

Bulk Metadata Structures in CERA

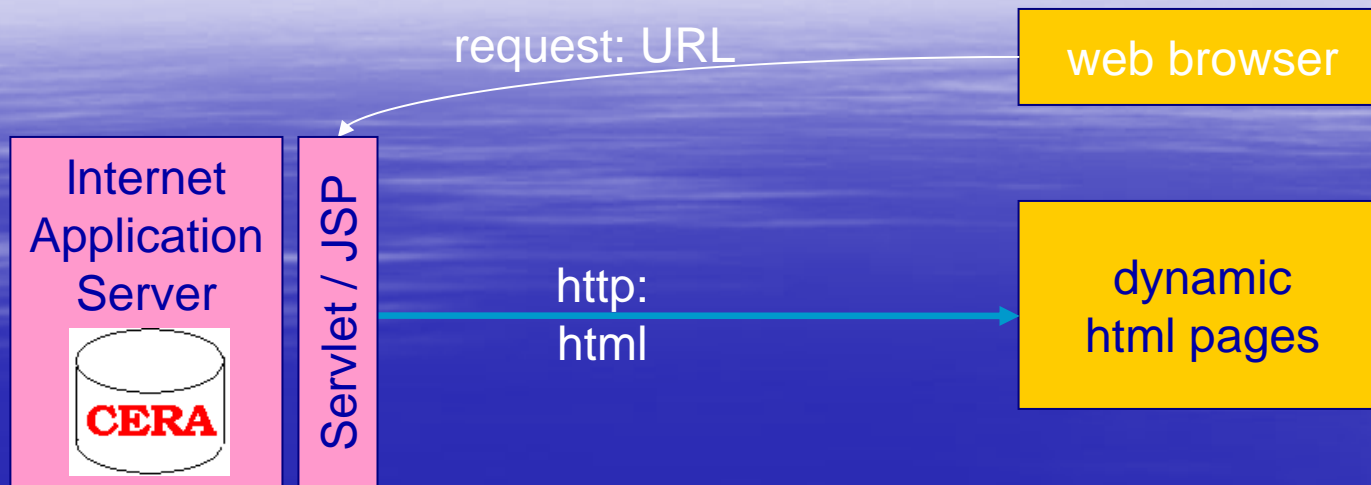
Frank Toussaint, Michael Lautenschlager

Max-Planck-Institut für Meteorologie
World Data Center for Climate

Contents

- Present activities
- General vs. specific metadata
- Present structure at WDC-Climate
- Structural changes in CERA
- Data flows
- Pros and cons of this mode
- What do we gain?

New Catalogue Access



Catalogue access via WWW

- URL parsed by JSP
- integrated DB retrieval by JSP
- response in standard html
- efficient administration of detailed meta information

The screenshot shows the World Data Center for Climate, Hamburg website. The main content area displays "CERA entry information for ERA40_PL00_6H". The page includes a table with the following data:

Variable	Value
Topic	oceans
Parameter	sea/ice
Variable	divergence
Variable description	divergence

Below the table, there is a section for "Code" with the following details:

Number	155
Type	ERA40
Acronym	D
Description	divergence

The "Unit" section shows:

Acronym	1/s
Name	1/second
Description	not filled

The "Aggregation" section shows:

