

# The CEOP Model Data Archive at the World Data Center for Climate as part of the CEOP Data Network

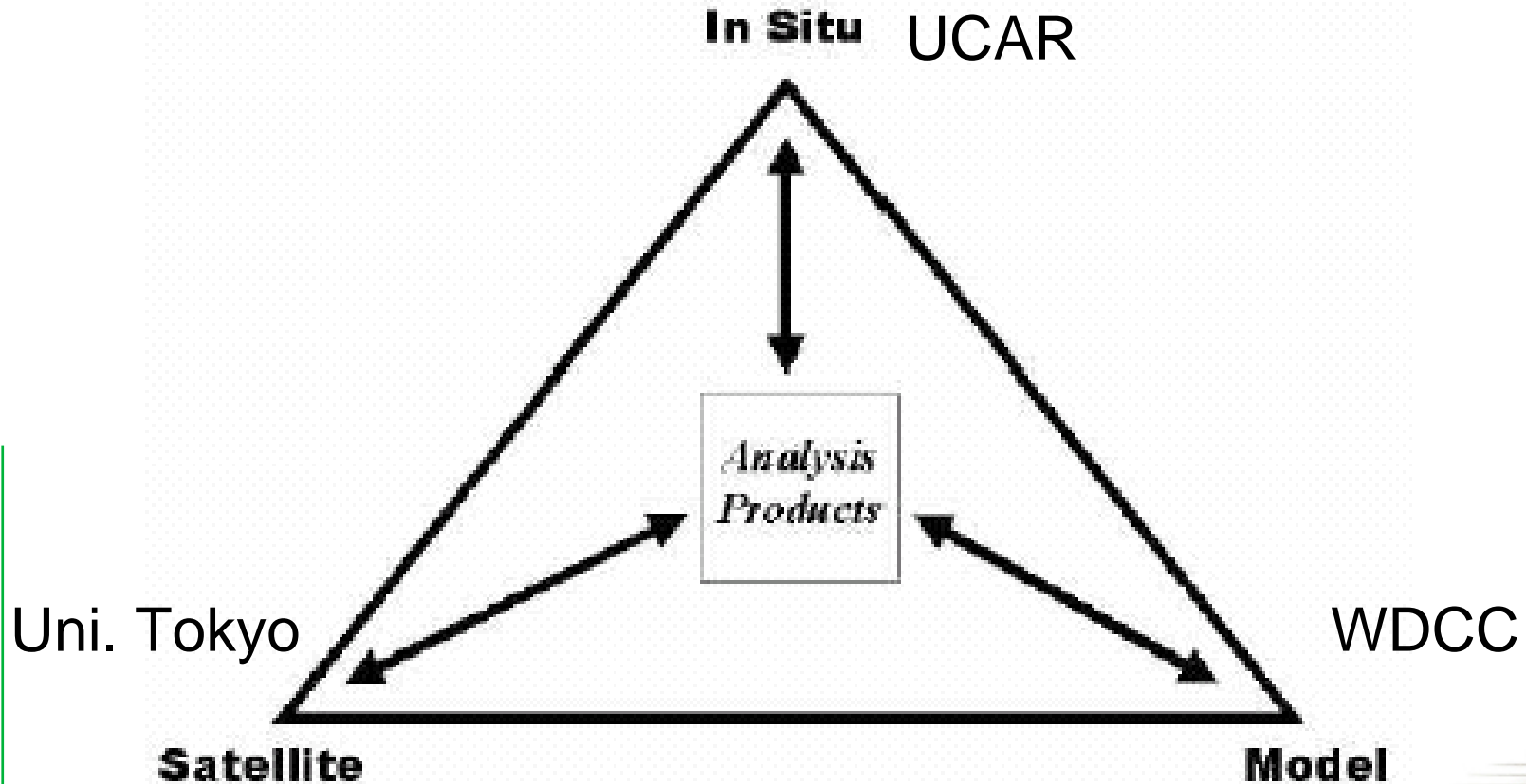
CEOP / IGWCO Joint Meeting  
Feb. 28th – Mar. 4th, 2005, Tokyo

**Michael Lautenschlager**

Frank Toussaint, Hans Luthardt

World Data Center for Climate

Max-Planck-Institut für Meteorologie / Modelle und Daten, Hamburg



**Start:** Approved in January 2003

**Maintenance:** Model and Data (M&D/MPI-M) 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 in CERA DB, 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.wdc-climate.de/>

