Max-Planck-Institut für Meteorologie
Max Planck Institute for Meteorology

The CEOP Model Output Archive – Status and Perspective

WDC-Climate at M&D / MPI-M

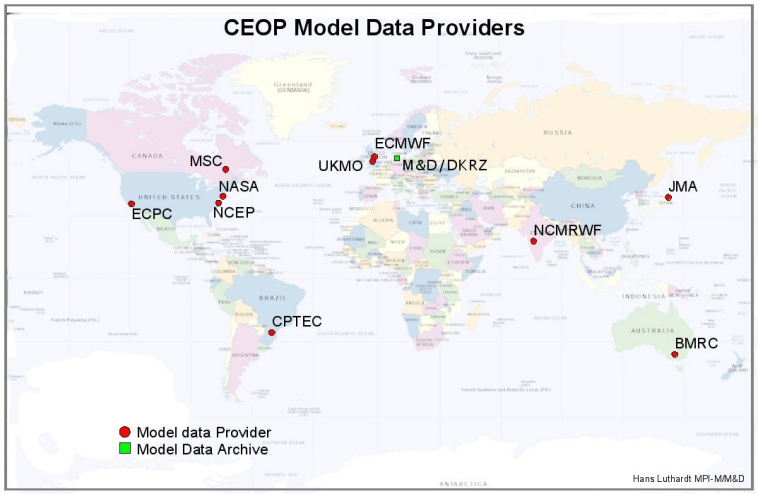
Status : February 2006

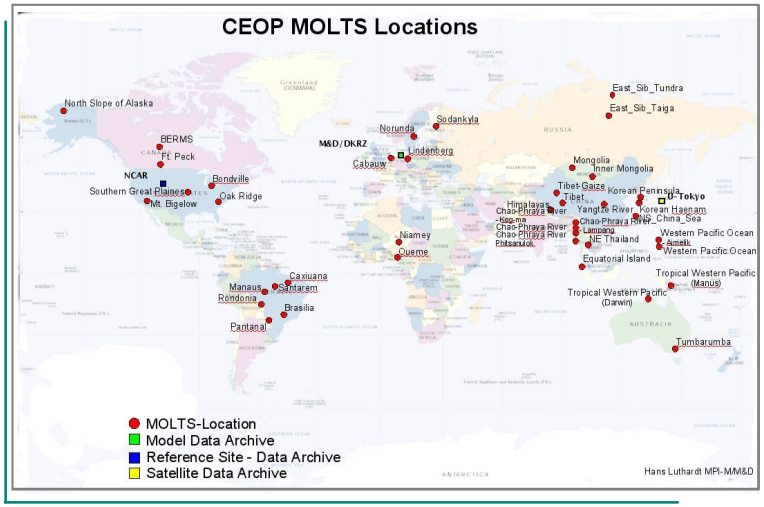
Hans Luthardt, Michael Lautenschlager, Frank Toussaint

CEOP Model Output Archive
World Data Center for Climate

contains :