Description	6 hours interval
-------------	------------------

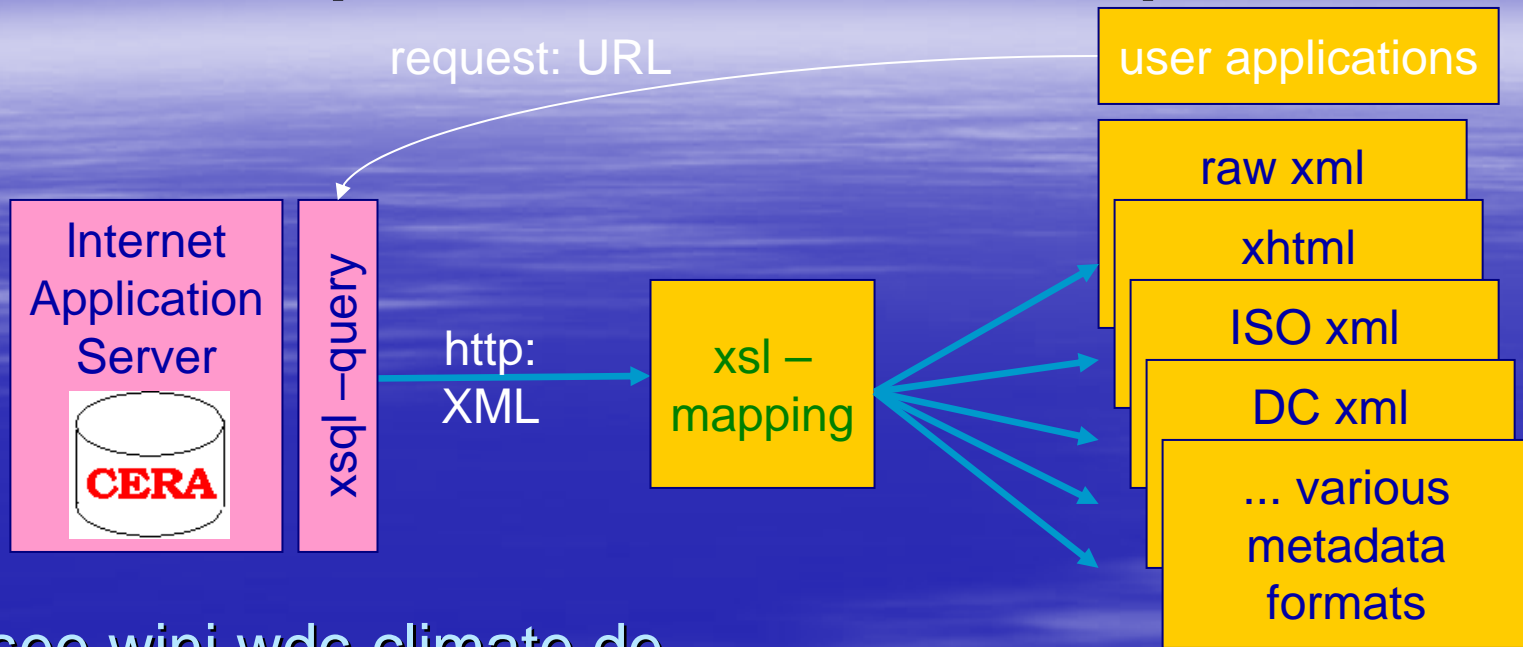
The "Temporal structure (period 1 of 1)" section shows a table with the following data:

	Start	Increment	End / Same information
Year	1957	0	Period start: 01/09/1957 00:00
Month	-9	0	Period end: 31/08/2002 16:00
Day	1	0	
Hour	0	6	
Minute	0	0	

The "Additional time information" section shows:

Description	real time
Type	

http Metadata Output



see wini.wdc-climate.de

Metadata access via WWW:

- xsql query to DB
- xml output from DB by integrated servlet
- xsl mapping to any metadata format

```
- <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" codeListValue="data"
      </role>
    </CI_ResponsibleParty>
  </contact>
```

