WDC Climate:
Web Access to Metadata and Data

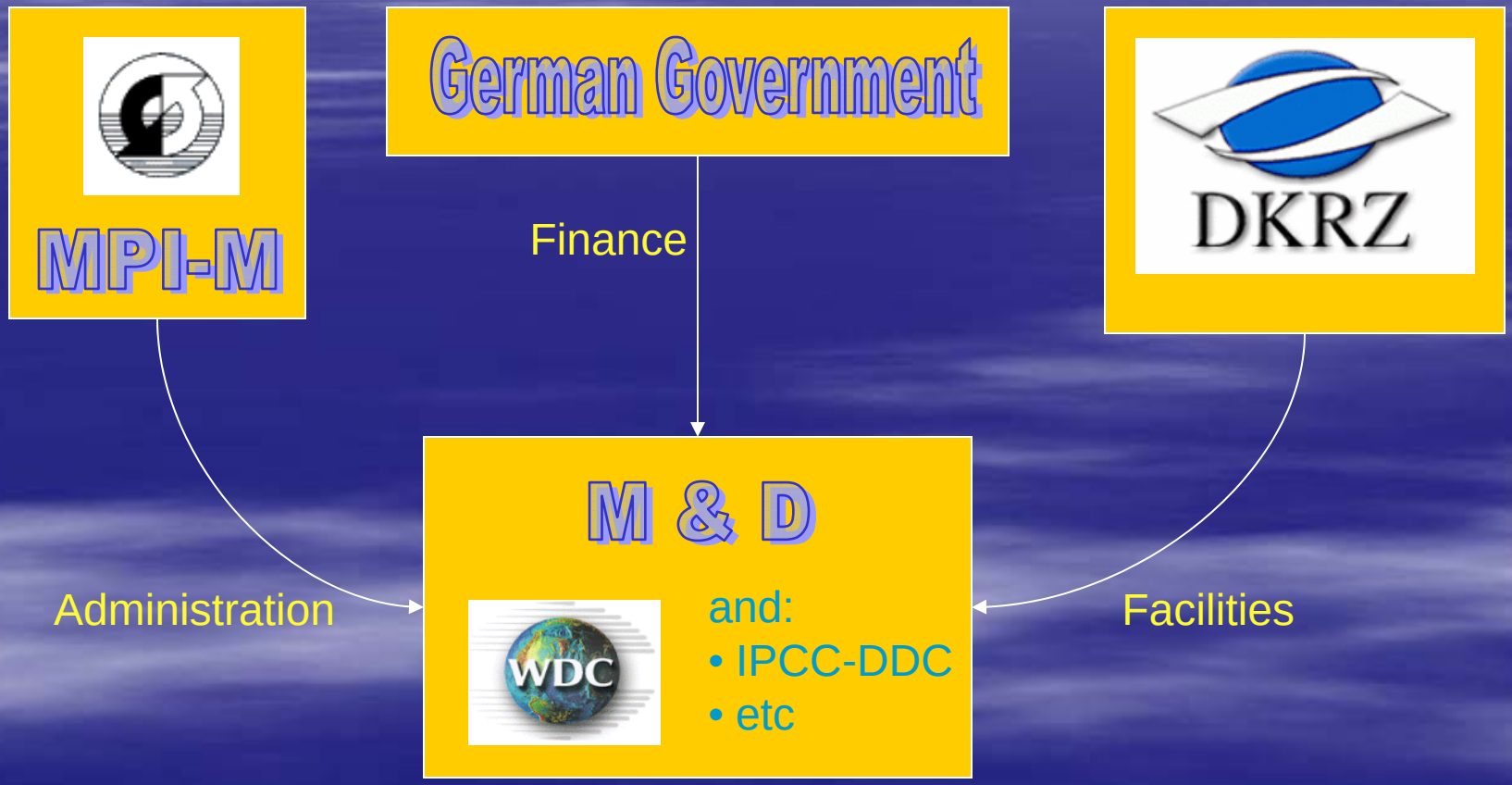
Frank Toussaint

World Data Center for Climate
(M&D/MPI-Met, Hamburg)

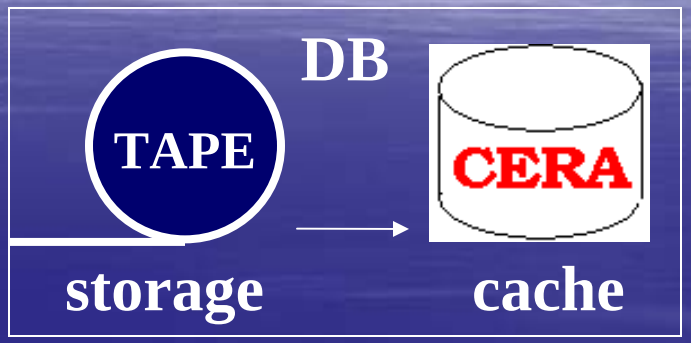
Contents

- The World Data Center for Climate (WDC-C) at Max-Planck-Institute
- Input to the DB
- Output of Metadata
- Output of Data

WDCC + M&D: Structure



Hardware and DB



The hardware

- > 2 PB data tape cartridges
- 90 TB disk cache

The database

- 170 TB web accessible via Oracle DB

Input to the DB

The metadata input: x m l

- xml interface to Oracle DB for:
  - xml metadata from data providers
  - “hand knit” xml metadata