- gridded global fields from numerical weather prediction centers





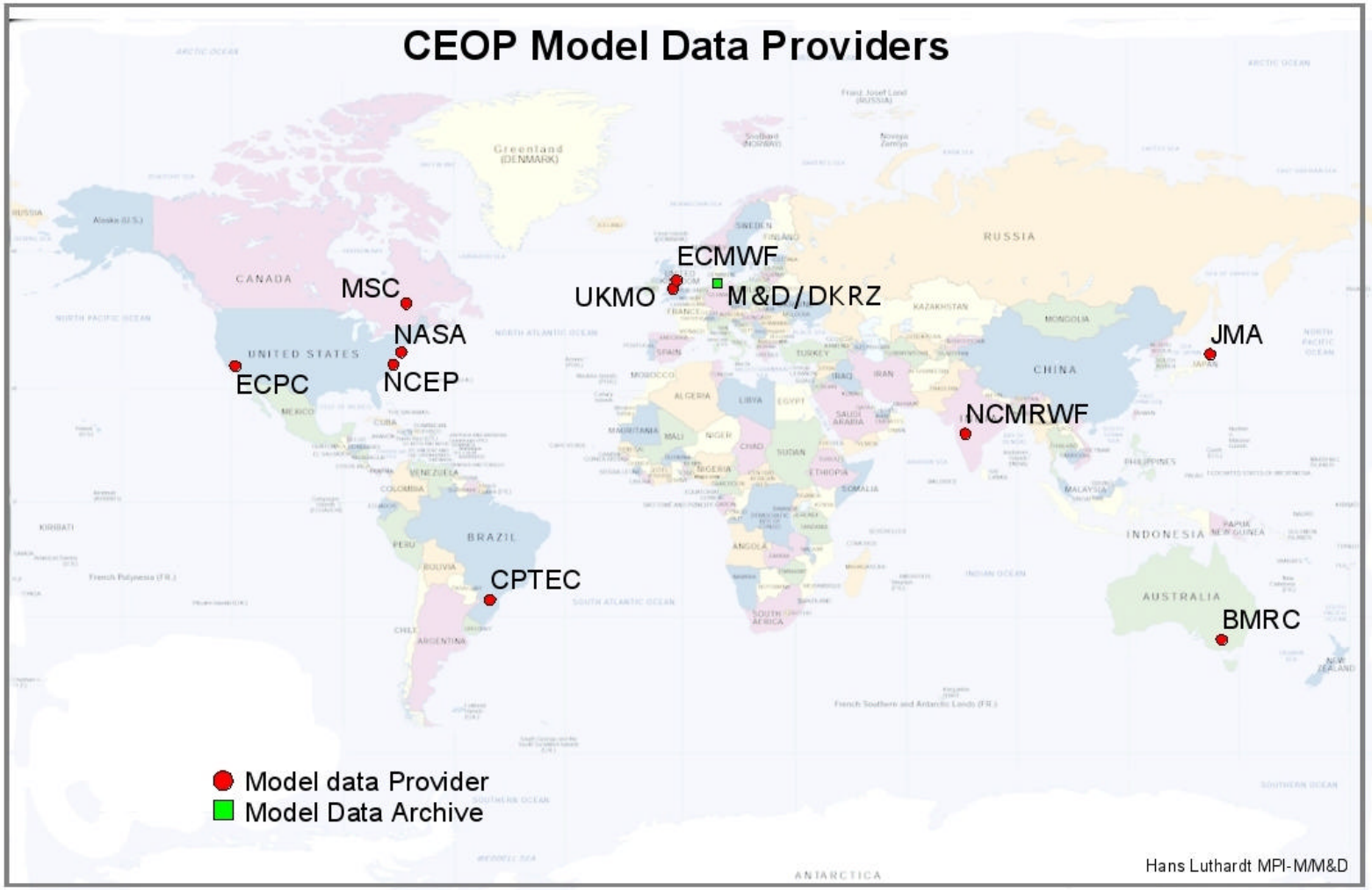
- time series for specific locations (MOLTS)

Storage of CEOP model data

Gridded data:

- Stored as binary large objects (blobs) in the CERA[1] data base
- 'As is' granularity (defined by the data providers)
- Time increment 6 h or 1 day
- Usually a blob contains one time/forecast step, various variables, different levels, global fields ->4-D time series
- GRIB data format
- Blob-size : between 400 KByte and 18 MByte
- Size of a time series : up to 22 Gbyte
- Therefore each data retrieval allows for file sizes between 400 KB (1 blob) and 22 Gbyte (complete TS)