## (I) Data catalogue and Unix files (pointer or BLOB-table-entry)

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

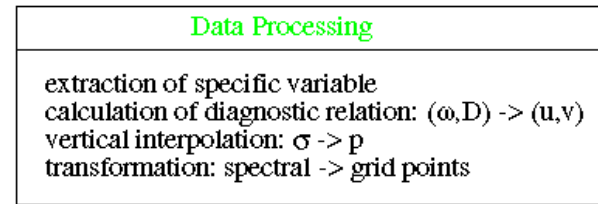
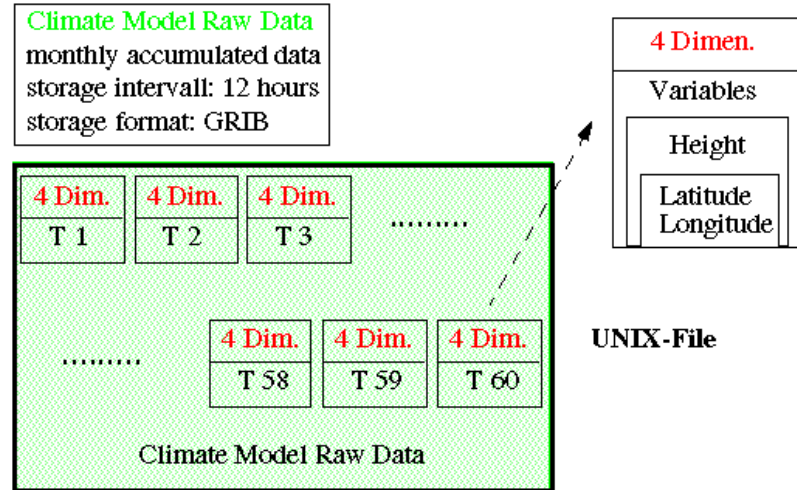
## (II) Application-oriented data storage

- **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, NetCDF)  
Allow for application of standard data processing routines (PINGOs)

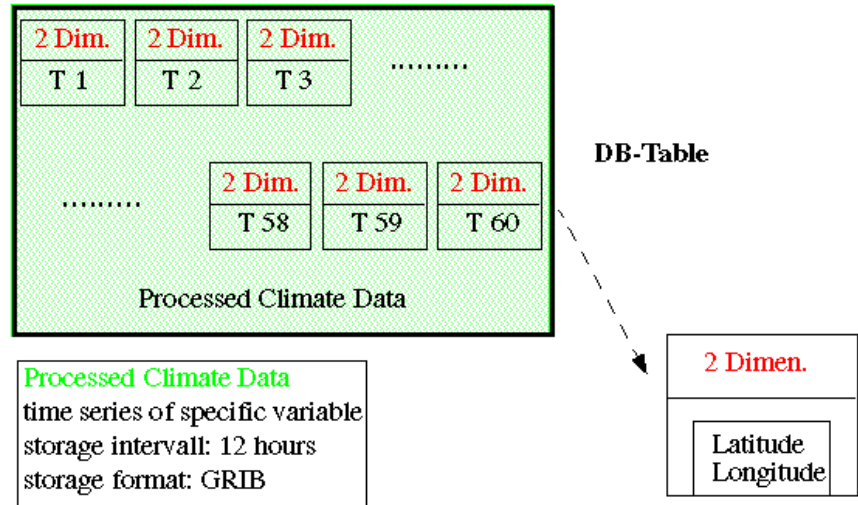
1) **C**limate and **E**nvironmental data **R**etrieval and **A**rchiving