The data input

- cutting into single time steps
  one step = 1 blob (binary large object)
- Java filling tool / JDBC

Output from the DB

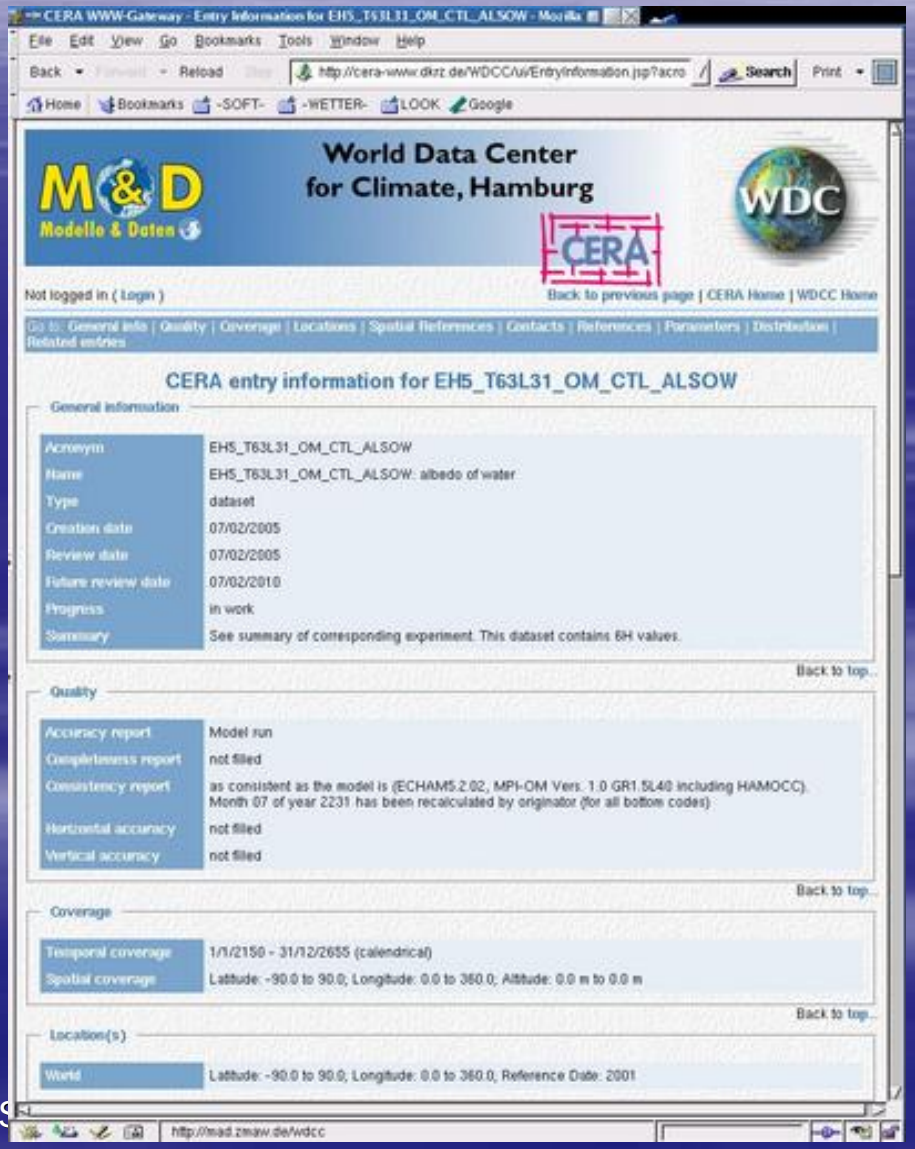
What is desired – and what is desirable ?

- browse, search, retrieve
- easy access from everywhere
- authentication and authorization for some data
- no problems with the damnable firewalls

Metadata Output: JSP + Servlets

Browsing

- online access to all CERA projects and experiments
- links to all datasets of any experiment



Metadata Output: JSP + Servlets

Search

- search on different CERA tables
- technique: JSP + html form

➔ in preparation

http – Access to Metadata

xsl mapping
from/to xml

and

JSP / Servletts

xml Web access as :

- xhtml table for quick view of raw data
- any type of xml as:
  - ISO 19115 / ISO 19139
  - Dublin Core
- project dependent xhtml design
- raw DB data

