



Data and Metadata Infrastructures at the World Data Centre for Climate (WDCC)



IN13A-1090

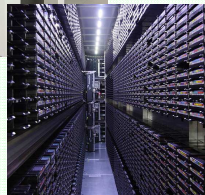
The World Data Centre for Climate is hosted by Deutsches Klimarechenzentrum GmbH in Hamburg, Germany

Data Archive and HPC Cluster – the Technical Basis

HLRE2 Data Archive: HPSS

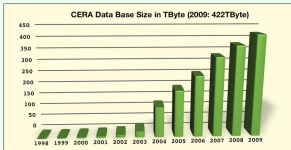
HPSS: High Performance Storage System

- 6 Sun StorageTek SL8500 tape libraries
8 robots per library
10 000 media slots per library
- 73 tape drives
- total capacity: 60 Petabytes
- 60.000 media (T10000 A/B, LTO4, 9940B and 9840C)
- 500 Terabytes disk cache
- projected fill rate: 10 Petabytes/year
- bidirectional bandwidth: 3 Gigabytes/s (sustained), 5 Gigabytes/s (peak)



The HPC Cluster water cooled IBM p575 "Power6" cluster

- 16 dual core CPUs per node
- Total size: 264 nodes, 8448 cores
249 compute nodes
12 I/O nodes
3 interactive nodes
- 18,8 GigaFlops/core
- Total system peak performance:
158 TeraFlops/s
20 Terabytes memory
3 (2011:6) Petabytes HD: GPFS



The Storage Layout at WDCC



Data Storage & Data Granularity

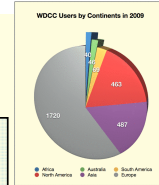
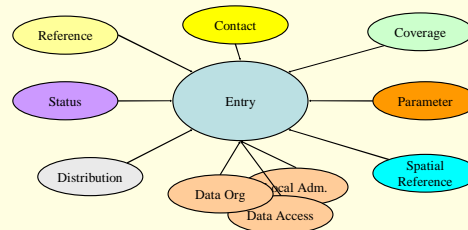
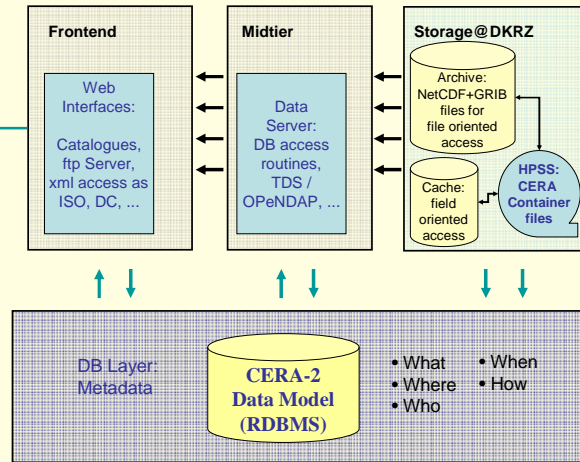
Data access of finest granularity while having a good performance is important for climate data. An intelligent interface to the mass storage (tapes) is needed.

So at DKRZ, an hierarchical storage management system (HSM) has been developed: time step data of a few dozens of kilobytes are packed to datasets, datasets are packed to container files. The CERA container files are stored on HPSS tapes, all access information is held in the RDBMS CERA-2.

As this code is open source, collaborators are welcome to enhance it further.

The CERA-2 Data Model

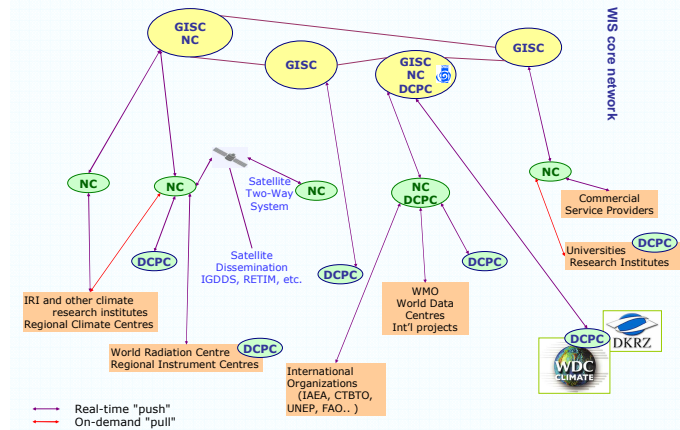
...can represent Metadata structures of ISO, Dublin Core (DC) and many other standards. It has been used at WDCC for more than ten years. Only minor changes and few supplements had to be made.



The WDCC as Part of the WMO Information System (WIS)

Components of the WIS

- NC National Centres
 - GISC Global Information System Centres (12..14)
 - DCPC Data Collection and Production Centres (>100)
- Some institutes are more than one of these (NC, GISC, DCPC).

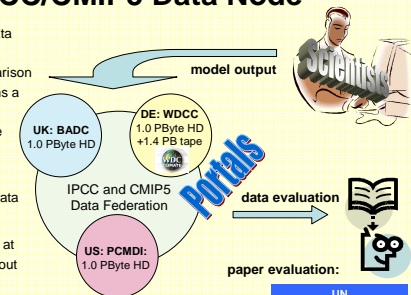


The IPCC/CMIP5 Data Node

Together with the British Atmospheric Data Centre (BADC, UK) and the Program for Climate Model Diagnosis and Intercomparison (PCMDI, USA) the WDC for Climate forms a data federation to hold and disseminate the climate data produced for the Climate Model Intercomparison Project (CMIP5) and IPCC-AR5 processes.

For web access via different portals, all data will be catalogued and quality checked.

A core part of the data will be held online at each node. It is supposed to make up about 1 Petabyte.



Earth System Science Data (ESSD)

The Data Publishing Journal

ESSD is an international, interdisciplinary journal for the publication of articles on original research data and datasets, furthering the reuse of high quality data of benefit for the Earth System Sciences.

Chief Editors: David Carlsson & Hans Pfeiffenberger

Submissions on original data are encouraged: www.earth-system-science-data.net

ESSD information flyers are available on request!



Frank Toussaint, Heinke Hoeck, Michael Lautenschlager, Martina Stockhause, Hannes Thiemann