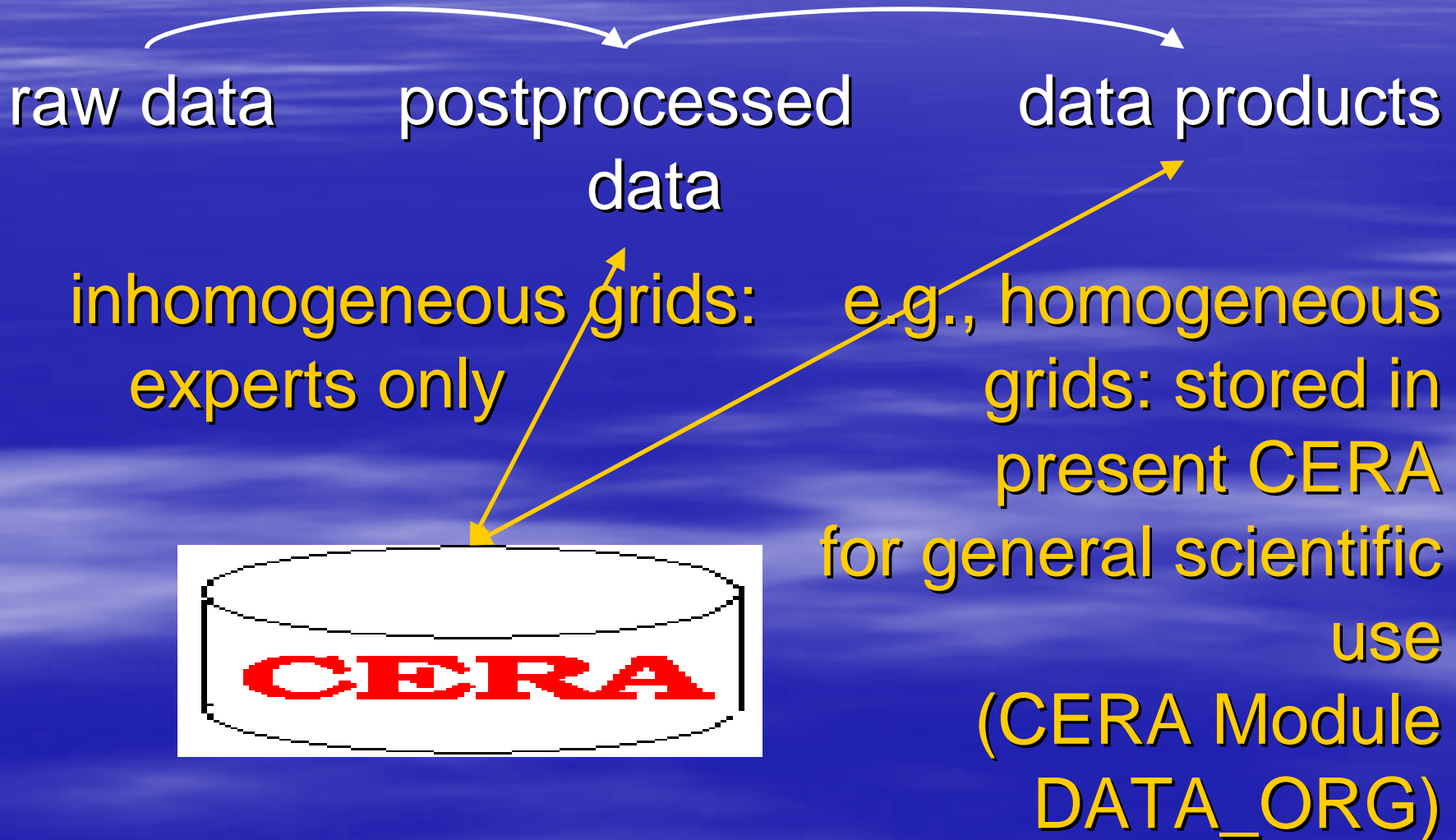
Types of Metadata

MD type	General/Catalogue MD	Specific/Use MD
properties	canonical, general	branch specific, detailed
use	browse, search, retrieval	interpretation, processing
content	title, contacts, dates, space-time coverage...	grids, setups, platform/sensor...
complexity	low diversity relatively high stability	high diversity low stability
user	catalogue visitors (all scientists)	user of the data (branch specific)

