# ICSU World Data Center Climate at DKRZ
## DKRZ - DATA INTENSIVE CLIMATE SCIENCE

The World Data Center for Climate (WDCC) at DKRZ, that is a member of the ICSU World Data System (WDS), provides a long-term archive with approximately 500 TeraByte fully documented climate data. Users are provided with a flexible search interface and the possibility for field based data access and server side data processing (sub-setting, format conversion).

ICSU WORLD DATA SYSTEM

DKRZ archive over time
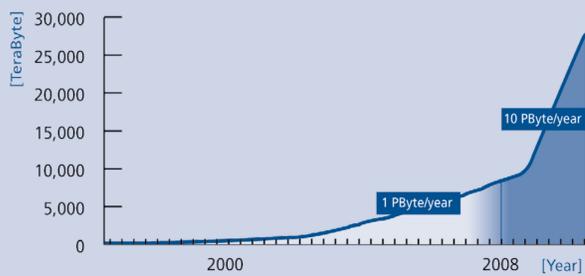
Exponential data growth at DKRZ. The DKRZ tape archive is designed for a growth of 10 PetaByte/year.

WDCC size over time

[Size in TeraByte]
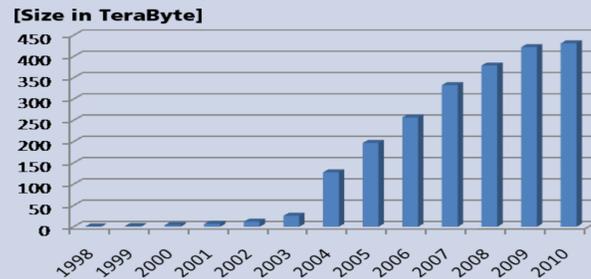
WDCC has grown by considerably over the last years and provides now access to more than 500 TeraByte of highly granular stored data. The WDCC is designed for a growth of 1 PetaByte /year.

IPCC data volume
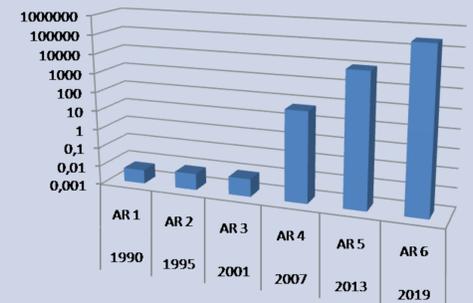
Data volumes from IPCC have increased by a factor of 30 by each report. For the current AR5 report a volume of 10 PetaByte will be produced. For the next report in 2019 a volume of 300 PetaByte is expected.

WDCC: Users

- 12% Campus
- 25% Germany
- 27% Europe
- 4% AF+OC+SA
- 13% North America
- 64% Asia
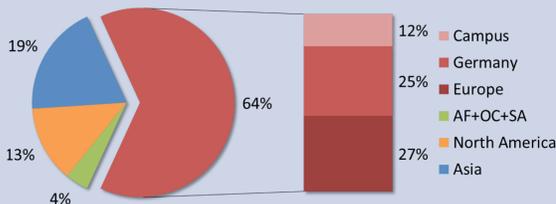
Data is disseminated to users worldwide. Only a minority of users is located on the campus.

Challenges

Climate model simulations produce high volume multidimensional data products not optimised for direct end-user access.

Data and descriptive metadata have to rearranged and made accessible supporting various usage scenarios and access patterns.

Thus the WDCC is based on a flexible metadata model as well as a generic data container system, to be able to efficiently serve data products ranging from few kiloBytes to hundreds of Gigabytes.

WDCC: Downloads

- Europa: 55,53%
- Nordamerika: 33,25%
- Asien: 10,73%
- Südamerika: 0,39%
- Afrika: 0,04%
- Ozeanien: 0,02%
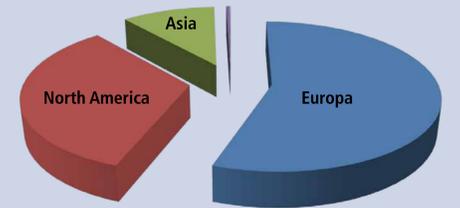
Downloads from Asia and North America represent a large part of the load.

Blades and disks

- 1.2 PetaByte storage system of DDN
- 10 Dell M605 Blades
- two AMD Opteron 4-Core 2.9 GHz
- ~32 or 64 GiagByte main memory
- 1 Dell M605 blade with two AMD Opteron
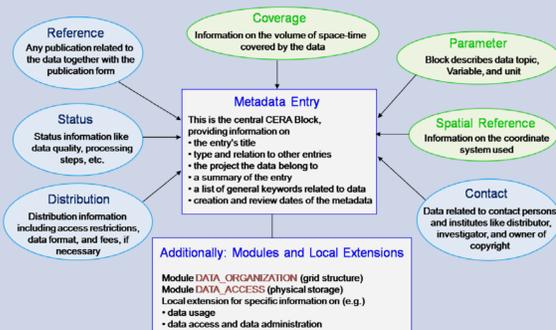
WDCC architecture and hardware

The WDCC is based on a complex metadata model (CERA), which is managed by an Oracle database system. Data is aggregated in container files, which are managed by so called LOB (large object) servers. These servers handle the staging of containers and (in cooperation with the CERA databases) support fine grained data access patterns.

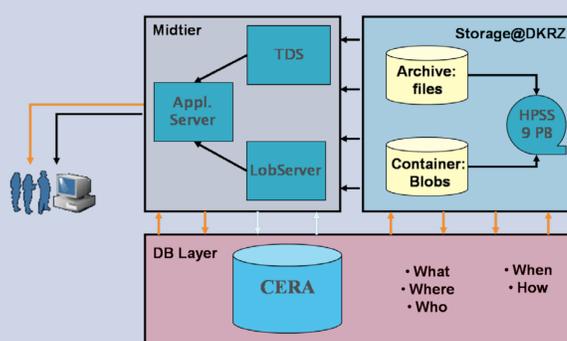## HPSS - High perfomance storage system

- 7 automated Oracle/StorageTek SL8500 tape libraries with more than 67 000 slots for magnetic tape cassettes
- 8 robots per library
- 88 tape drives
- more than 100 PetaByte total capacity
- bidirectional bandwidth of 5 GigaByte/s (peak)

CERA

Reference Any publication related to the data together with the publication form

Coverage Information on the volume of space-time covered by the data

Parameter Block describes data topic, Variable, and unit

Status Status information like data quality, processing steps, etc.

Metadata Entry This is the central CERA Block, providing information on
- the entry's title
- type and relation to other entries
- the project the data belong to
- a summary of the entry
- a list of general keywords related to data
- creation and review dates of the metadata

Spatial Reference Information on the coordinate system used

Distribution Distribution information including access restrictions, data format, and fees, if necessary

Contact Data related to contact persons and institutes like distributor, investigator, and owner of copyright

Additionally: Modules and Local Extensions
Module DATA_ORGANIZATION (grid structure)
Module DATA_ACCESS (physical storage)
Local extension for specific information on (e.g.)
- data usage
- data access and data administration

Data access

Midtier — TDS, Appl. Server, LobServer

Storage@DKRZ — Archive: files, Container: Blobs, HPSS 9 PB

DB Layer — CERA
- What
- Where
- Who
- When
- How

Container

Oracle RDB server

Application

Lobster object manager
- show-container
- read-record
- fetch-records

Cache

www.dkrz.de

Poster 3

DKRZ DEUTSCHES KLIMARECHENZENTRUM