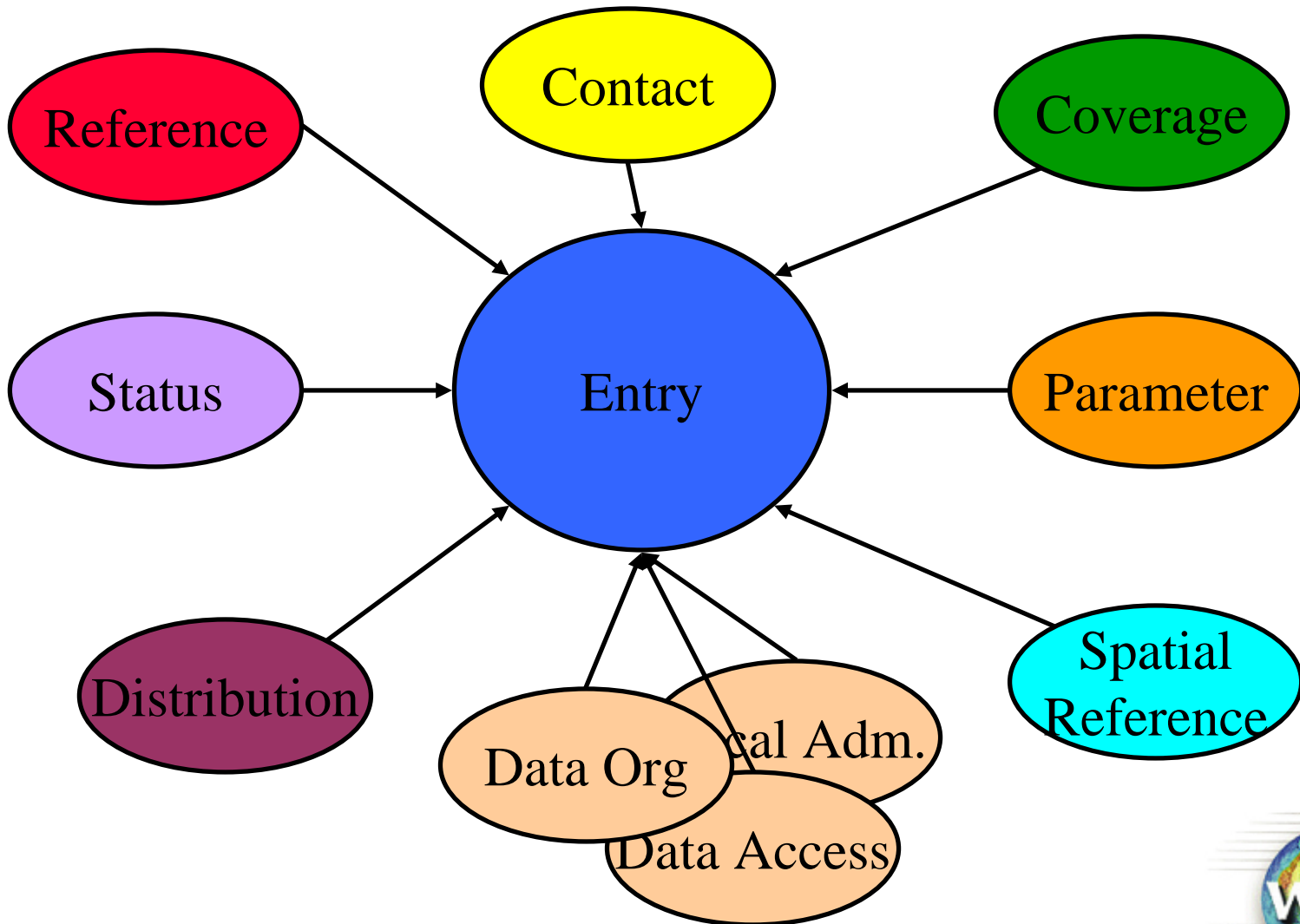


## Climate Model Raw Data



## Application-oriented Data Storage (Interface level 2)





## The CERA2 data model ...

allows for **data search** according to discipline, keyword, variable, project, author, geographical region and time interval and for **data retrieval**.

allows for **specification of data processing** (aggregation and selection) without attaching the primary data.

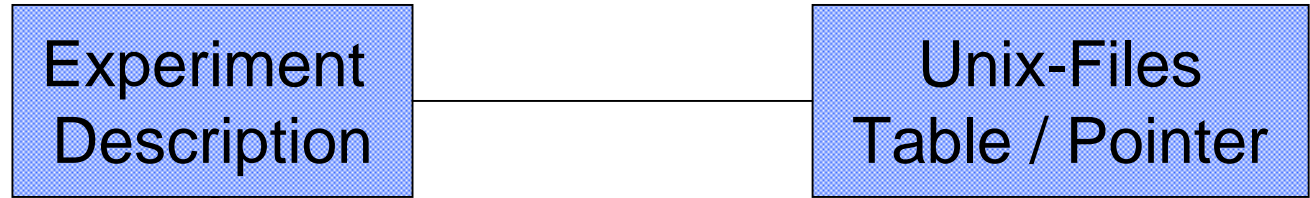
is flexible with respect to **local adaptations**, to storage of **different types** of geo-referenced data, and to definition of **data topologies** (hierarchical, network, ....).

is open for cooperation and **interchange with other database** systems (e.g. FGDC metadata standard and ISO 19115 included).

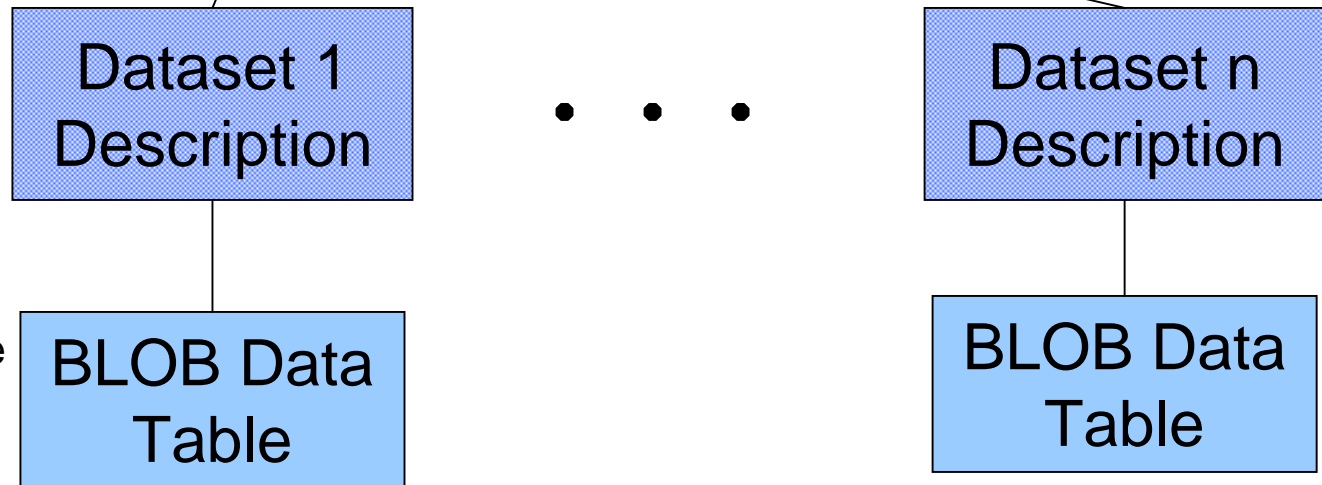
**But:**

is **not the simplest** data model for each single application.