Forms of Specific Metadata

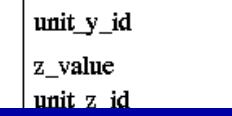
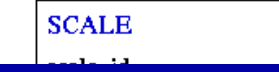
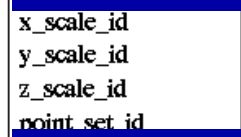
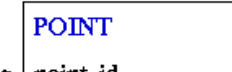
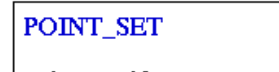
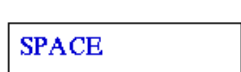
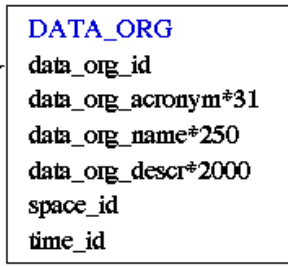
- grib headers with/without code tables
 - NetCDF(-CF) headers
 - xml files – structure definitions in xsd
- ...in addition:
- hand written notes
 - programme inline comments
 - phone calls

Present Structure at WDCC



PARAMETER

data_org_id



CERA Modules

5 Modules (3 in use):

- DATA_ACCESS for automatted data access
- DATA_ORG organization of grid data
- CODE model code numbers

1 submodule

8 ENTITIES

1 LISTS OF VAL

Submodule PATCHED_DIM

to allow for patchwise equidistant gridded

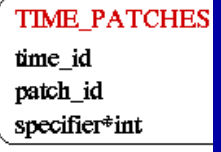
SCALE

scale_id



TIME

time_id

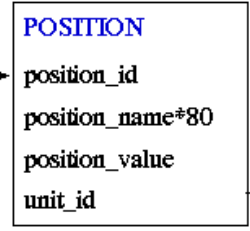
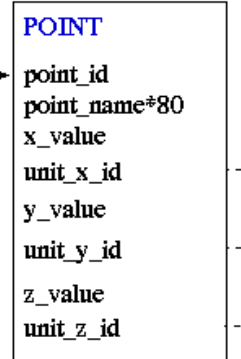
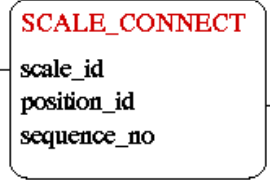
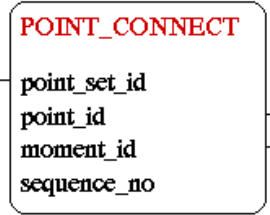
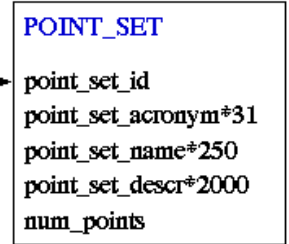
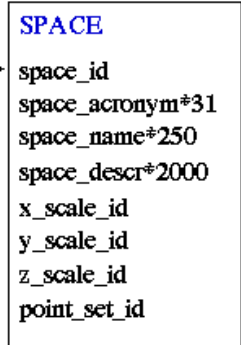
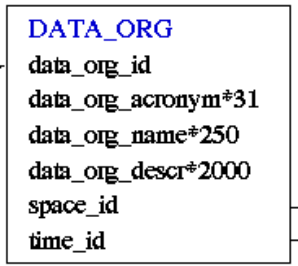


2 RELATIONS

CERA Module DATA_ORGANIZATION, Version 1.3

PARAMETER

data_org_id

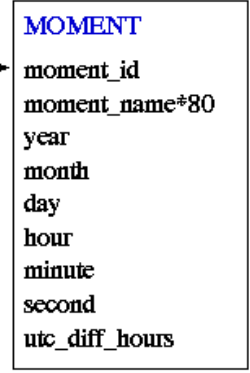
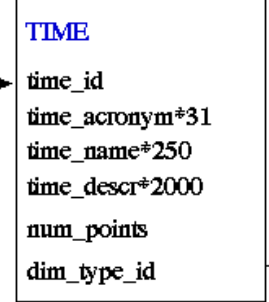
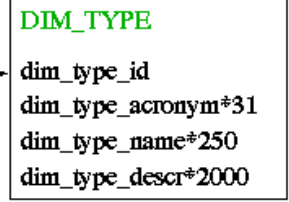


