

# IPCC-DDC at WDC Climate / DKRZ

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

# IPCC-DDC at WDCC / DKRZ (1)

## World Data Center for Climate

### Long-Term Archive for Climate Data

- 1992: Long-term archive for climate data
- 2003: regular member of the ICSU World Data System, 2011 renewed ICSU WDS membership/certification
- 2010: WDCC moved to Deutsches Klimarechenzentrum



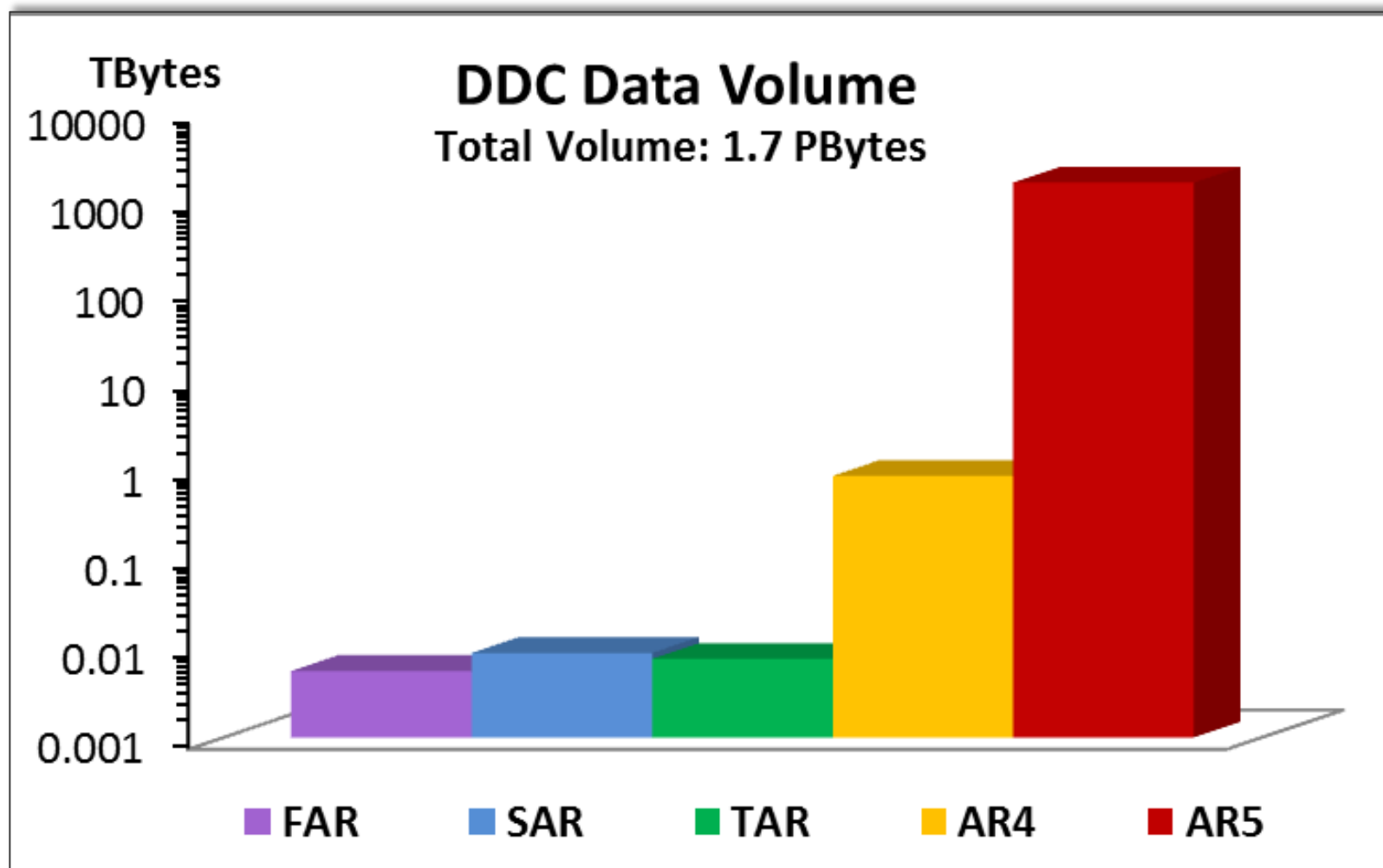
## IPCC DDC at WDCC

### Reference Archive for Climate Model Output Data

- 1995: LTA for IPCC climate model data since SAR
- 1997: IPCC DDC repository
- 2008: parts of FAR added to DDC
- 2013/14: LTA of IPCC AR5