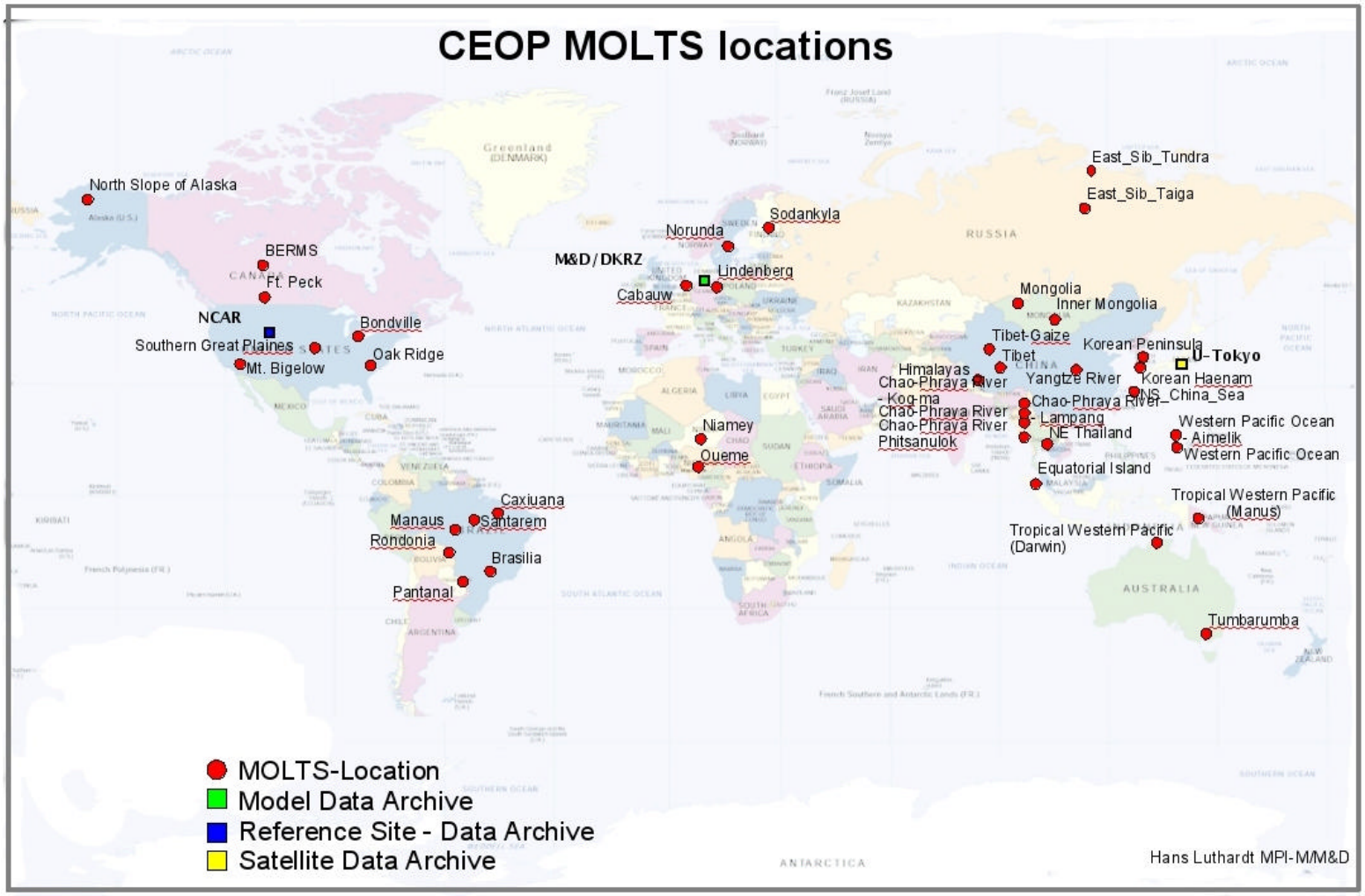
[1] Climate and Environmental data Retrieval and Archiving


CEOP Model Data Providers
ECMWF
UKMO
M&D/DKRZ
MSC
NASA
ECPC
NCEP
JMA
NCMRWF
CPTEC
BMRC
Model data Provider
Model Data Archive
Hans Luthardt MPI-M/M&D

Storage of CEOP model data

MOLTS data:

- As binary objects (blobs) in the CERA data base
- Split up into individual station data
- Usually a blob contains one time/forecast step, various variables, different levels, individual station -> 2-D time series
- Time increment 6 h or 1day
- various data formats (ASCII, binary, netCDF)
- Blob size : between 5 KByte and 2 MByte
- Size of a time series : up to 2 Gbyte
- Therefore each data retrieval allows for file sizes between 5 KB (1 blob) and 2 Gbyte (complete TS)


CEOP MOLTS locations
North Slope of Alaska
BERMS
Ft. Peck
NCAR
Southern Great Plaines
Bondville
Oak Ridge
Mt. Bigelow
M&D/DKRZ
Norunda
Sodankyla
Lindenberg
Cabauw
East_Sib_Tundra
East_Sib_Taiga
Mongolia
Inner Mongolia
Tibet-Gaize
Tibet
Himalayas
Chao-Phraya River - Kog-ma
Chao-Phraya River
Chao-Phraya River - Lampang
Phitsanulok
NE Thailand
Yangtze River
Korean Peninsula
U-Tokyo
Korean Haenam
KNS_China_Sea
Western Pacific Ocean - Aimelik
Western Pacific Ocean
Equatorial Island
Tropical Western Pacific (Manus)
Tropical Western Pacific (Darwin)
Tumbarumba
Niamey
Oueme
Caxiuana
Manaus
Santarem
Rondonia
Brasilia
Pantanal
MOLTS-Location
Model Data Archive
Reference Site - Data Archive
Satellite Data Archive
Hans Luthardt MPI-M/M&D

Storage of CEOP model data

MOLTS data :

Additional data processing :