**Level 1 - Interface:**  
Metadata entries  
(XML, ASCII)  
+ Data Files



**Level 2 – Interf.:**  
Separate files  
containing BLOB  
table data in  
application  
adapted structure  
(time series of  
single variables)

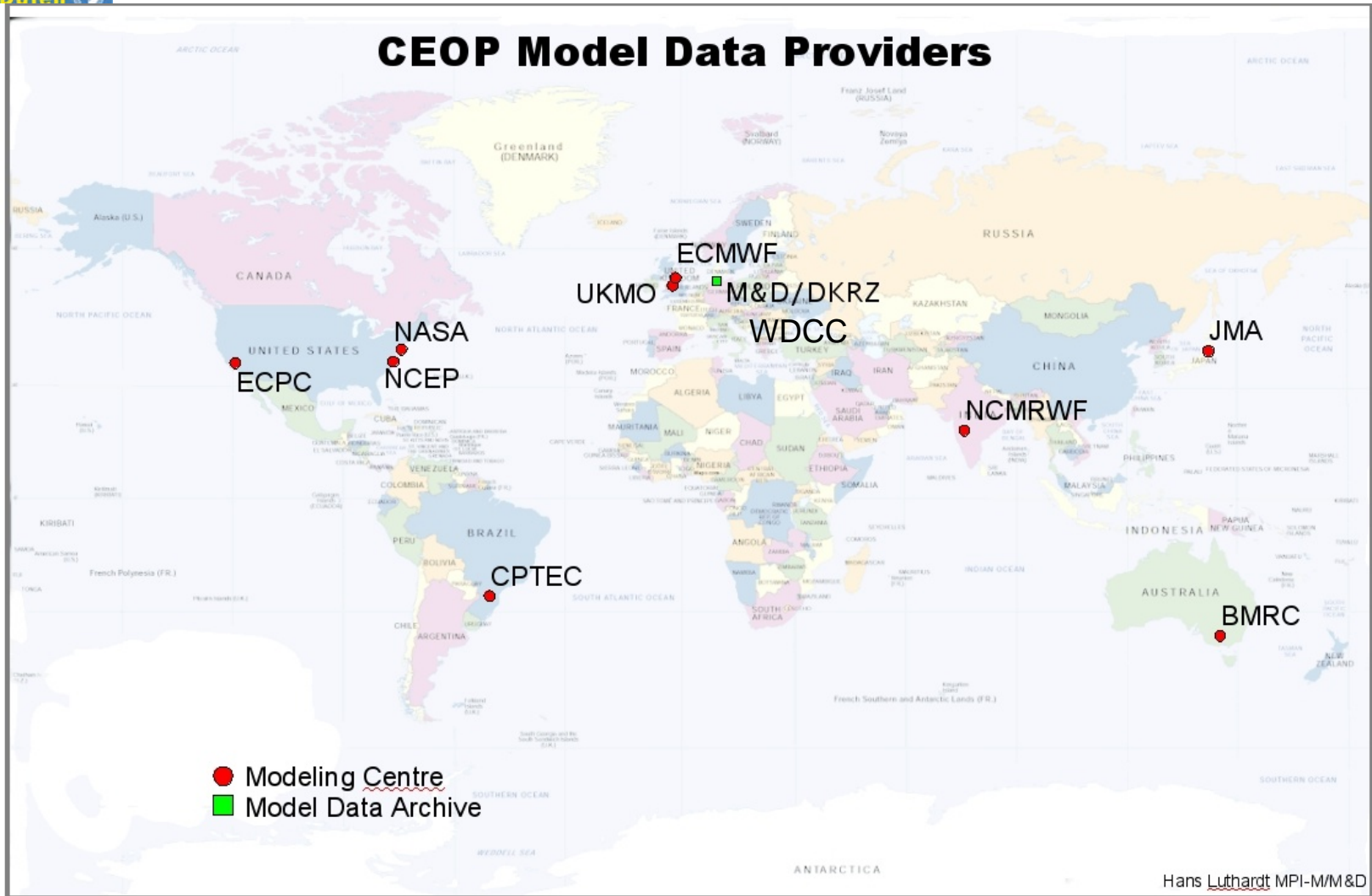


BLOB DB Table corresponds to scalable,  
virtual file at the operating system level.