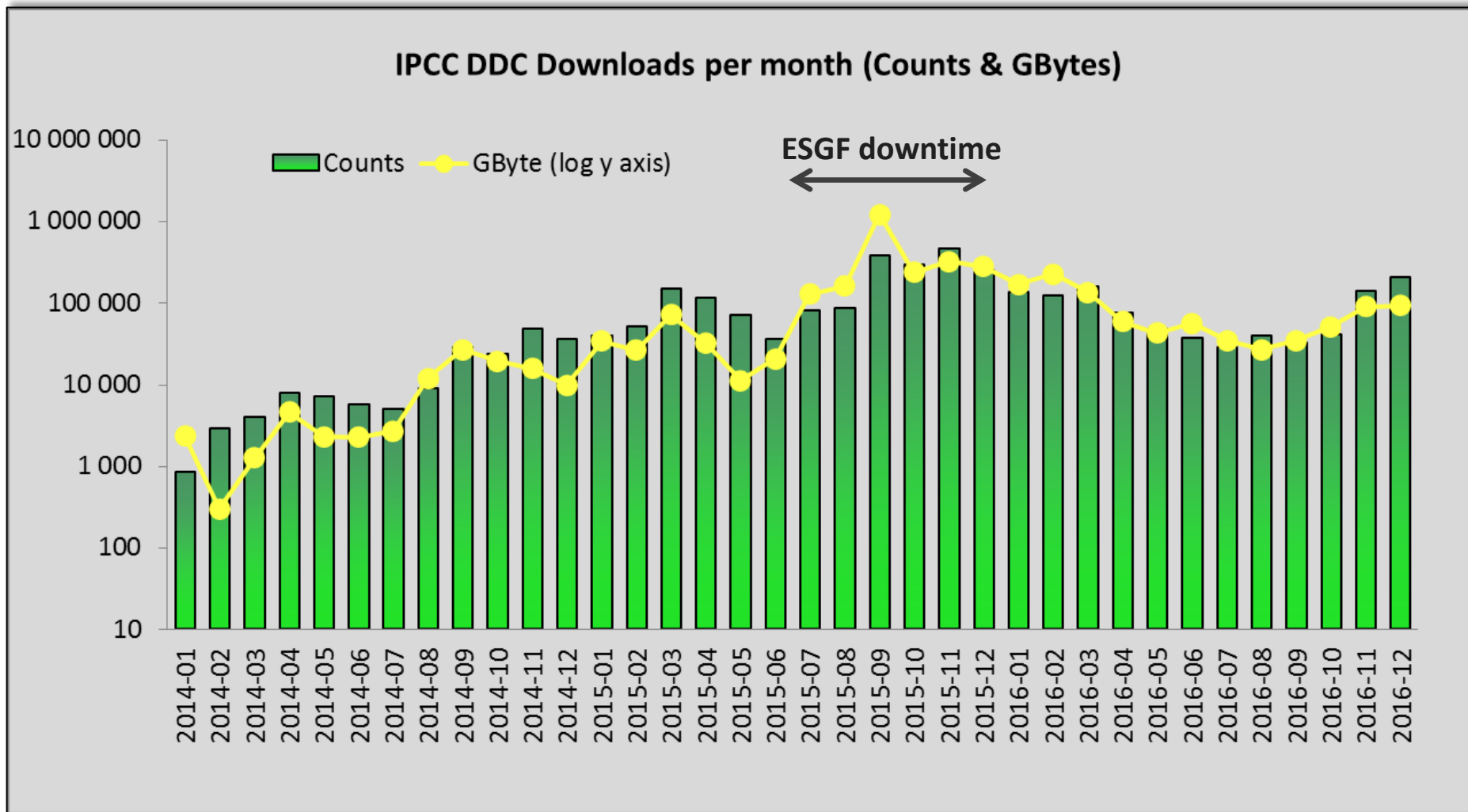
# Reference Data Archive (1)