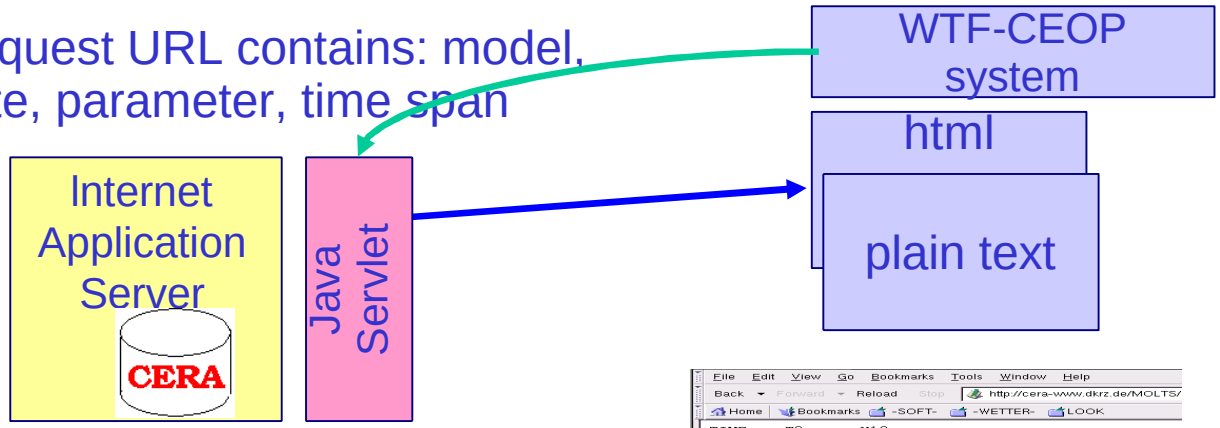
- homogenizing the MOLTS data due to format (netCDF) and code (netCDF-CF standard) (*under way --> GKSS*)
- storing the data (as single values) in data base tables (for WTF-CEOP access)

MOLTS Data Access for WTF-CEOP

WDC-Climate
MPI-M
Hamburg

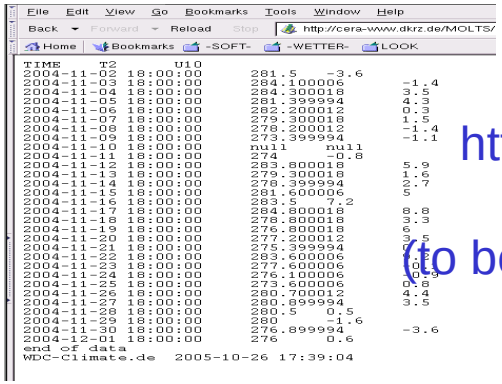
WTF-CEOP
RESTEC
Tokyo

request URL contains: model, site, parameter, time span



Data access via WWW:

- URL parsed by servlet, query: direct DB access
- 100 GB online - good performance

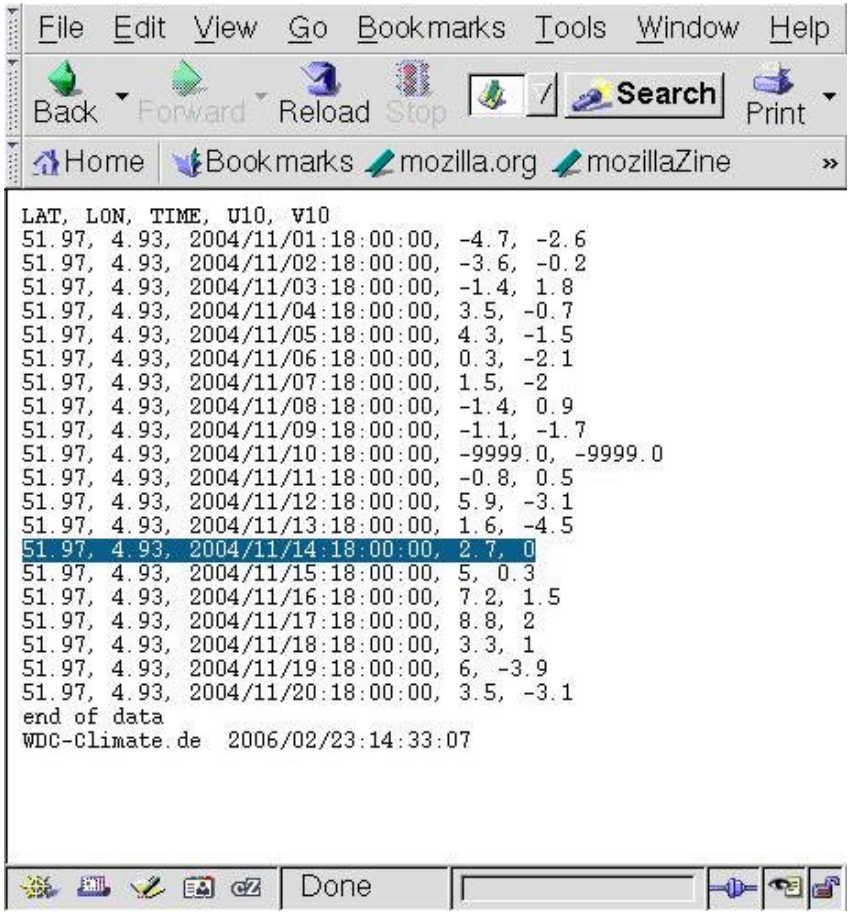


```
TIME        T2          U10
2004-11-02 18:00:00       281.5      -3.6
2004-11-03 18:00:00       284.100006      -1.4
2004-11-04 18:00:00       284.300018      3.5
2004-11-05 18:00:00       281.399994      4.3
2004-11-06 18:00:00       282.200012      0.3
2004-11-07 18:00:00       279.300018      1.5
2004-11-08 18:00:00       278.200012      -1.4
2004-11-09 18:00:00       273.399994      -1.1
2004-11-10 18:00:00       null      null
2004-11-11 18:00:00       274       -0.8
2004-11-12 18:00:00       283.800018      5.9
2004-11-13 18:00:00       279.300018      1.6
2004-11-14 18:00:00       278.399994      2.7
2004-11-15 18:00:00       281.600006      5
2004-11-16 18:00:00       283.5       7.2
2004-11-17 18:00:00       284.800018      8.8
2004-11-18 18:00:00       278.800018      3.3
2004-11-19 18:00:00       276.800018      6
2004-11-20 18:00:00       277.200012      3.5
2004-11-21 18:00:00       275.399994      0.6
2004-11-22 18:00:00       283.600006      6.1
2004-11-23 18:00:00       277.600006      [illegible]
2004-11-24 18:00:00       276.100006      [illegible]
2004-11-25 18:00:00       273.600006      0.8
2004-11-26 18:00:00       280.700012      4.4
2004-11-27 18:00:00       280.899994      3.5
2004-11-28 18:00:00       280.5       0.5
2004-11-29 18:00:00       280       -1.6
2004-11-30 18:00:00       276.899994      -3.6
2004-12-01 18:00:00       276       0.6
end of data
WDC-Climate.de  2005-10-26 17:39:04
```

