

CERA: Database System and Data Model

**Michael Lautenschlager
Frank Toussaint**

World Data Center for Climate
(M&D/MPIMET, Hamburg)

***NINTH WORKSHOP ON
METEOROLOGICAL OPERATIONAL SYSTEMS
ECMWF, Reading/Berks., 10 - 14 November 2003***

Content:

- ICSU WDC for Climate
- DKRZ archive development
- CERA¹⁾ concept
- WDCC Content

Second Part

Data model and architecture
(Frank Toussaint)

1) Climate and Environmental data Retrieval and Archiving

Start: Approved in January 2003

Maintenance: Model and Data (M&D/MPIMET) and German Climate Computing Centre (DKRZ)

Mission: Data for climate research are collected, stored and disseminated

ICSU Policy: long-term archiving and unrestricted data access for scientists

Restriction: Only climate data products, no raw data storage.

Content: Emphasis is spent on climate modelling and related data products.

Co-operation: with thematically corresponding data centres like WDC-MARE (Bremen) and WDC-RSAT (Oberpfaffenhofen)

URL: <http://www.mad.zmaw.de/wdcc/>

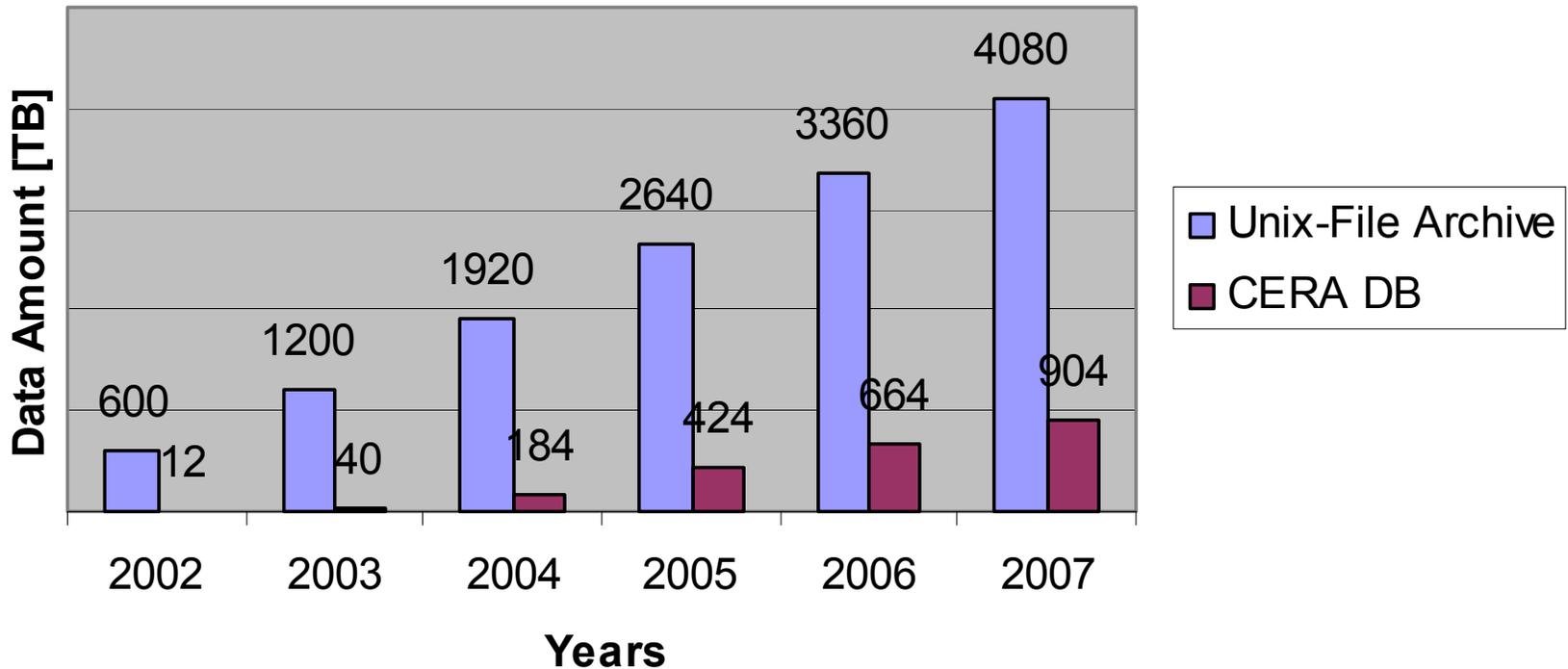
Basics observations and assumptions:

- 1) **Unix-File archive** content end of 2002: 600 TB including Backup's
- 2) Observed archive rate (Jan. - May 2003): 40 TB/month
- 3) **System changes**: 50% compute power increase in August 2003
- 4) **CERA DB** size end of 2002: 12 TB
- 5) Observed Increase (Jan. - May 2003): 1 TB/month
- 6) **Automatic fill process** into CERA DB is going to become operational with 4 TB/month this year and should increase from 10% of the archiving rate to approx. 30% end of 2004

DKRZ Archive Development



DKRZ's Archive Increase (Estim. 09.03)



DKRZ: "Conservative Estimate"



Year	2003	2004	2005	2006	2007
Estimated File Archive Size	1,2 PB	1,9 PB	2,6 PB	3,4 PB	4,1 PB

Problems in file archive access:

- **Missing Data Catalogue**

Directory structure of the Unix file system is not sufficient to organise millions of files.

- **Data are not stored application-oriented**

Raw data contain time series of 4D data blocks.

Access pattern is time series of 2D fields.

- **Lack of experience with climate model data**

Problems in extracting relevant information from climate model raw data files.

- **Lack of computing facilities at client site**

Non-modelling scientists are not equipped to handle large amounts of data (1/2 TB = 10 years T106 or 50 years T42 in 6 hour storage intervals).