## Data Content



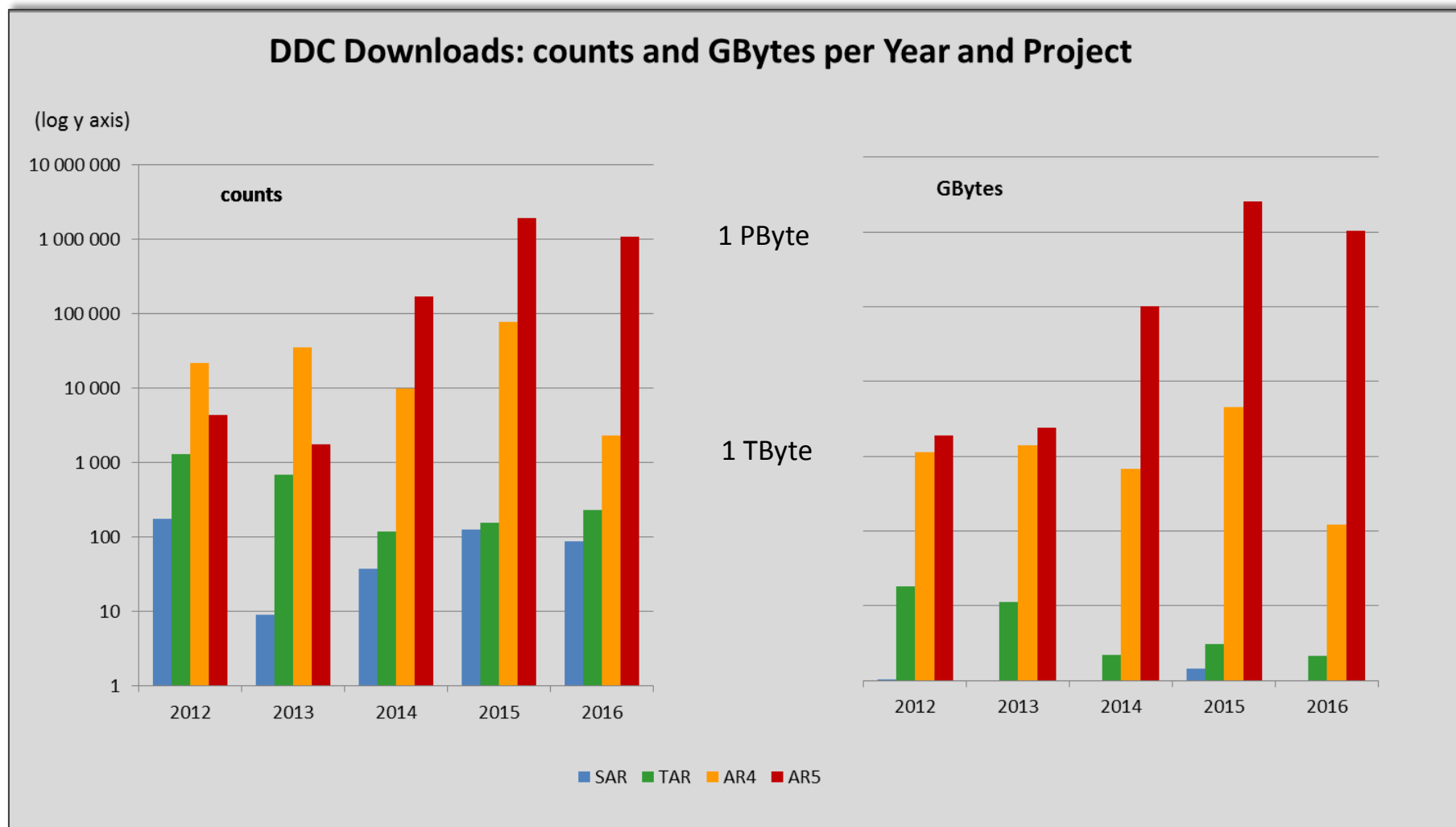
# Reference Data Archive (2)

## Monthly Data Download Rates



# Reference Data Archive (3)

## Annual Data Download Rates per AR



# IPCC-DDC for AR6 (1)

**IPCC DDC AR5** consists of two data collections (snapshots 15.03.2013):

- DDC Reference snapshot (transferred from ESGF into LTA)
- IPCC Working Group I snapshot (collected by the ETH Zurich)