http response: table (to be visualized)

MOLTS Data Access for WTF-CEOP

http://cera-www.dkrz.de/MOLTS/SelectMOLTS?
t=molts_ncep_cabauw_06&
c=LAT,LON,TIME,U10,V10&
r=date_time%28%222004/11/1:13
%22,%222004/11/21:06:00:00%22%29

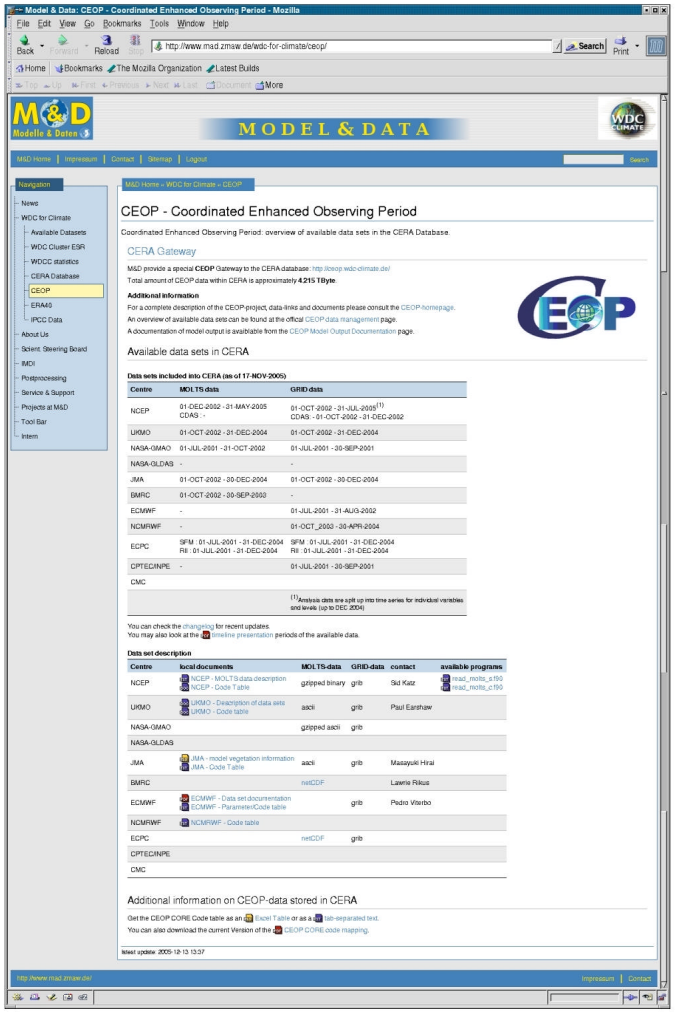


Statistics

Total:

- Size of CEOP data in WDC-Climate: **4.324 TByte** in Feb. 2006 (accessible by internet)
- Number of data downloads : **730 000** (total)

Web Access to CEOP model data

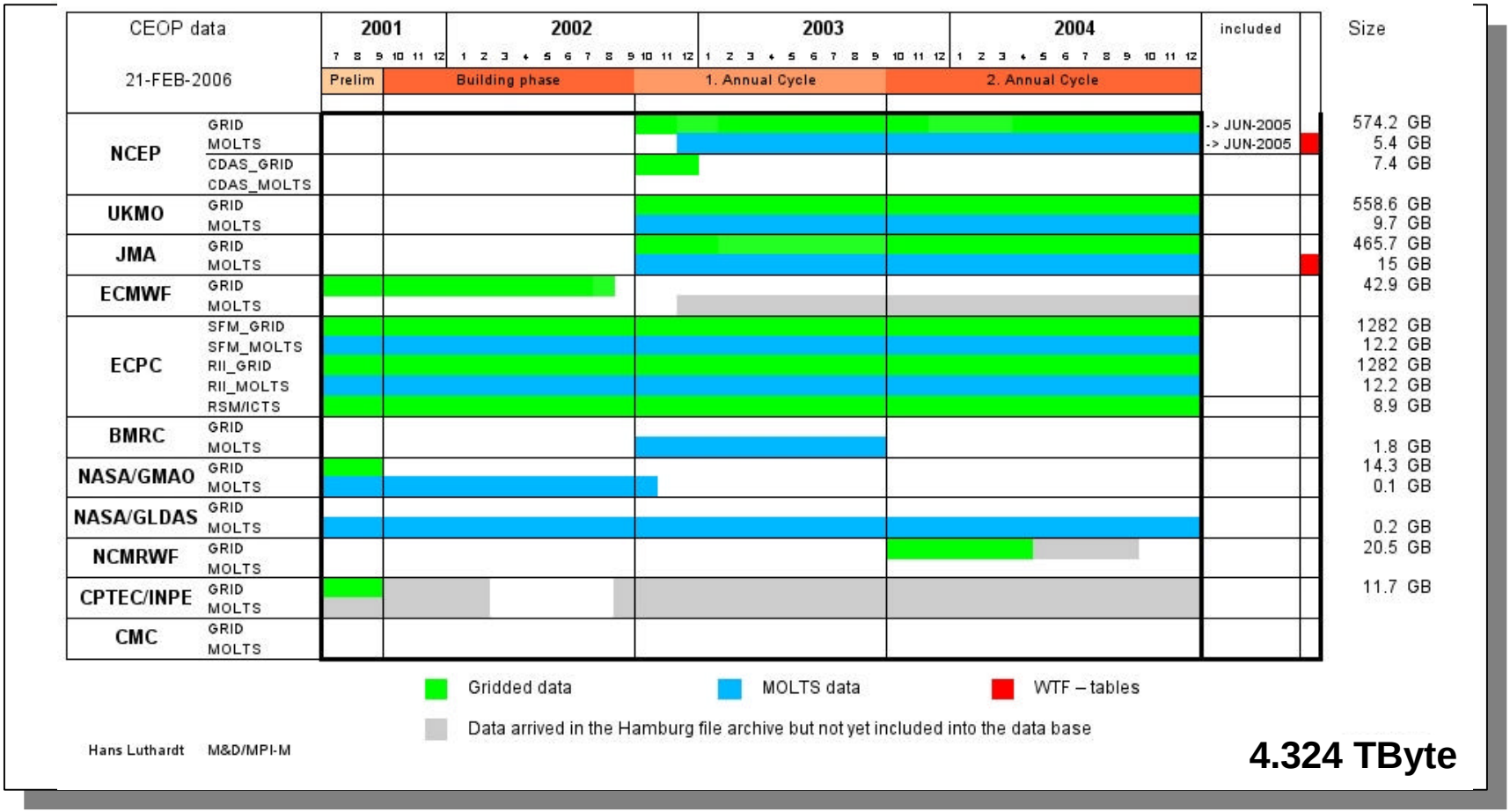


Access by Internet (web browser) to the CERA Gateway
http://ceop.wdc-climate.de/
http://www.mad.zmaw.de/wdc-for-climate/ceop/
or
by java command line tool ('jblob')

CERA Gateway page for CEOP also provides summary information (e. g. time line representation of available data) on the data base content and data descriptions