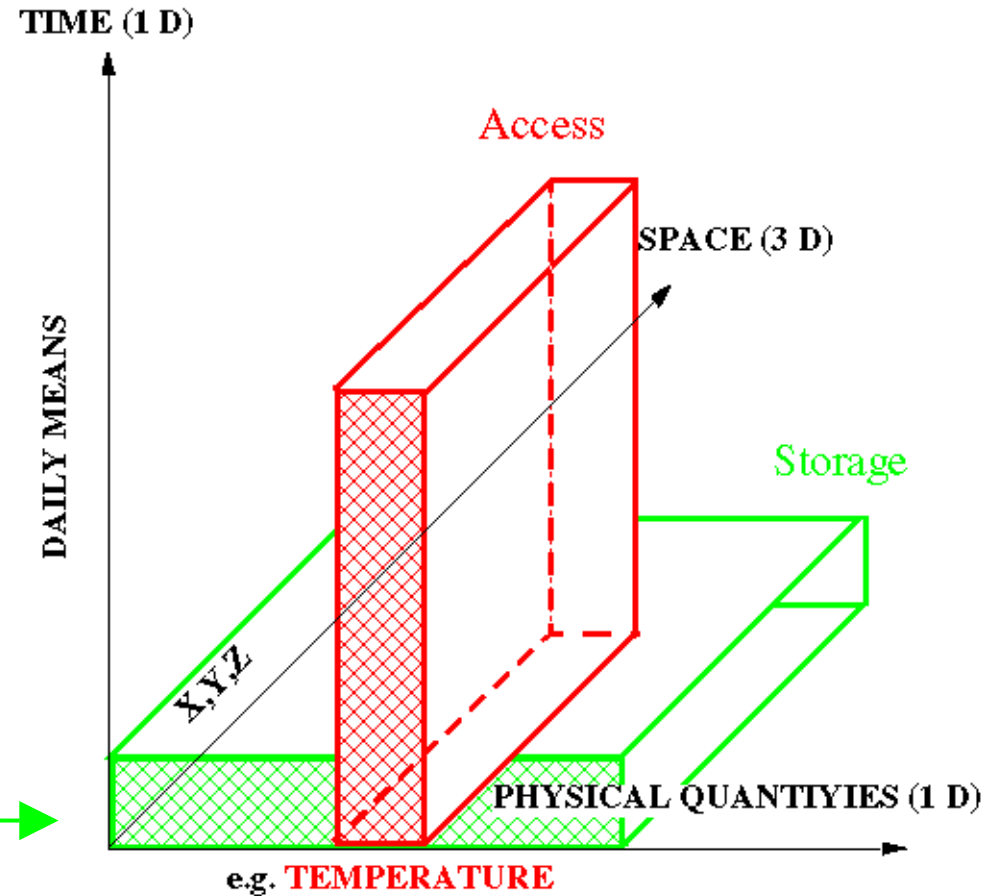
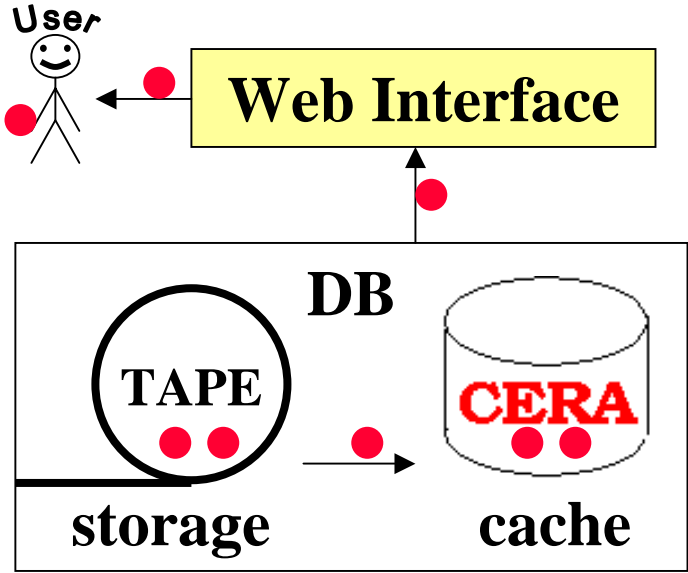


# WDCC Data Content

- ◆ **Climate model data**
  - IPCC (Hamburg model runs)
- ◆ **IPCC DDC**
- ◆ **CEOP**
- ◆ **Observational Data**
  - ERA15/40 (ECMWF), NCEP 40Y
  - WOCE
- ◆ **Project support**
  - HOAPS
  - CARIBIC
  - BALTEX
  - SFB512
  - Different model applications
- ◆ **Size of CERA: 140 Tbyte**
- ◆ **Number of experiments: 382**
- ◆ **Number of datasets: 51000**
- ◆ **No. of blobs: 4 billion**
- ◆ **Downloads last 12 months > 330.000**







## Binary Large Objects

- (BLOBS)

Global coverage of 1 (more) parameter at 1 (more) level for 1 moment

- Finest granularity for DB data access (SQL) is on BLOB level
- Sub-BLOB data access requires data processing outside the WDCC DB system
- BLOB sizes within WDCC
  - **Fine granularity:** global field of 1 variable at 1 vertical level for 1 moment  
20 – 100 KB / BLOB depending on model resolution
  - **Coarse granularity:** global fields of all variables at all vertical levels for 1 moment  
around 25 MB / BLOB for CEOP model output
- Model output **file sizes:** global fields of of all variables at all vertical levels for many moments  
500 MB – 1 GB / File for CEOP model output