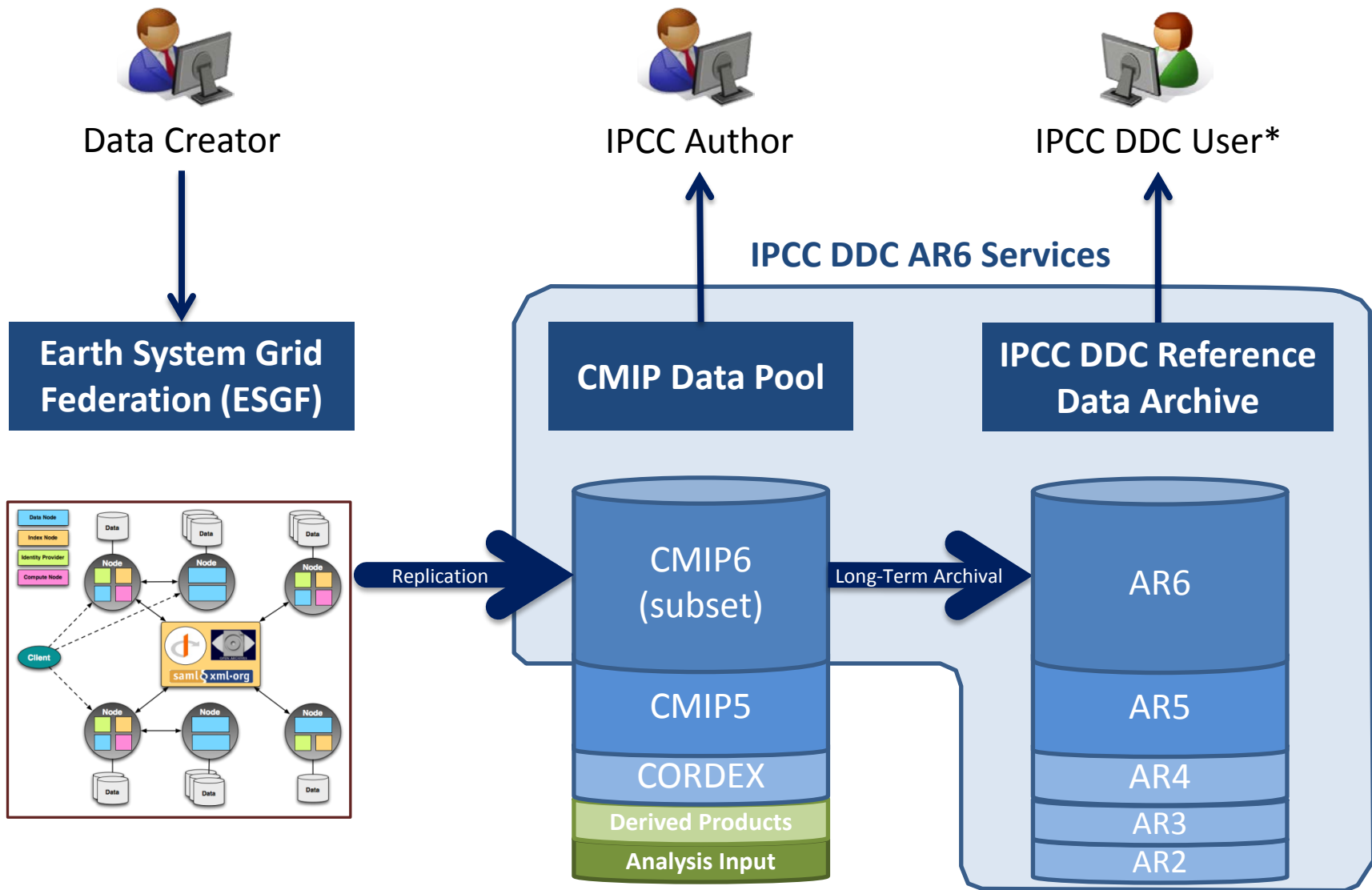


**IPCC DDC AR6** should consist of one data collection:  
IPCC AR6 Reference Data Archive

**IPCC DDC support of WGs:**  
CMIP6 Data Pool with subset of latest dataset versions (similar to ETH Zurich for AR5) and access to DKRZ's HPC

The screenshot shows the IPCC Data Distribution Centre website. The header includes the IPCC logo and the text 'INTERGOVERNMENTAL PANEL ON climate change'. Below the header, there are search options and a navigation menu. The main content area is titled 'AR5 GCM data' and contains text about climate model results and a list of bullet points: (i) the IPCC Working Group I snapshot, and (ii) the DDC Reference snapshot.

# IPCC-DDC for AR6 (2)



\* Data access via WDCC and ESGF portals

# IPCC-DDC for AR6 (3)

CMIP6 Data Pool was requested by our national users and became part of the national CMIP6 project.