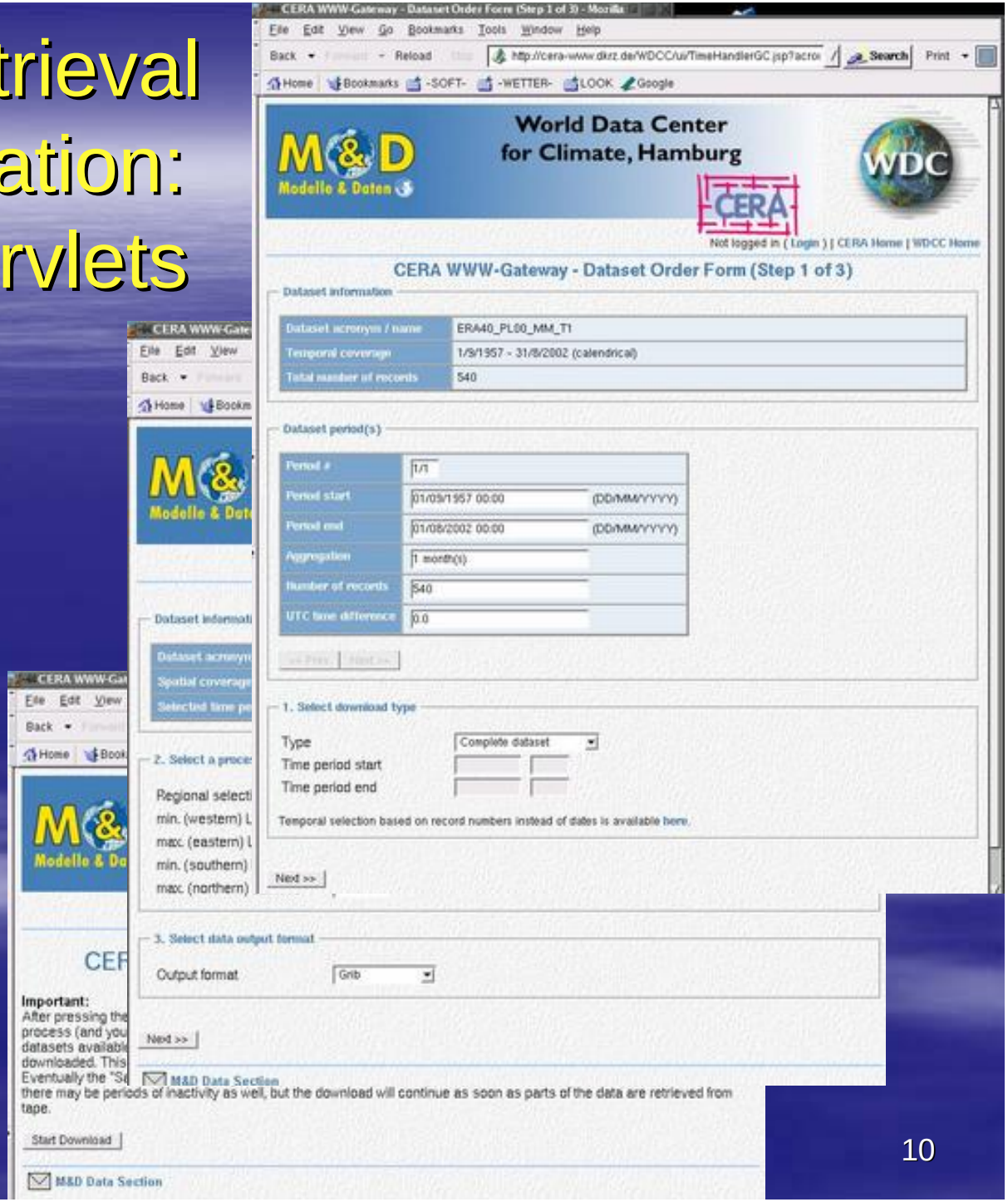
Data Retrieval Specification: JSP+Servlets

html forms

- dataset selection from CERA catalogue
- three steps of specification: bounding box and period

➔ GRIB & NetCDF

➔ download region in space and time



Data Download

http / ftp data download:

hyperlink to ftp request integrated into an html page

docking to DB by JDBC

…but :

- piping ?
  easy for GRIB - difficult for NetCDF

- postprocessing ?
  presently: cuts in time of 1-parameter-1-level global data

➔ regional cuts: prototype ready

Data Download (http/ftp)

Authentication and authorization :

html embedded hyperlink:
DB login required by servlet

DB login allows for “shopping cart”

➔ future: certificated access

DODS – an Alternative ?

http data download: e.g., a DODS Server

- multidimensional data selection and step width
  in an http request

  but (mainly open NetCDF files) :
- does not work on piped GRIB - positioning
- does not work directly on tape archives - positioning
- does not work indirectly on tape archives – performance
- has presently no access control
  at maximum: security by obscurity

➔ possibly use on DB tables

Finally :

Questions ?

?

h

- a

  ➔ p

Metadata: CERA-2 and ISO

CERA meta data and data is web-enabled


```
- <MD_Metadata xsi:schemaLocation="http://metadata.dgiwg.org/smXML ..\smXML\metadataEntity.xsd">
  + <fileIdentifier></fileIdentifier>
  + <language></language>
  + <hierarchyLevel></hierarchyLevel>
  - <contact>
    + <CI_ResponsibleParty></CI_ResponsibleParty>
    </contact>
  + <dateStamp></dateStamp>
  - <metadataStandardName>
      <CharacterString>ISO 19115</CharacterString>
    </metadataStandardName>
  + <metadataStandardVersion></metadataStandardVersion>
  - <identificationInfo>
    - <MD_Identification>
      - <citation>
        - <CI_Citation>
          - <title>
              <CharacterString>ECHAM3 T21 Early Industrial Run</CharacterString>
            </title>
          + <alternateTitle></alternateTitle>
          + <date></date>
          </CI_Citation>
        </citation>
      - <abstract>
        - <CharacterString>
            The atmosphere-ocean general circulation model ECHAM3/LSG is used for climate change ar
            integration, the CO2 concentration from 1880 to 2084 were prescribed. Up to 1985 the concen
            (Schoenwiese et al., 1990) and for the following years the IPCC scenario A (Business as Usua
            ECHAM3 and LSG are described in the DKRZ reports No.6 and No.2, respectively.
          </CharacterString>
        </abstract>
        <status/>
```

WDCC: CERA-2 DB

The Database

- 100 TB in Oracle RDBMS web accessible
  > 2 PB data in files
- distributed system (5 DB servers)

CERA-2 holds metadata & data

- CERA-2 meta data model: PIK & DKRZ / M&D

Access by web applet

- data download by applet or servlet

WDCC: CERA-2 Features

What are the typical features of CERA-2 ?