8 ENTITIES

1 LISTS OF VALUES

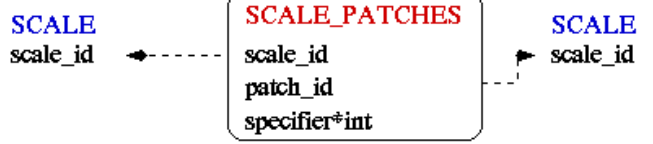
3 RELATIONS

Dimension types (e.g.):
 equidistant grid
 irregular grid
 single position(=projection)
 spectral
 -- other --



Submodule PATCHED_DIMENSION 1.0

to allow for patchwise equidistant gridded dimensions



2 RELATIONS

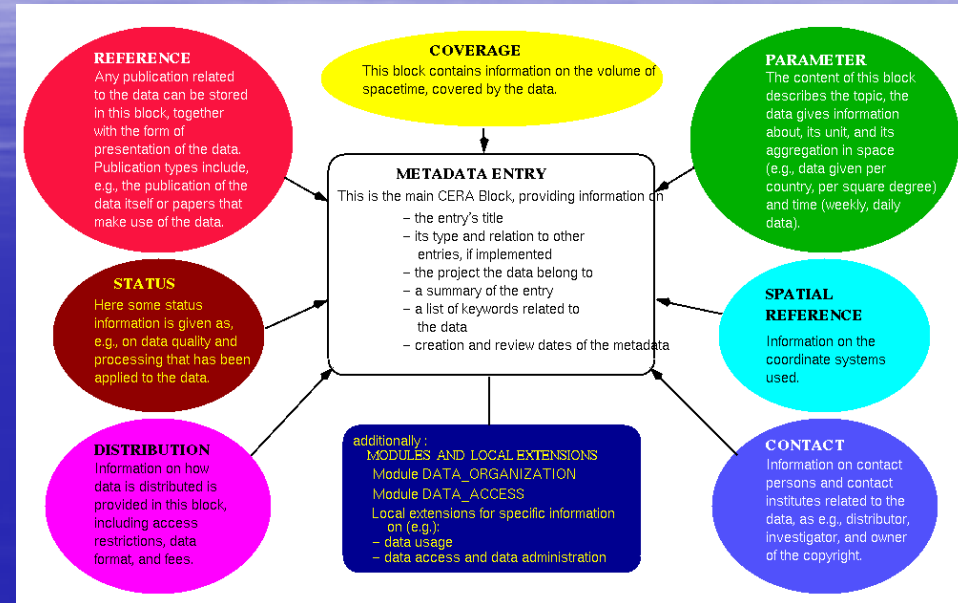
Version 1.0, 2001-03-24

Version 1.3, 2001-03-22

Appendix for Bulk MD: XML and other

Appendix for bulk data

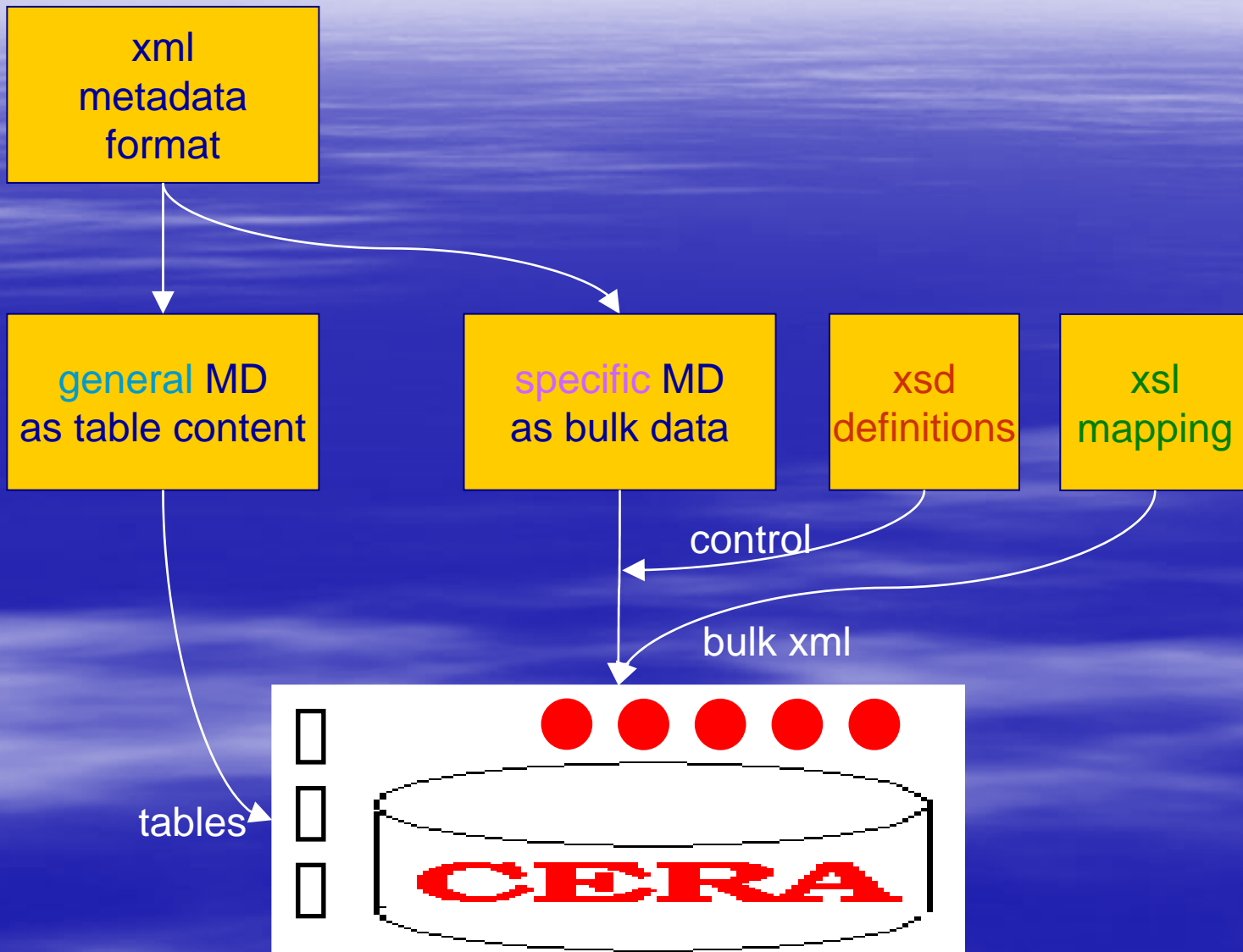
- type of appendix
incl version
- xml, xsd, xsl
techniques for
catalogue display
- txt files to view
- other formats for
download