→ Open CMIP6 Data Pool for IPCC Authors

## Contacts between AR6 WGs and DDC at DKRZ

- Discussions between DDC Manager and WG I and II Co-Chairs have been started in **01/2016** at the IPCC EM on the future of TGICA  
→ data pool idea mentioned with positive feedback
- Continued during coffee breaks at a Future Earth – PROVIA – IPCC workshop **09/2016**  
→ data pool plan: contacts for TSUs of WG I and II given
- Short contacts at IPCC-DE Annual Meeting with WG II and at EGU with WG I AR6  
→ **06/2017**: initial list of variables for CMIP6 data pool sent (based on requests from national project partners)



# The DKRZ CMIP Data Pool (1)

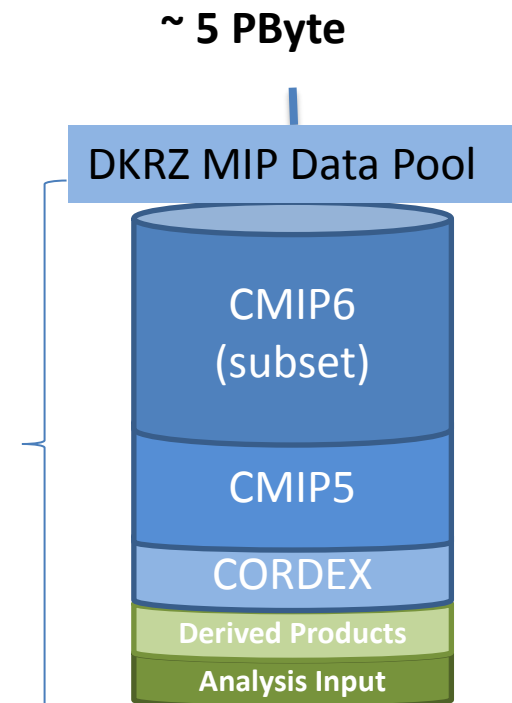
## Which data is in the pool ?

- Data often required in MIP analysis activities
- Core CMIP data collections for ESGF based data distribution (national + replicated)
- An advisory board decides on priorities

# The DKRZ CMIP Data Pool (2)

## Which data is in the pool ?

- ~1 PByte German CMIP6 contributions (first estimate)
- ~2 PByte CMIP6 replicas (**content defined by European replication strategy**)
- ~600 TByte CMIP5 data (now > 1.3 PByte, will be removed)
- ~400 TByte data for analysis and evaluation
- ~ 1 PByte, data management cache



# Access the DKRZ CMIP Data Pool (1)

## How to access ?

### ■ Direct posix file access

- From any HPC (Mistral) node
- On request from dedicated VMs
- Path: /work/kd0956/...
  - e.g. in /work/kd0956/CMIP5/data/cmip5 a large portion of the overall CMIP5 data collection is accessible )

```
bash-4.1$ tree -d
/work/kd0956/CMIP5/data/cmip5/output1 | more
/work/kd0956/CMIP5/data/cmip5/output1
├── BCC
│   ├── bcc-csm1-1
│   │   ├── 1pctCO2
│   │   │   ├── day
│   │   │   │   ├── atmos
│   │   │   │   └── day
│   │   │   │       ├── r1i1p1
│   │   │   │       └── v1
│   │   │   │           ├── pr
│   │   │   │           ├── psl
│   │   │   │           ├── tas
│   │   │   │           ├── tasmax
│   │   │   │           └── tasmin
│   │   └── mon
│   │       ├── atmos
│   │       └── Amon
│   │           ├── r1i1p1
│   │           └── v1
│   │               ├── ccb
│   │               ├── cct
│   │               ├── ci
│   │               ├── cl
│   │               ├── cli
│   │               ├── clivi
│   │               └── ct
```

# Access the DKRZ CMIP Data Pool (2)

## How to access ?

ESGF published data in the data pool can be accessed also via the EGF interface

### ESGF Node at DKRZ

The screenshot shows the ESGF Node at DKRZ interface. On the left, there is a sidebar with various filters: Project, Product, Institute, Model, Experiment, Experiment Family, Time Frequency, Realm, CMIP Table, Ensemble, Variable, Variable Long Name, CF Standard Name, Driving Model, and Datanode. The 'Datanode' filter is selected, showing 'esgf1.dkrz.de (19299)'. On the right, there is a search bar with 'Enter Text:' and a search button. Below the search bar, there are search constraints: 'esgf1.dkrz.de'. There are also checkboxes for 'Show All Replicas', 'Show All Versions', and 'Search Local Node Only (Including All Re...)', which is circled in red. Below the search bar, there is a 'Display' dropdown set to '10' results per page. The search results show a total of 19299 results. The first two results are listed, showing project, model, and experiment details.

- Select „Search Local Node Only“
- Select data node esgf1.dkrz.de for CMIP5/CORDEX
- Select data node esgf3.dkrz.de for CMIP6 (coming)
- (esgf2.dkrz.de only serves data from the WDCC – thus stored on tape and not in the datapool)

# Access the DKRZ CMIP Data Pool (3)

How to access ?

