

IPCC DDC AR6 Services

Earth System Grid Federation (ESGF)

CMIP Data Pool at DKRZ

IPCC DDC Reference Data Archive at DKRZ



Replication

CMIP6 (subset)

CMIP5

CORDEX

Derived Products

Analysis Input

Long-Term Archival

AR6

AR5

AR4

AR3

AR2

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

→ Use CMIP Data Pool for CMIP6 to build the Reference Data Archive AR6

→ Stephan's Talk

Challenges of the Long-Term Archival (1)

Gather what you can…

…as long as it is available.

LTA challenges:

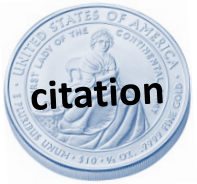
Ancillary metadata is diverse in respect to:

- Granularity,
- Format,
- Access,
- Stability

ESGF Search API used for discovery:

Ancillary metadata URLs registered in ESGF are used to relate it to the data.

→ **ESGF registration of ancillary metadata URL implemented within the ESGF-QCWT**



Challenges of the 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.

- Description of ancillary metadata file
  → see living doc on specification at: http://bit.ly/1XsVOoz
- Specification of mapping to local database

Mapping

Challenges of the Long-Term Archival (3)

Check everything at least twice…

…before archival.

The automated process is interrupted at several stages in order to ensure the valid mapping and consistency of the metadata. After archival, such a **quality assurance (QA)** is unfeasible.
A final QA is applied prior to DataCite DOI assignment.

Long-Term Archival Workflow

LTA Workflow

1. Access and map detailed ESGF metadata
2. Start of Data Archival from the CMIP Data Pool
3. Gather ancillary metadata URLs via ESGF Search API
4. Download ancillary metadata
5. Interpret and map ancillary metadata content
6. Final quality assurance
   after completion of LTA and metadata enrichment
7. DataCite DOI assignment making the data citable

IPCC DDC: http://ipcc-data.org

DDC at DKRZ: http://ipcc.wdc-climate.de

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