Model Data Content

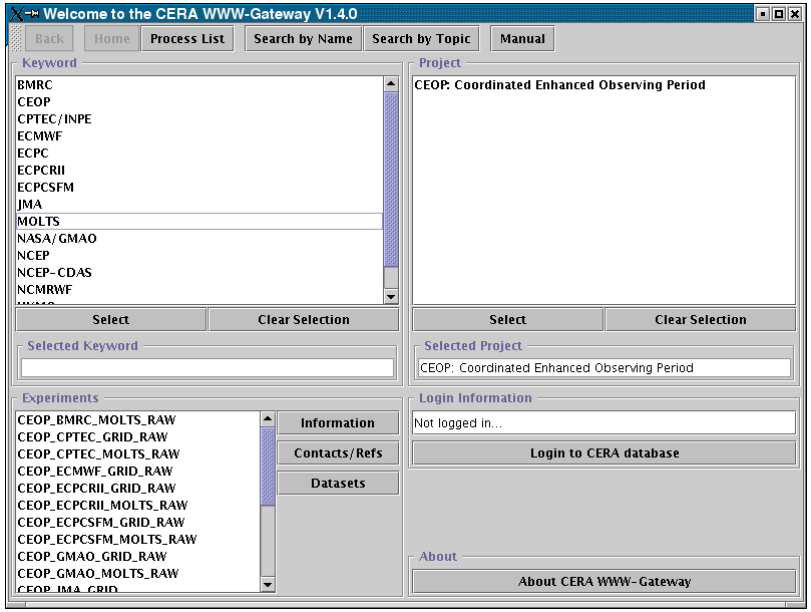
(as of 16.2.2006)



The most recent version of the data status (time line) can be found from:

http://www.mad.zmaw.de/wdc-for-climate/ceop/

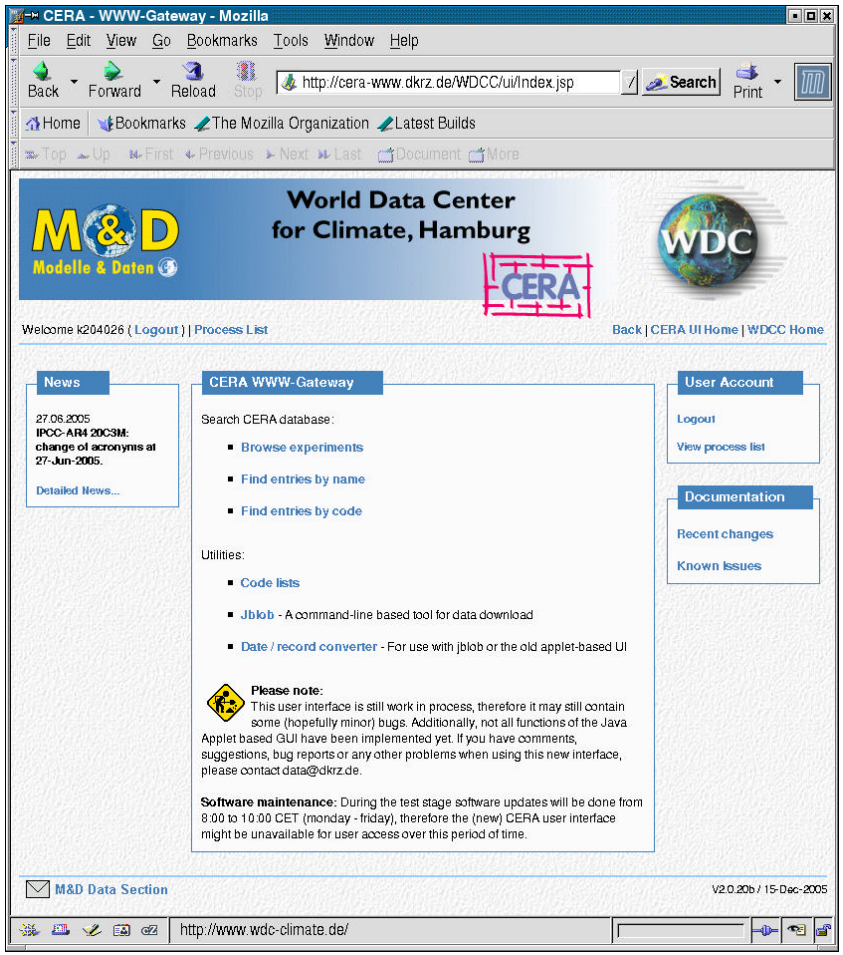
Graphical User Interface to WDC-Climate



GUI to CERA:

- java applet
- free access to all meta data
- download of public data after login
- free user account (basic registration necessary)
- download of specific data on request

New Graphical User Interface to WDC-Climate



GUI to WDCC:

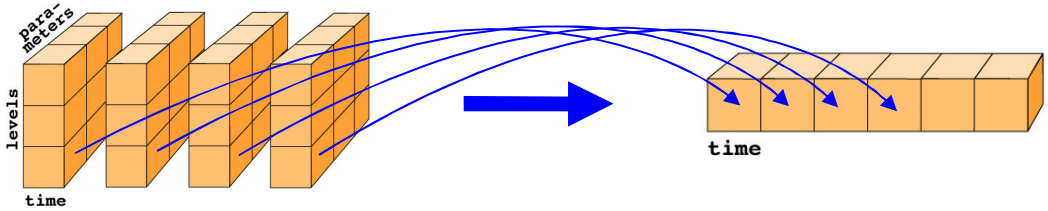
- java servlet
- will reduce firewall problem for data download
- provides additional functionality for grib formatted data (cutting out regions, conversion into ASCII format)
- Export of meta data in HTML or XML format

http://cera.wdc-climate.de

Perspectives

- Pursuit of hommogeneity of data structure, data content, formats, meta data , ...
- MOLTS also in netCDF/CF

- 2-dimensional storage of gridded data, (single variable / single level time series) (as e.g. Analysis data from NCEP & JMA)



original data structure (4-D)
(3 MByte / blob)

Application related data structure (2-D)
(180 KByte/blob)

- WTF conform data storage structure (MOLTS, gridded data)

Thanks

+ Experiences

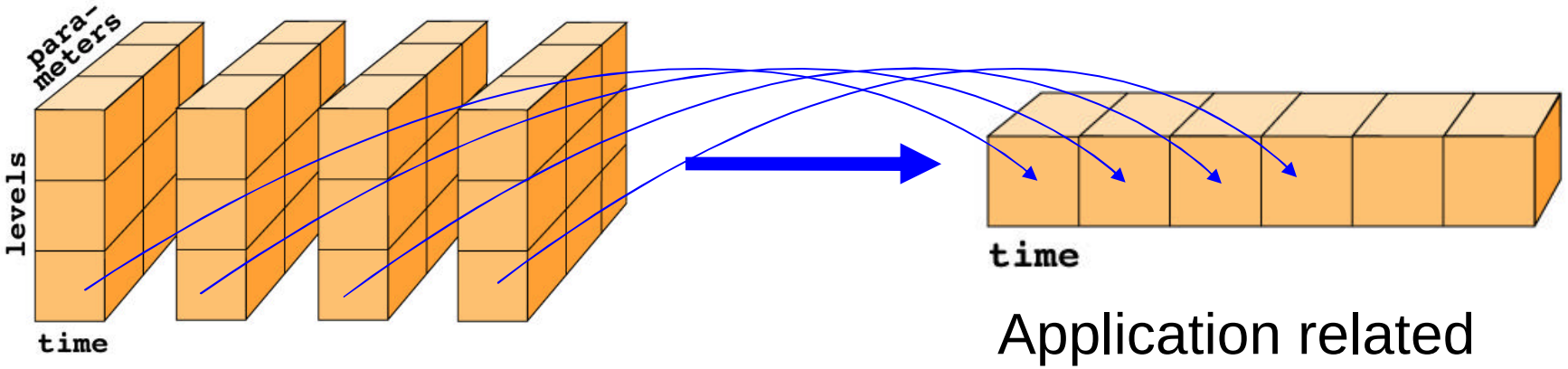
- Data transfer (data providers -> MPI) by internet works fine
- Data volumes cause no problems
- User access problems due to firewall should vanish with the new database interface (java servlets)
- Command line tool available
- Meta data available from the data base
- User access control by user accounts
- WTF-access works fine (requires additional processing efforts)
- Postprocessing capability (format conversion, cut out regions for grib data) implemented (only 2-dim grib data)
- NetCDF file format supported
- Conversion of MOLTS data into netCDF format (using CF standard variable descriptions) is under way (GKSS/BMRC)

- Experiences

- Original data very inhomogeneous in structure and format
- Individual variable selection and code tables
- Heterogeneous meta data informations provided
- Large data blocks (gridded data)
- 'Zoo' of formats for MOLTS data
- Considerable processing effort is required for DB-storing the data even in the (near) original structure
- Postprocessing needs for application related storage of the data requires (too many) personal resources

Potential improvements

- Standard fromats and data structures for data provision
- Unified metadata provision
- Unified code table
- Finer granularity of the data
- Standard formats (grib and/or netCDF)
- Semi-automatice data transfer and inclusion into the Model Data Archive
- . . .



original data structure
(3 MByte / blob)

Application related
data structure

(180 KByte/blob)