Use the Freva interface  
at <http://cmip-eval.dkrz.de>

The screenshot shows the 'Data-Browser' interface of the CMIP Data Pool. The header includes the CMIP logo (The World Climate Research Programme's Coordinated Multi-Model Project), the ESMValTool logo (Earth System Model Evaluation Tool), and the Freva logo (by freva). Navigation links include Result-Browser, Data-Browser, Help, Contact, and Terms of Use. A 'logout (guest)' link is also present.

The main content area is titled 'Data-Browser' and features a 'Clear all' button. Below this, several filter categories are displayed with their respective counts:

- project: **cmip5**
- product: **output1**
- institute: **bnu** (with a globe icon)
- model: **bnu-esm**
- experiment (16)
- time\_frequency (6)

The 'time\_frequency' filter is expanded, showing a search input field and a grid of options:

Search time_frequency name			
3hr [434]	6hr [85]	day [492]	fx [14]
mon [2085]	yr [78]		

Additional filter categories shown below are 'realm (7)' and 'variable (173)'.

# Access the DKRZ CMIP Data Pool (4)

## How to access ?

Outlook (currently in testing phase):

Use the DKRZ jupyterhub at <http://jupyterhub.dkrz.de> to start interactive notebooks with direct access to the data pool.

# Access the DKRZ CMIP Data Pool (5)

## Four options / services

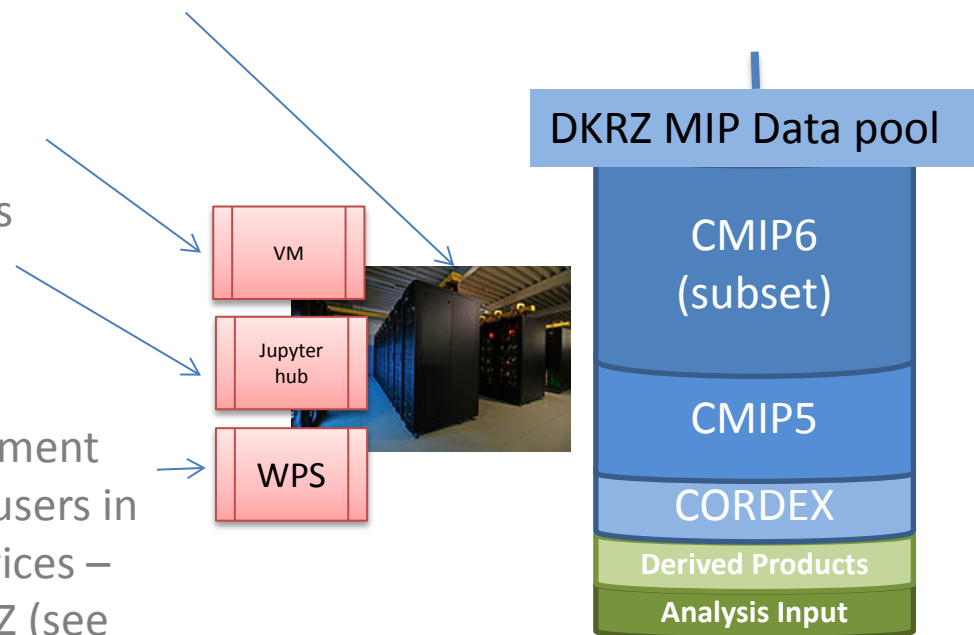
A) User logs into DKRZ cluster and processes data

B) User rents a VM with data mounted

C) User uses jupyterhub interactive notebooks to spawn processing services

D) User accesses web processing services

- see [mouflon.dkrz.de](http://mouflon.dkrz.de) for a test deployment
- The birdhouse framework<sup>1)</sup> supports users in building and deploying own WPS services – which then can be e.g. hosted at DKRZ (see option B. ) or in cloud or ..



Options C) and D) are currently in testing / ramp up phase

<sup>1)</sup> <http://bird-house.github.io/>

# Insert into the DKRZ CMIP Data Pool

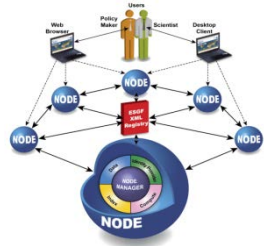
## How to get data into the pool (data ingest) ?