- integration of catalogue data and mass data
- web access to catalogue data and mass data
- adequate for all types of data
- capacity: >1000 daily requests
- statistics by data volume and country
- cutting of data in spacetime
- formats: grib, ASCII, netCDF, IEEE
- integrated netCDF visualization under development

WDCC: CERA-2 Structure

What are the typical structures of CERA-2 ?

- structural flexibility:
  - definable fields, tables, entry types, & various other
  - flexible lists of values (LOV): extensible but controlled
  - layouted for automated data access

- simple structure:
  - blockwise table groups
  - CERA-2 Blocks have similar structure
  - more difficult structures go into CERA Modules

WDCC: CERA-2 Structure

Coverage

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

Reference

Any publication related to the data togehter with the publication form

Parameter

Block describes data topic, variable and unit

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

Status

Status information like data quality, processing steps, etc.

Spatial Reference

Information on the coordinate system used

Distribution

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

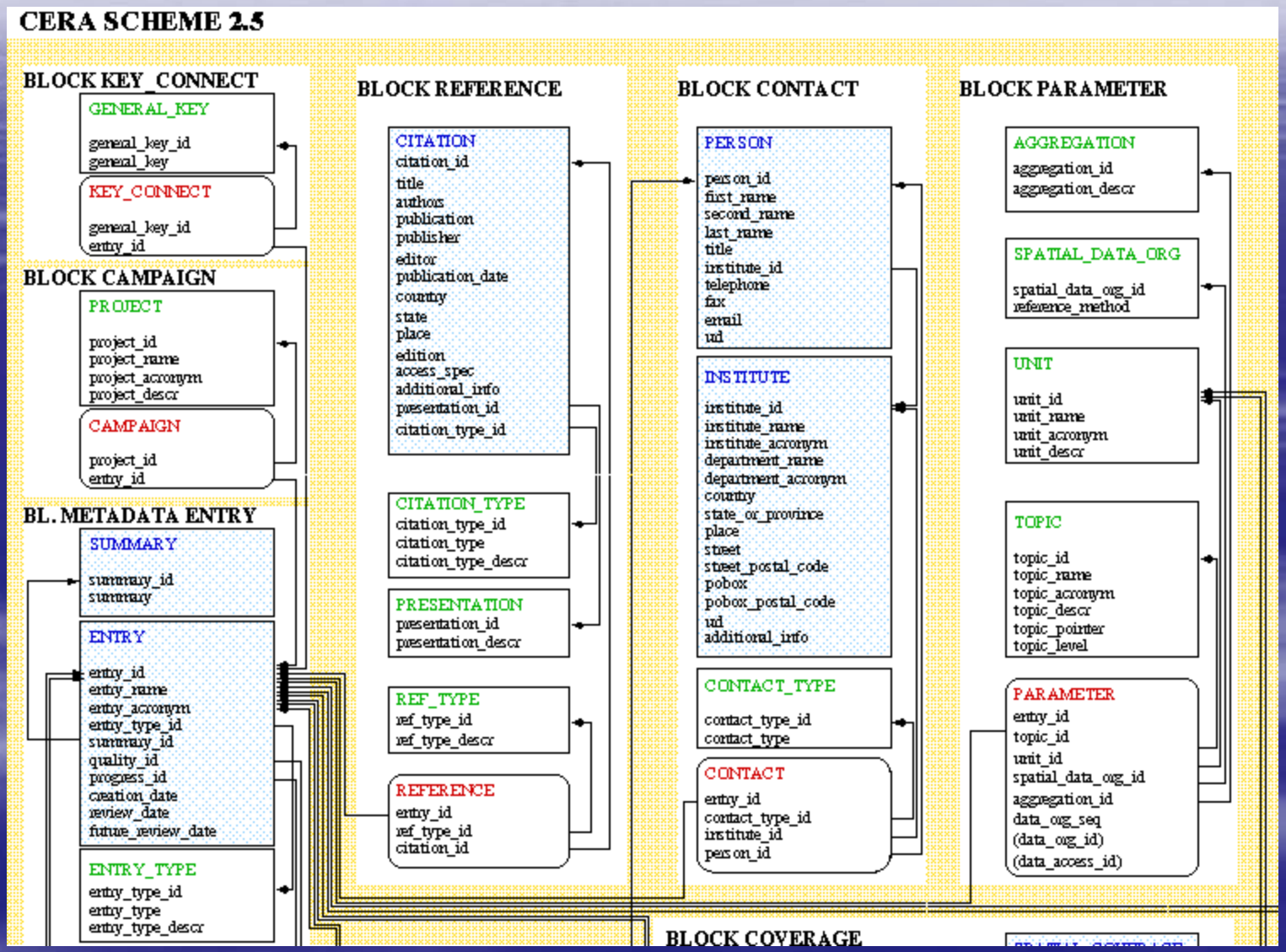
Contact

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

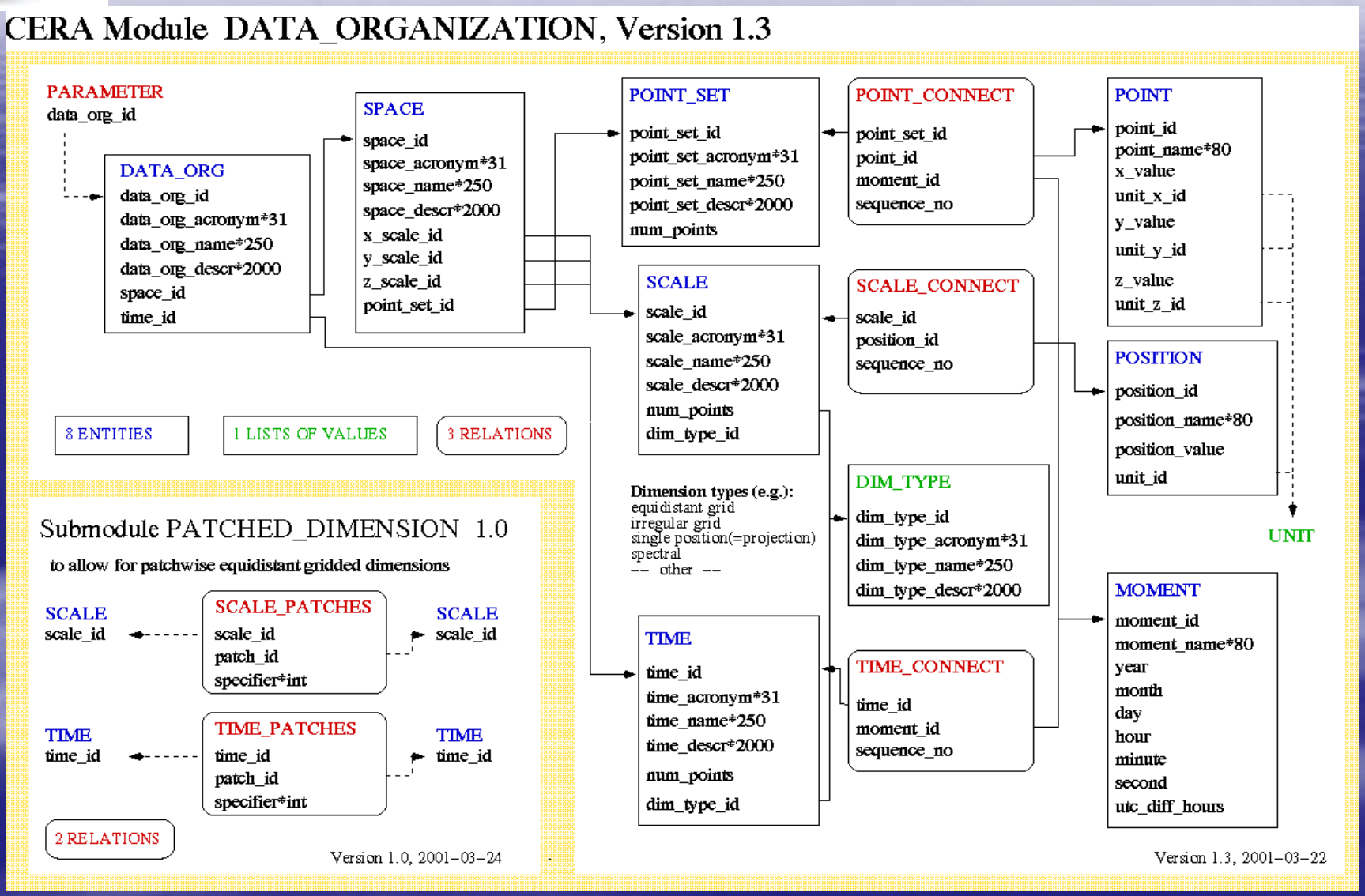
Additionally: Modules + Local Extensions