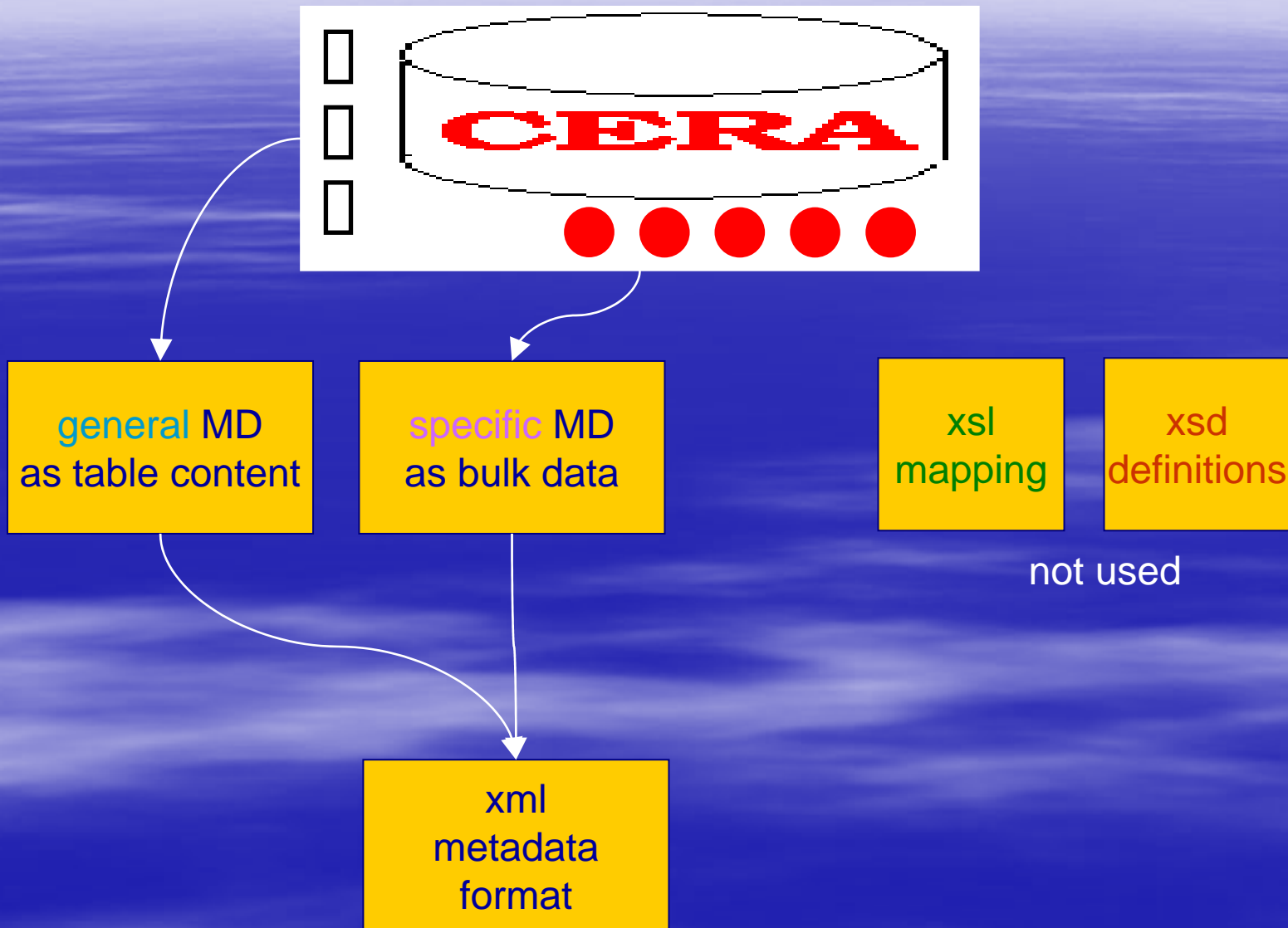
possible types:

- numerical grid description
- model/experiment description

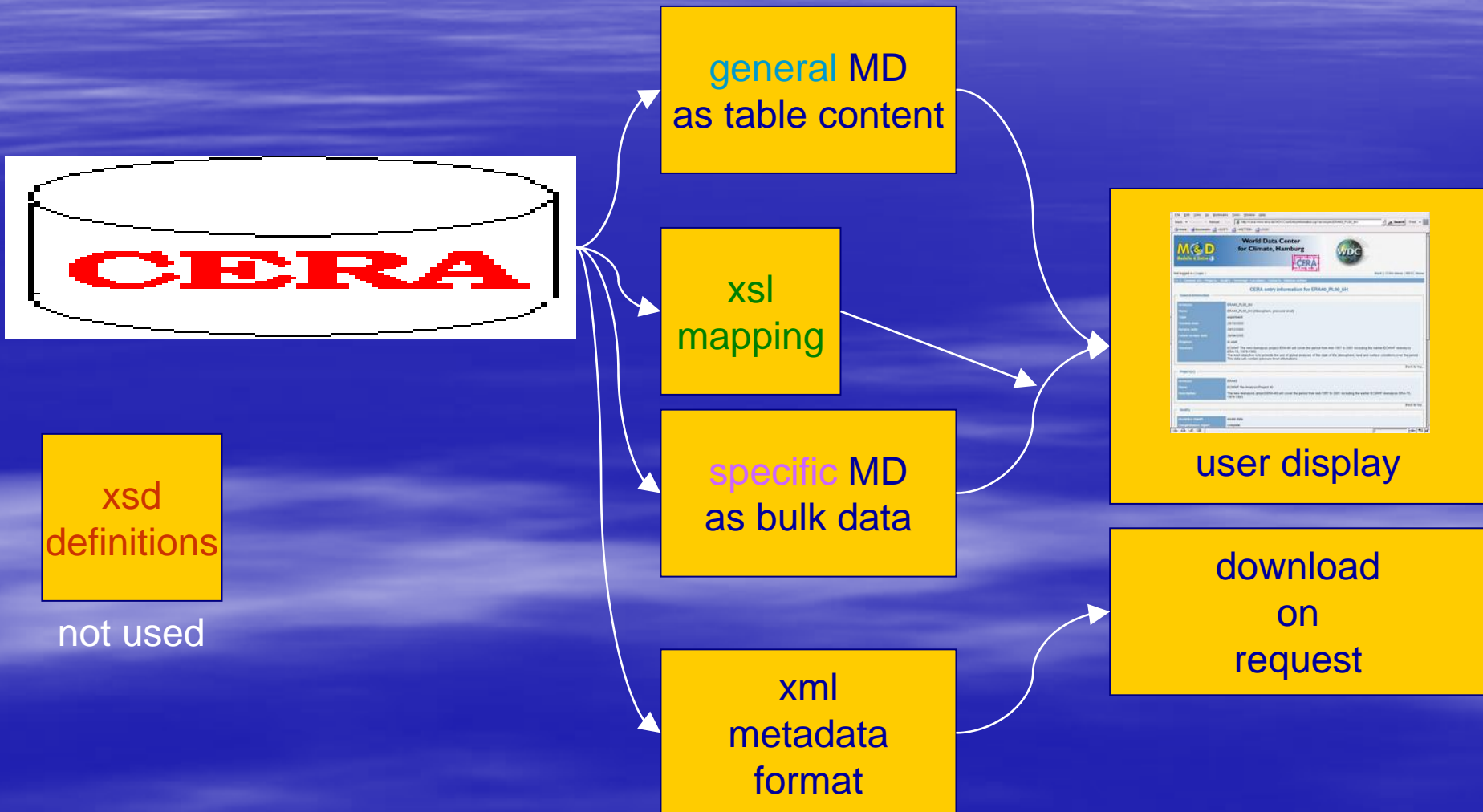
Data flows: Input



Data flows: xml Output



Data flows: Catalogue Output



Concept of Appendix for Bulk Data

Pros

- Data structure discussion decouples from data storage technique
- maximum flexibility
- easy catalogue integrated display for xml
- low effort
- access rights separate from main metadata
- stable xml structures later can be migrated to table structures

Concept of Appendix for Bulk Data

Cons

- search mechanisms on stored data are between crude and none
- ...

Data Storage Problem and Numerical Model Description Problem

- Which problems are solved by this concept ?
diversity & time changes of specific data
- Which problems are created ?
...
- Which problems persist ?
we do not yet have a structural concept...
...responsibility of scientific specialists ?

The Structures of Bulk Data

Minimum requirements

- every data bulk needs a **name, format, size**
- it may have **contact persons, access constraints ...**

Bulk metadata as XML

- **extraction & display** of defined information
- undefined data is stored but not displayed

Other bulk metadata

- text files **as name lists, source codes, run scripts, ...** → for display
- docs as pdf

Thank you!

Questions?