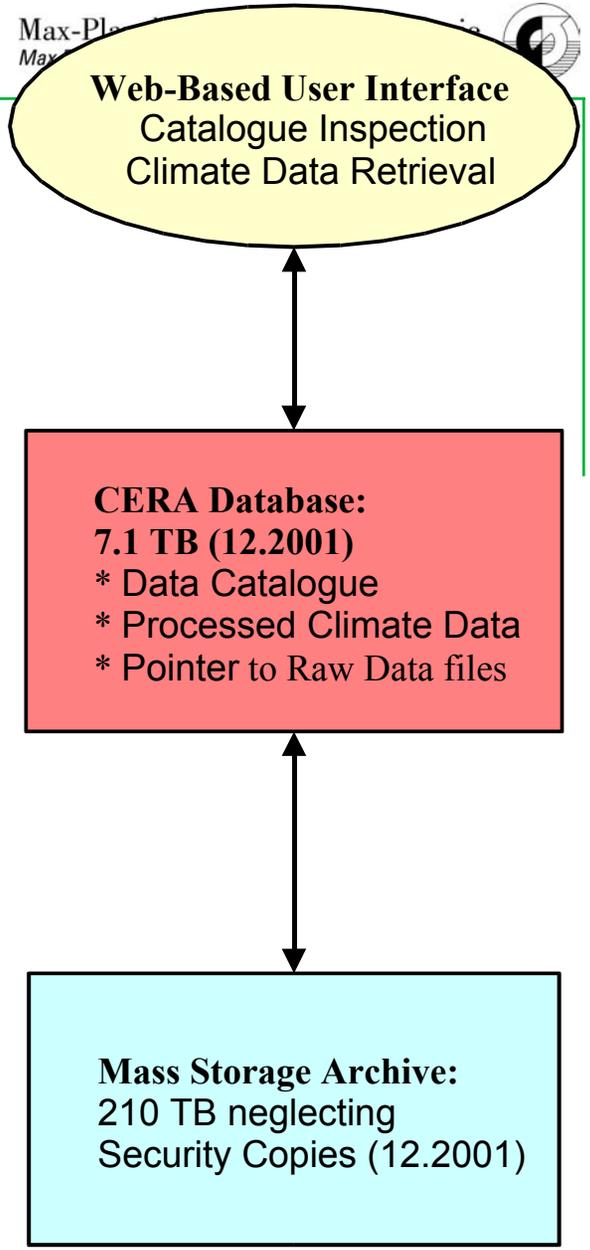
CERA Concept: Semantic Data Management

(I) Data catalogue and pointer to Unix files

- Enable search and identification of data
- Allow for data access as they are

(II) Application-oriented data storage

- **Time series** of individual variables are stored as BLOB entries in DB Tables
Allow for fast and selective data access
- Storage in **standard file-format** (GRIB)
Allow for application of standard data processing routines (PINGOs)



Current database size is

23.251 Terabyte

Number of experiments: **304**

Number of datasets: **30201**

Number of blob within CERA
at 29-OCT-03: **1403519699**

Typical BLOB sizes:

17 kB and **100 kB**

Number of data retrievals:

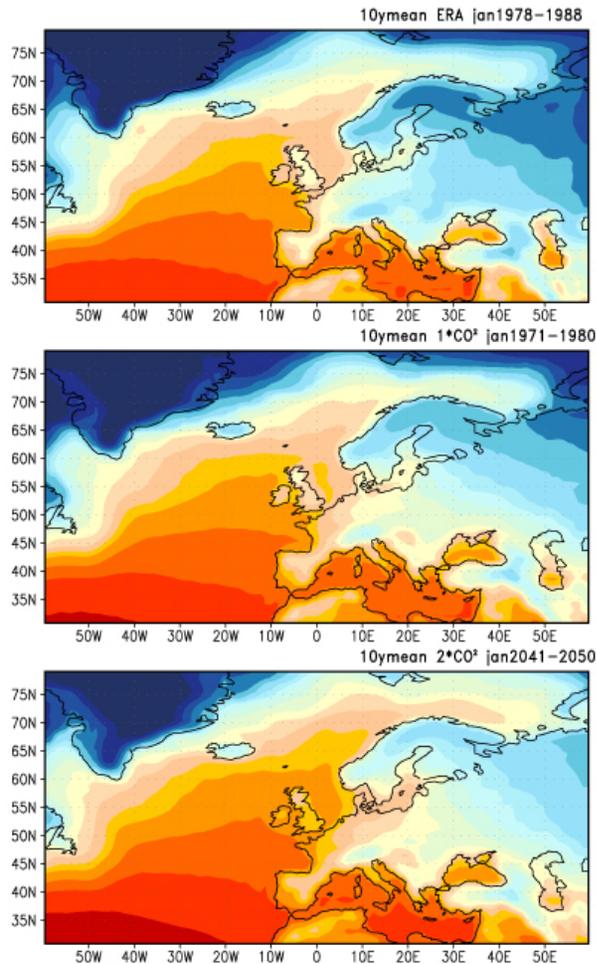
1500 – 8000 / month

More than **23 TB**
climate (model) data
web accessible!

- ◆ **Climate Model Data** (Continuous stream of **new data**)
- ◆ **IPCC DDC** (Data Distribution Centre)
Will be continued for the Fourth Assessment Report
- ◆ **CEOP** (Coordinated Enhanced Observing Period) **Model output retention and handling Centre**
Part of WCRP that was motivated by GEWEX with focus on water and energy cycles within the climate system (01.10.2002 – 31.12.2004)
- ◆ **Observational Data**
 - Model related observations:** ERA15/40 (ECMWF), NCEP 40 Y. Reanal.
 - Instrumental data:** WOCE (World Ocean Circulation Experiment)
 - Earth observations:** Access to SST's from NOAA AVHRR in cooperation with WDC RSAT (**distributed archive**)
- ◆ **Project Support** (encourage Good Scientific Practice)
 - HOAPS** (Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data)
 - CARIBIC** (Civil Aircraft for Regular Investigation of the Atmosphere Based on an Instrumentation Container), MPI Mainz
 - Different model applications**