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

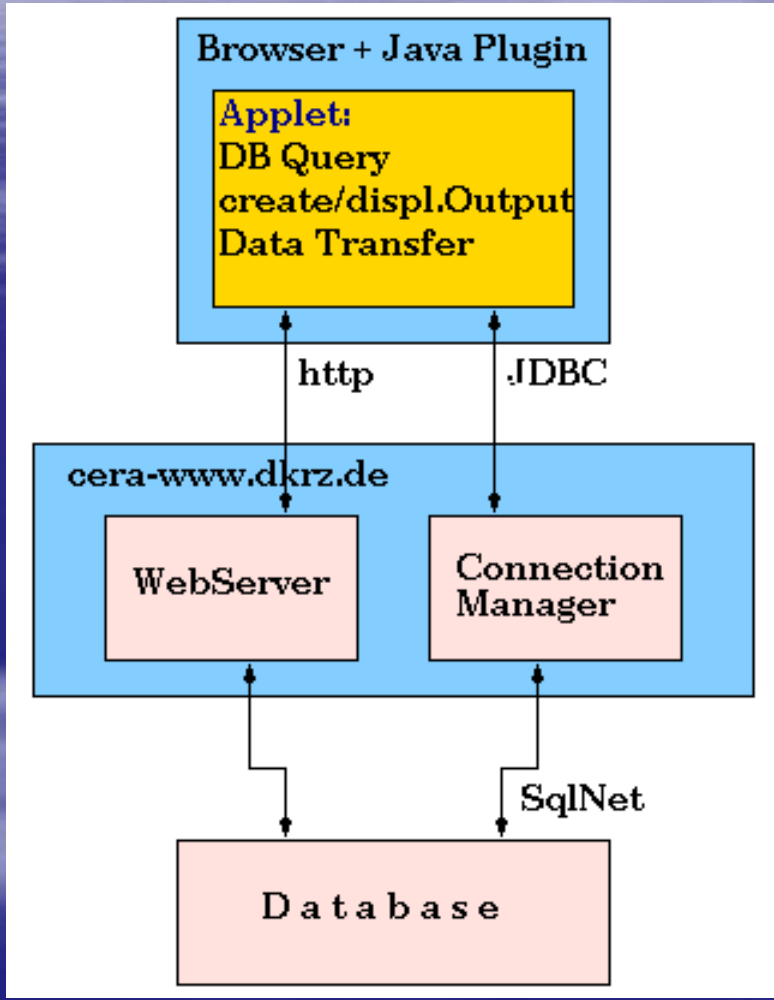
WDCC: CERA-2 Structure



WDCC: CERA-2 Module



CERA-2 Data Streams

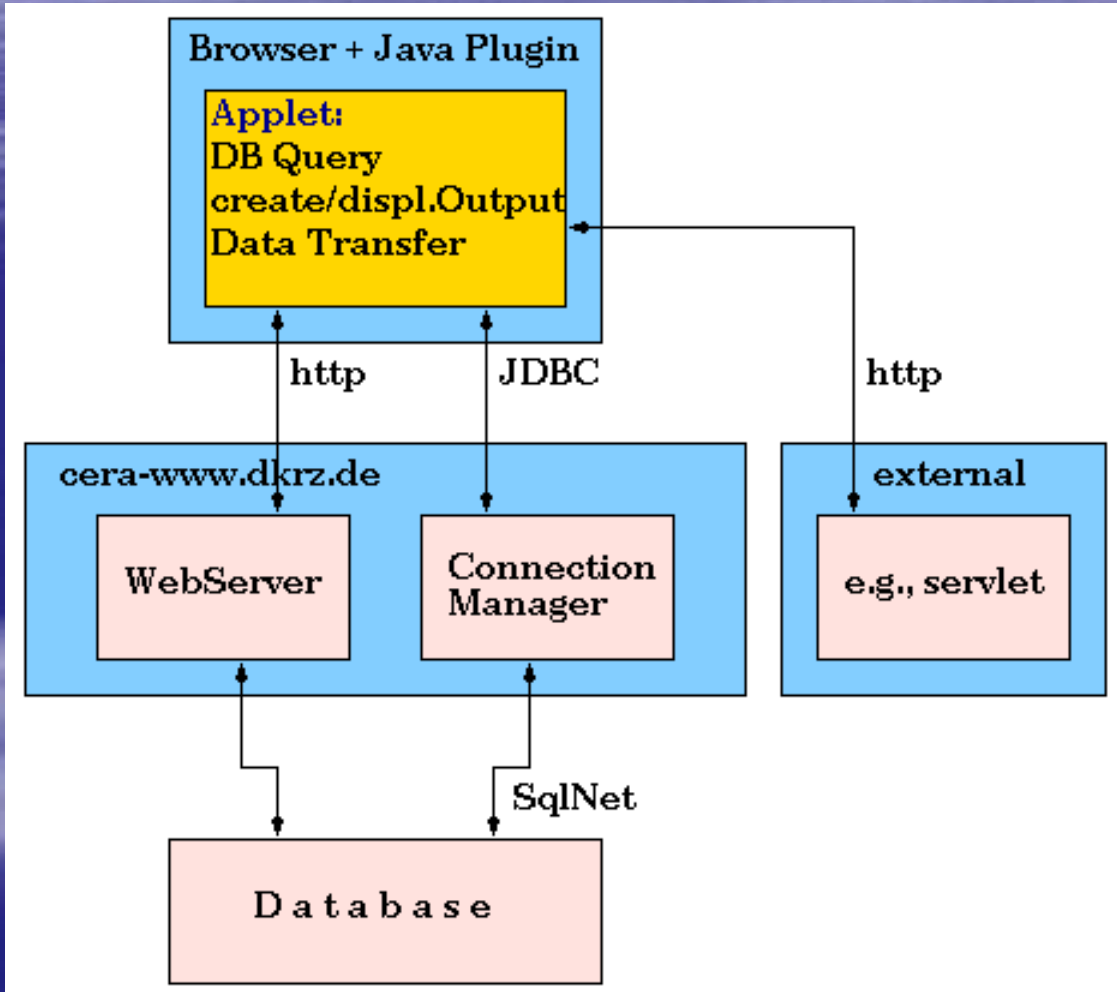


Access Client realised as web-based Java Applet.

Middleware layer provides applet and DB connection

DB-Server for catalogue operations and climate data retrieval.