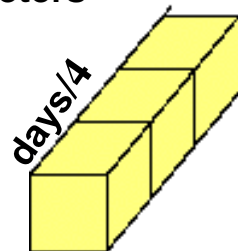
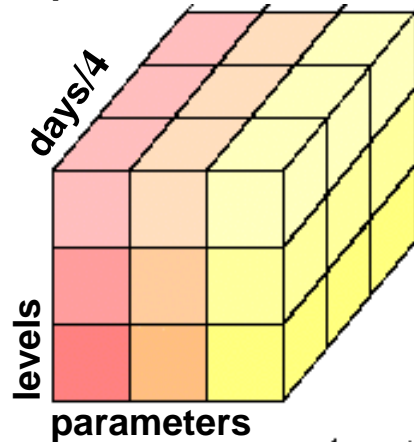
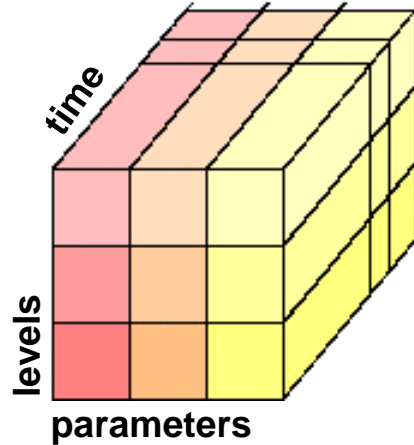
how we get the grid data:  
Files from prediction model



postprocessing step 1:  
homogenizing time  
10 - 50 MB per blob



postprocessing step 2:  
isolation of levels &  
parameters



**Storage of global coverages per file or BLOB :**

all levels, all parameters  
arbitrary time intervals

all levels, all parameters  
1 moment (6 by 6 hours)

1 level, 1 parameter  
1 moment

METADATA:

DATA:

GUI:

display in applet (1)

JDBC (2)



jblob-script:

Search for  
DS names (2)

JDBC (2)

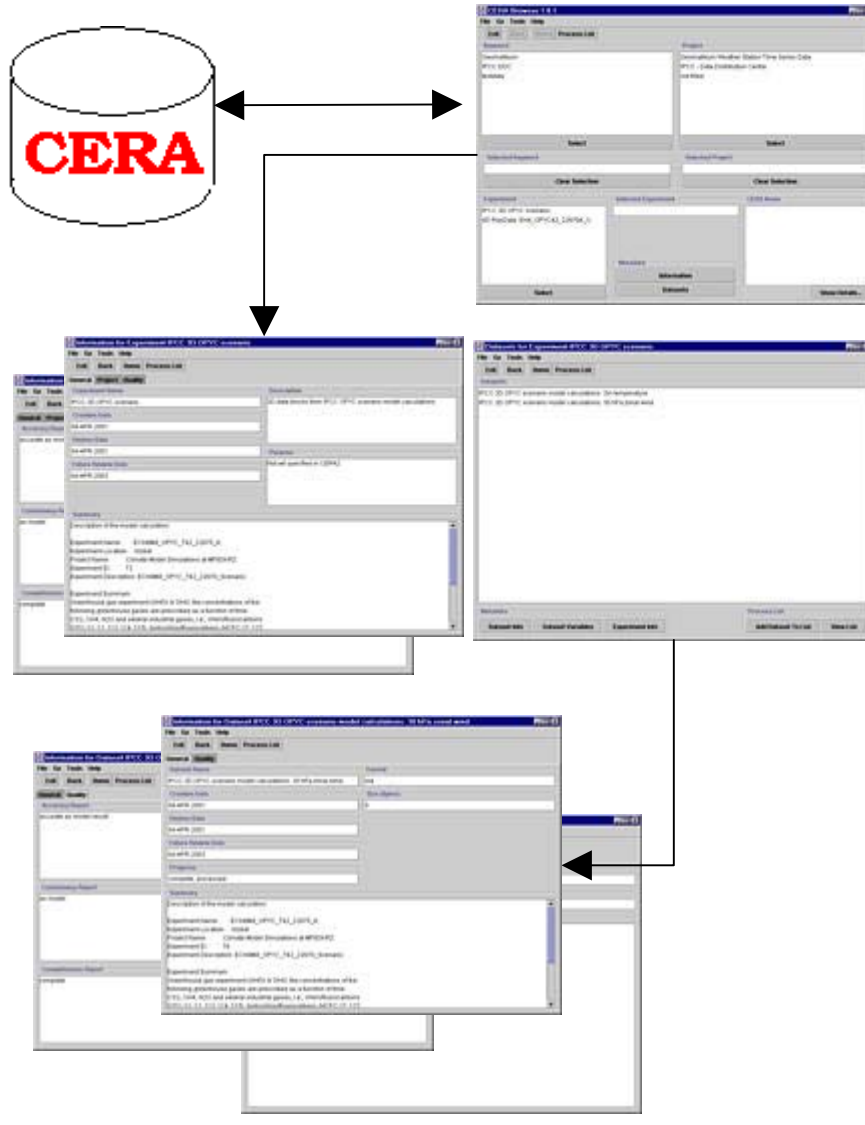
**jblob -f ...**

http:

- html-display (3)  
- xml-download  
(ISO, DC, ...) (4)

download  
http (5)

URL:  
**http://...**



## Selection via CERA meta data:

- selection of the experiment (=model run)
- display of meta data: experiment, quality, datasets
- selection of the dataset
- display of dataset information
- add datasets to “process list”
- select time span and download from tape archive to data server and to the client