SCENARIO IS92a ECHAM4 T106

temperature in °C: 10y mean, january

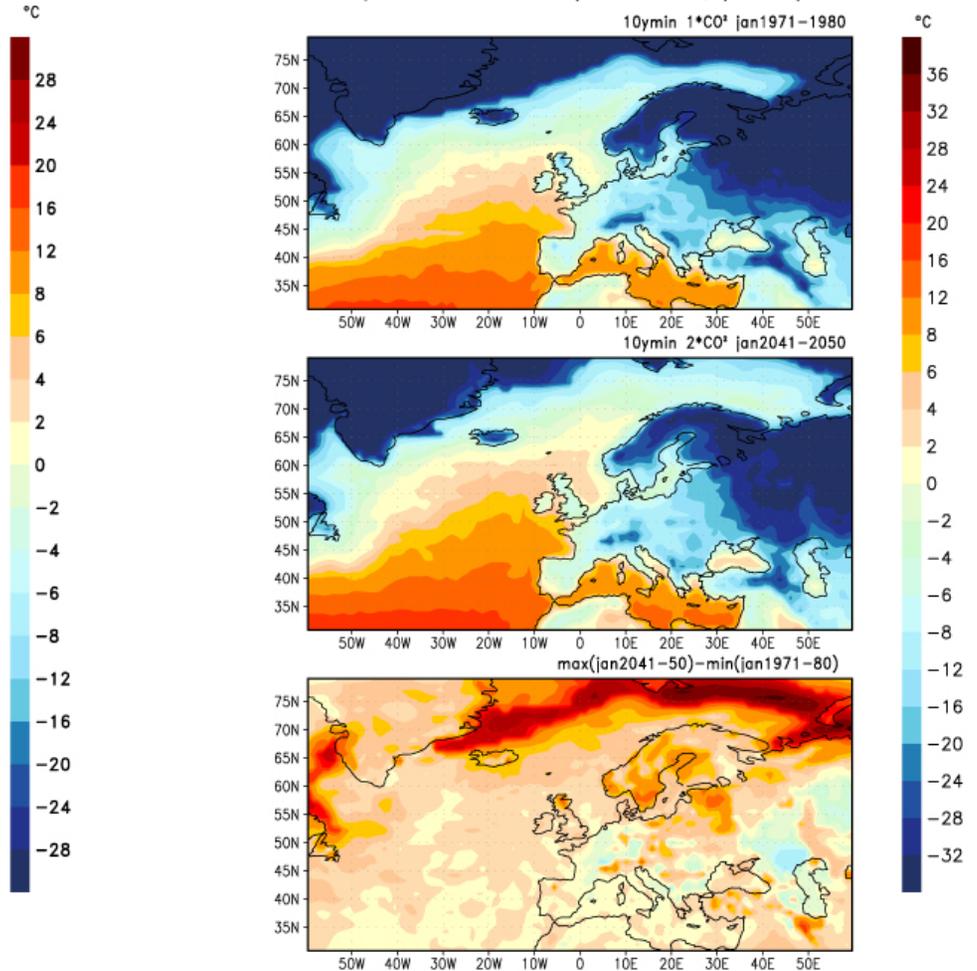


GrADS: COLA/IGES

M&D: M. Lautenschlager, H. Winter, J. Wegner

SCENARIO IS92a ECHAM4 T106

temperature in °C: 10y abs. min, january



2002-05-30-13:37

GrADS: COLA/IGES

M&D: M. Lautenschlager, H. Winter, J. Wegner

2002-05-30-13:18