CERA-2: Collaboration as WebService



Client applet receives foreign data URI from CERA-2 DB

Foreign server provides DB data by http:

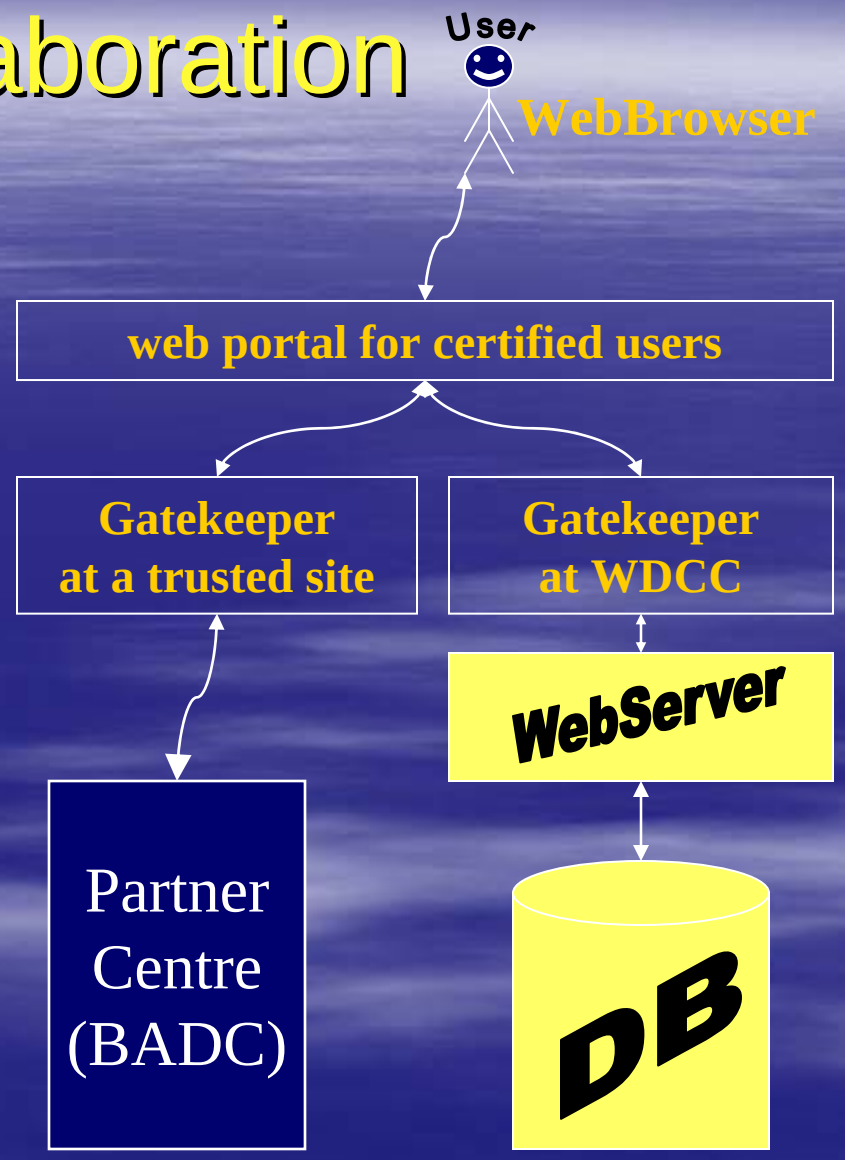
- German Aerospace Centre (DLR)
- CEOP in situ data ??