# User Interface 2: jblob Command for data download

## Retrieval by script :

- Java based – for Unix/Linux and Windows
- JDBC connection

- examples :

```
jblob -datasetname <name> -options
```

```
jblob -datasetid <id> -options
```

```
jblob -showdatasets "<search_string>"  
(% as wildcard)
```

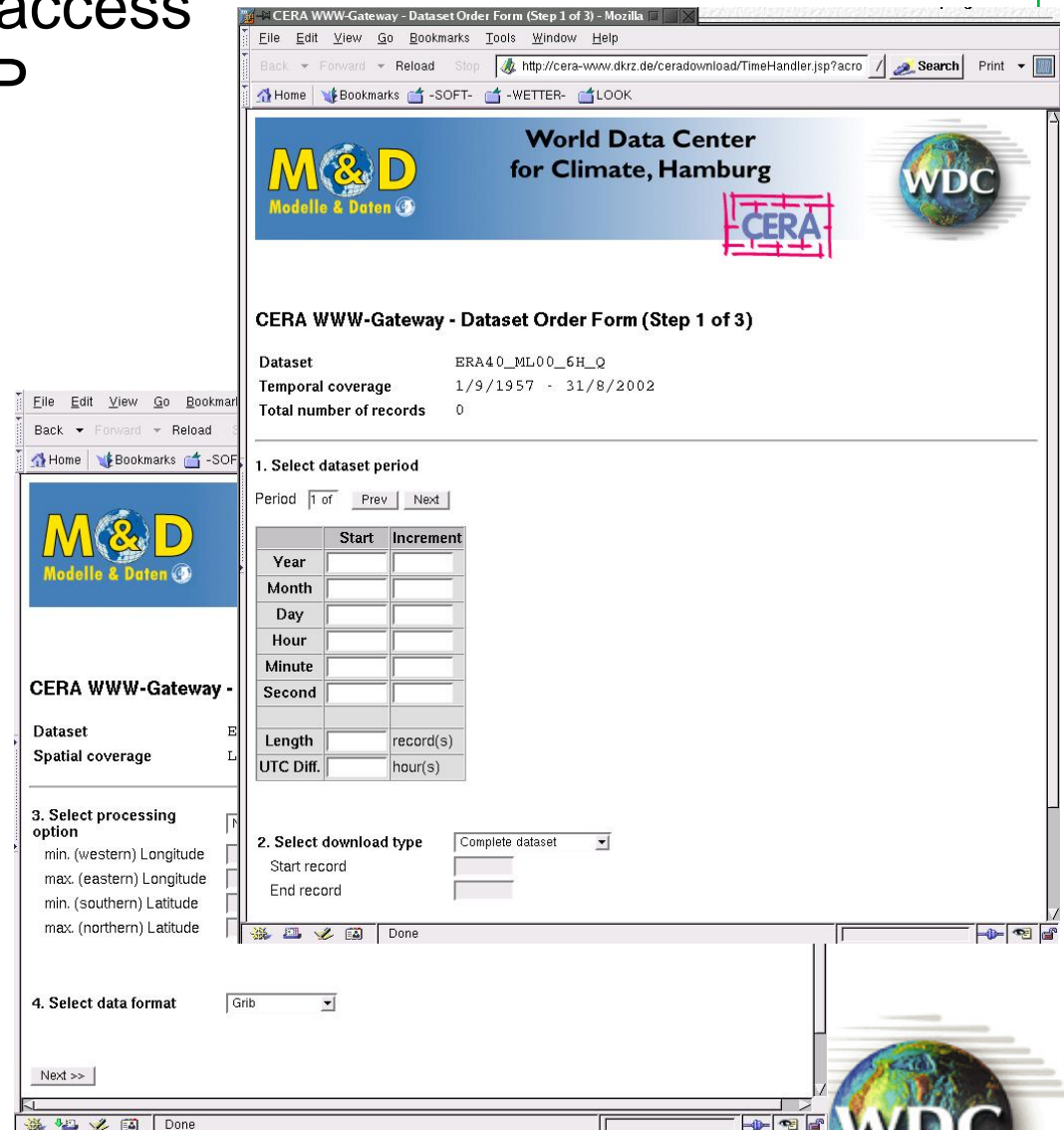
## Retrieval specification

- dataset selection from CERA catalogue

- three steps of specification:

selected region in space and time

download by http



**CERA WWW-Gateway - Dataset Order Form (Step 1 of 3)**

Dataset: ERA40\_ML00\_6H\_Q  
 Temporal coverage: 1/9/1957 - 31/8/2002  
 Total number of records: 0

**1. Select dataset period**

Period 1 of 1 | Prev | Next

	Start	Increment
Year	<input type="text"/>	<input type="text"/>
Month	<input type="text"/>	<input type="text"/>
Day	<input type="text"/>	<input type="text"/>
Hour	<input type="text"/>	<input type="text"/>
Minute	<input type="text"/>	<input type="text"/>
Second	<input type="text"/>	<input type="text"/>
Length	<input type="text"/>	record(s)
UTC Diff.	<input type="text"/>	hour(s)

**2. Select download type** | Complete dataset

Start record:   
 End record:

**3. Select processing option**

min. (western) Longitude:   
 max. (eastern) Longitude:   
 min. (southern) Latitude:   
 max. (northern) Latitude:

**4. Select data format** | Grib

Next >>

Retrieval by browser (or Unix command wget):

```
wget http://mad.dkrz.de/Daten/ ...
... XML/CERA2WINIq.xsql?&id=50
```

- XML  
ISO format  
http
- or:  
Browser display  
as html file

```
- <MD_Metadata>
- <fileIdentifier>
  <CharacterString>EH4_OPYC_SRES_B2_TEMP2EH4_OPYC_SRES_B2_TEMP2:</CharacterString>
</fileIdentifier>
- <language>
  <CharacterString>en</CharacterString>
</language>
- <parentIdentifier>
  - <CharacterString>
    ECHAM4_OPYC_SRES_B2: 110 YEARS COUPLED B2 RUN 6H VALUES
  </CharacterString>
</parentIdentifier>
- <hierarchyLevel>
  <MD_ScopeCode codeList="http://mad.dkrz.de/Daten/Metadata_Fill/scope.html" codeListValue="data"
</hierarchyLevel>
- <contact>
  - <CI_ResponsibleParty>
    - <organisationName>
      - <CharacterString>
        World Data Center for Climate http://www.mad.zmaw.de/wdcc/
      </CharacterString>
    </organisationName>
    - <role>
      <CI_RoleCode codeList="http://mad.dkrz.de/Daten/Metadata_Fill/contact_type.xsql" codeListValu
    </role>
    </CI_ResponsibleParty>
  </contact>
```

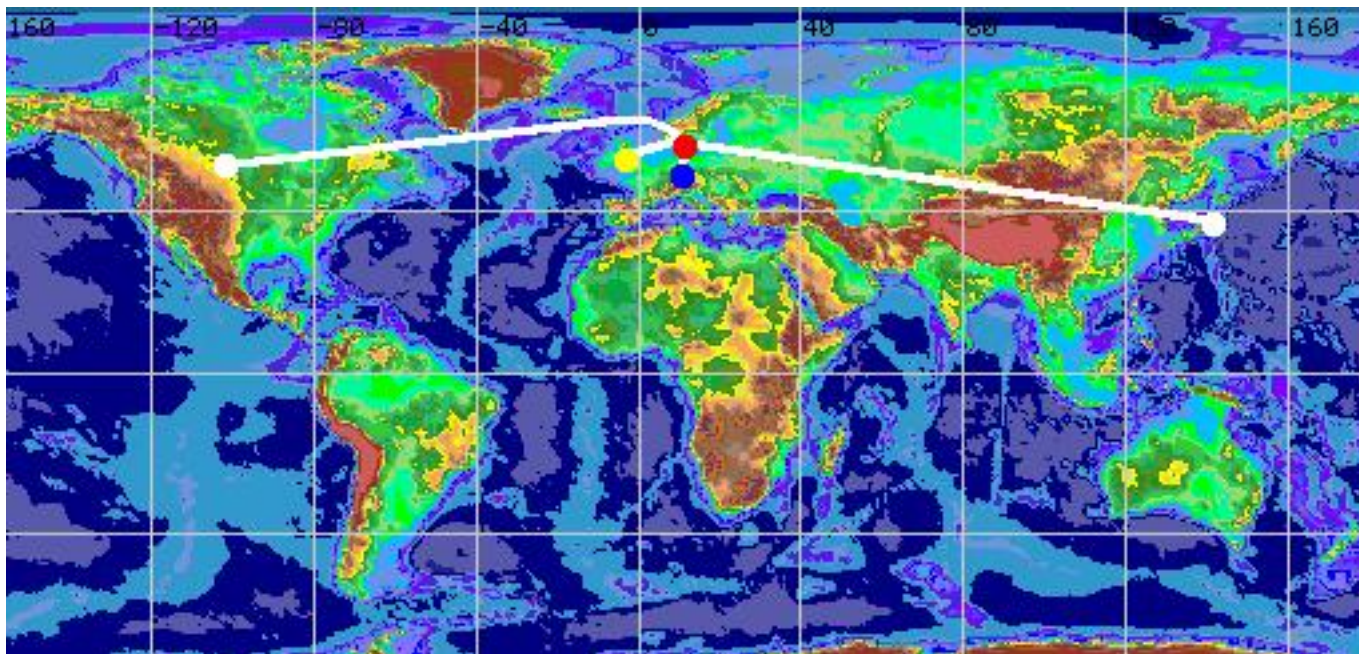
# User Interface 5: http data download



- Java servlet
- URL includes:
  - specification of dataset (parameter)
  - specification of records (time span)
  - encrypted user/password
- in preparation: control by certificate



- DLR: WDCC catalogue links to external satellite data, data download by servlet
- **CEOP: mutual data access for in-situ, model and satellite data**
- BADC: distributed data holding (ERA40 and IPCC), certificates for authentication



Future activities may be summarized as

- improving the data granularity,
- harmonizing the metadata and
- completing the CEOP data sets.

## Problem:

Model data provision does not fit the data inclusion interface of the WDCC in all details.

How to organise the extra work?