User specifies request using online forms in IPython notebook ([‘ESGF replication’ form](#)) or directly contacts [esgf-replication@dkrz.de](mailto:esgf-replication@dkrz.de)

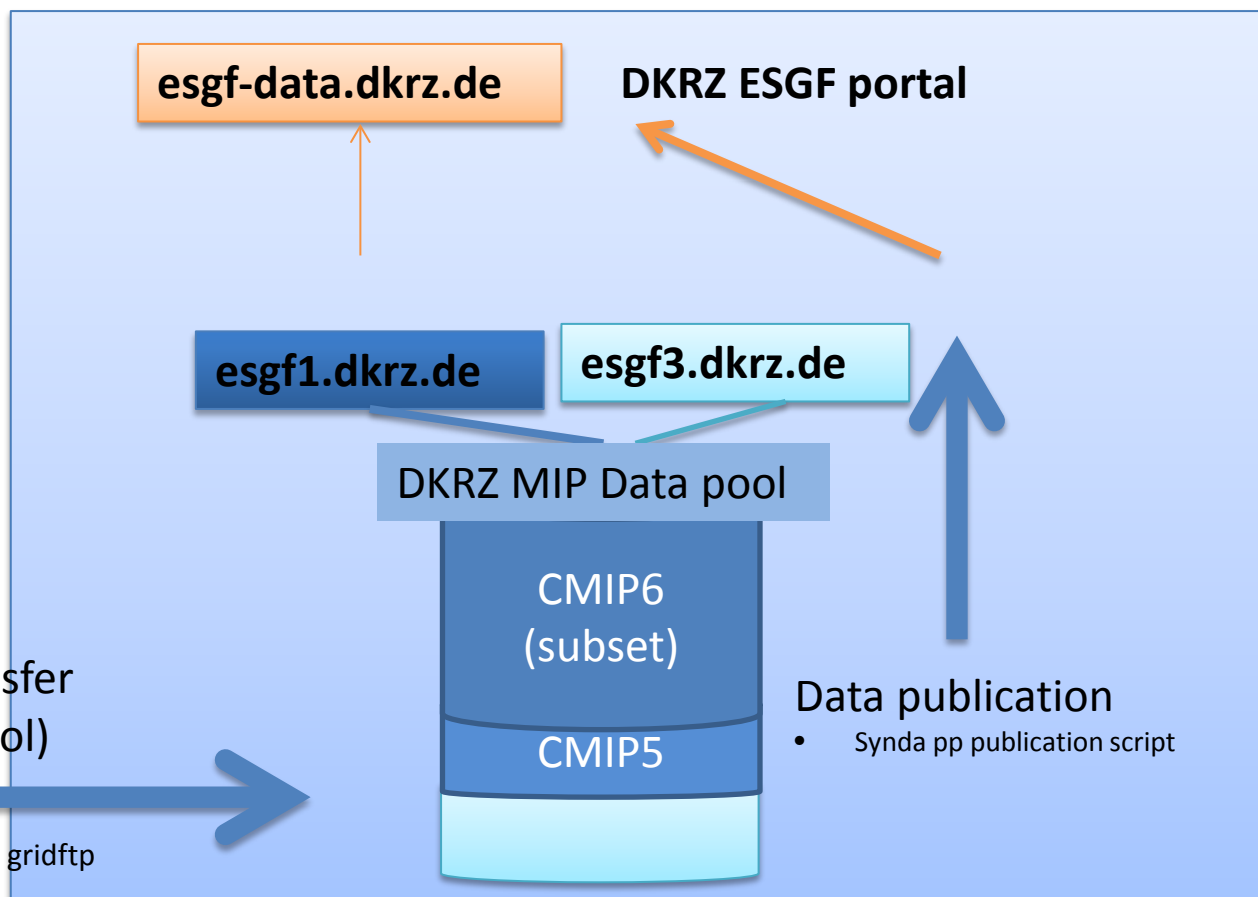
↓  
synda selection files

```
project=CMIP5
model=CNRM-CM5 CSIRO-Mk3-6-0
experiment=historical amip
ensemble=r1i1p1
variable[atmos][mon]=tasmin tas psl
variable[ocean][fx]=areacello sftof
variable[land][mon]=mrsos nppRoot nep
```