CERA-2: Certificate Based Collaboration

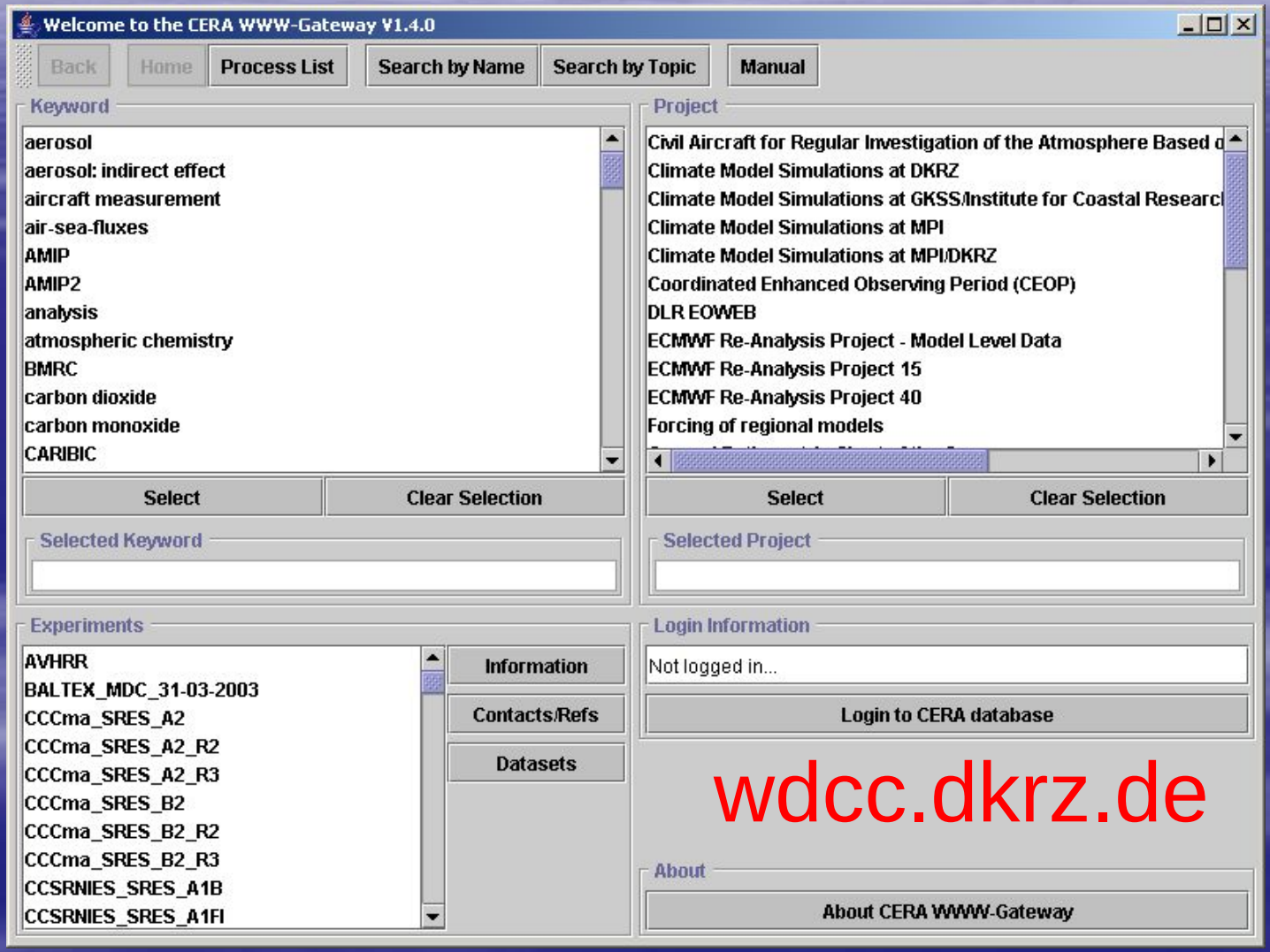
Certification of users

Access for users from trusted authorities

Exchange of authorisation information

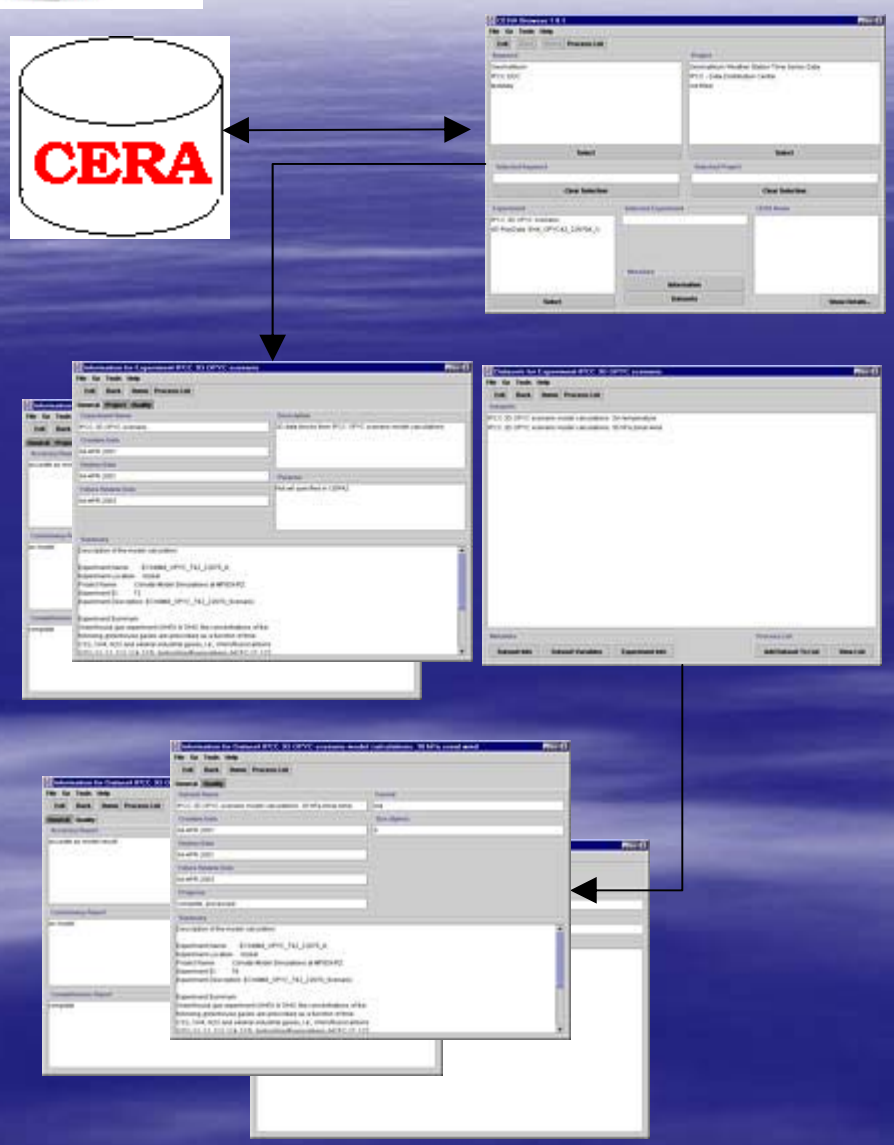


CERA-2: Applet



wdcc.dkrz.de

CERA-2: User Interface



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”
- download from tape archive to data server and to the client

0

- ?
- ?