Data transfer  
(synda tool)

http, gridftp





# Synchronize with the DKRZ CMIP Data Pool

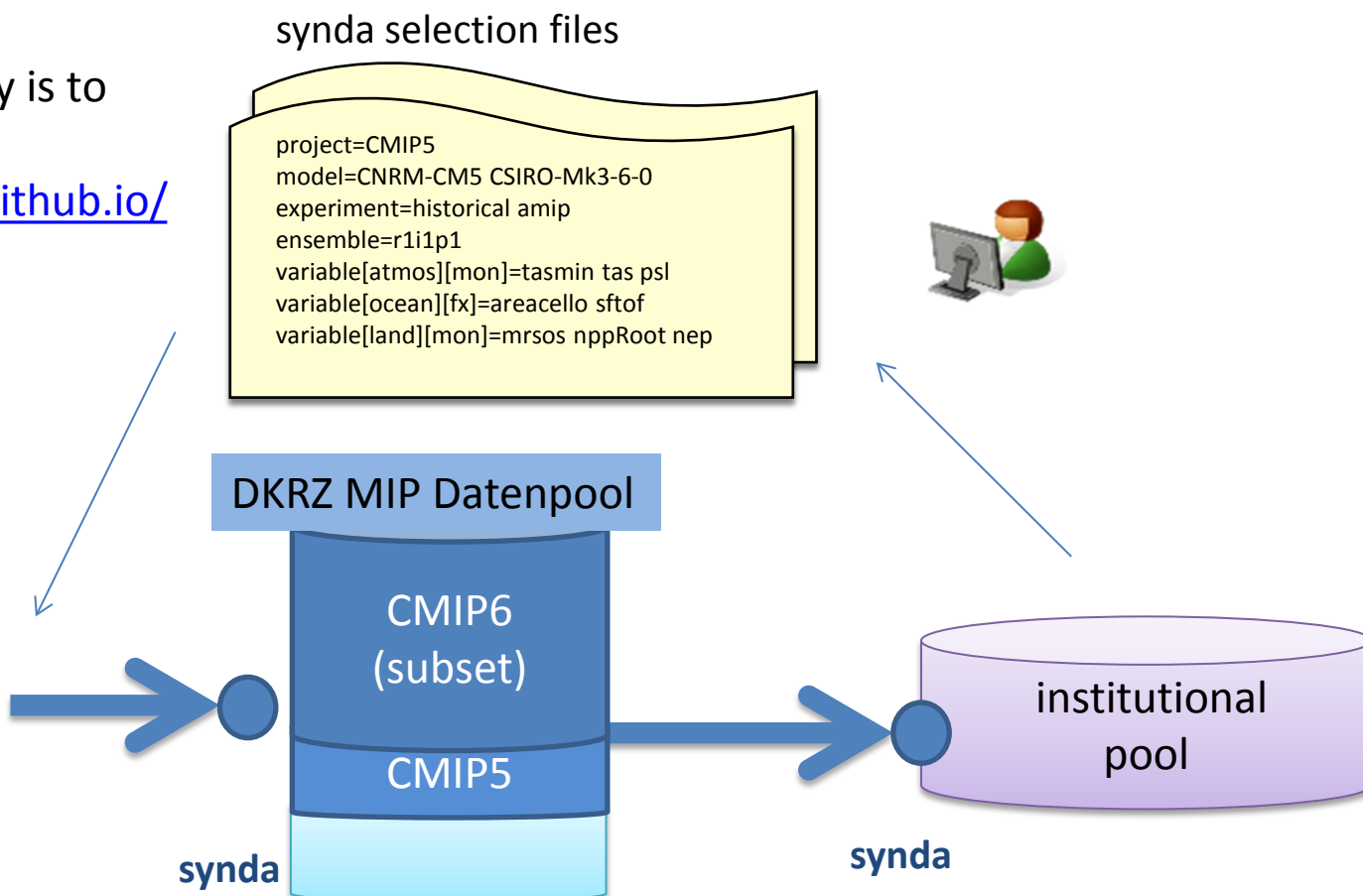
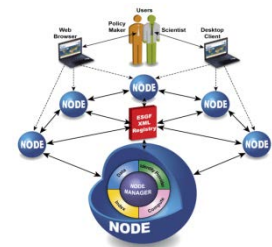
## How to synchronize a local pool with the DKRZ CMIP data pool ?

Recommended way is to use the synda tool

(<http://prodiguer.github.io/synda/>)

synda selection files

```
project=CMIP5
model=CNRM-CM5 CSIRO-Mk3-6-0
experiment=historical amip
ensemble=r1i1p1
variable[atmos][mon]=tasmin tas psl
variable[ocean][fx]=areacello sftof
variable[land][mon]=mrsos nppRoot nep
```



IPCC-DDC: [http://www.ipcc-data.org/sim/gcm\\_monthly/](http://www.ipcc-data.org/sim/gcm_monthly/)

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

Data Pool Request: <https://data-forms.dkrz.de:8080/login>

Data Pool Portal: <http://cmip-eval.dkrz.de/>