



EUCLIPSE@WDCC

Michael Lautenschlager, Hans Luthardt, WDCC/DKRZ

Kick-Off Meeting, 27/28 Sep. 2010 in Utrecht/NL



M.Lautenschlager, H.Luthardt



WDCC



World Data Centre on Climate

The **mission** of the WDCC corresponds with the basic tasks of M&D. **Data for climate research are collected, stored and disseminated** in order to serve the scientific community. M&D has experience in running international data centres. The WDCC is **restricted to climate data products**. ... **Emphasis is spent on climate modelling and related data products**. ...



WDCC



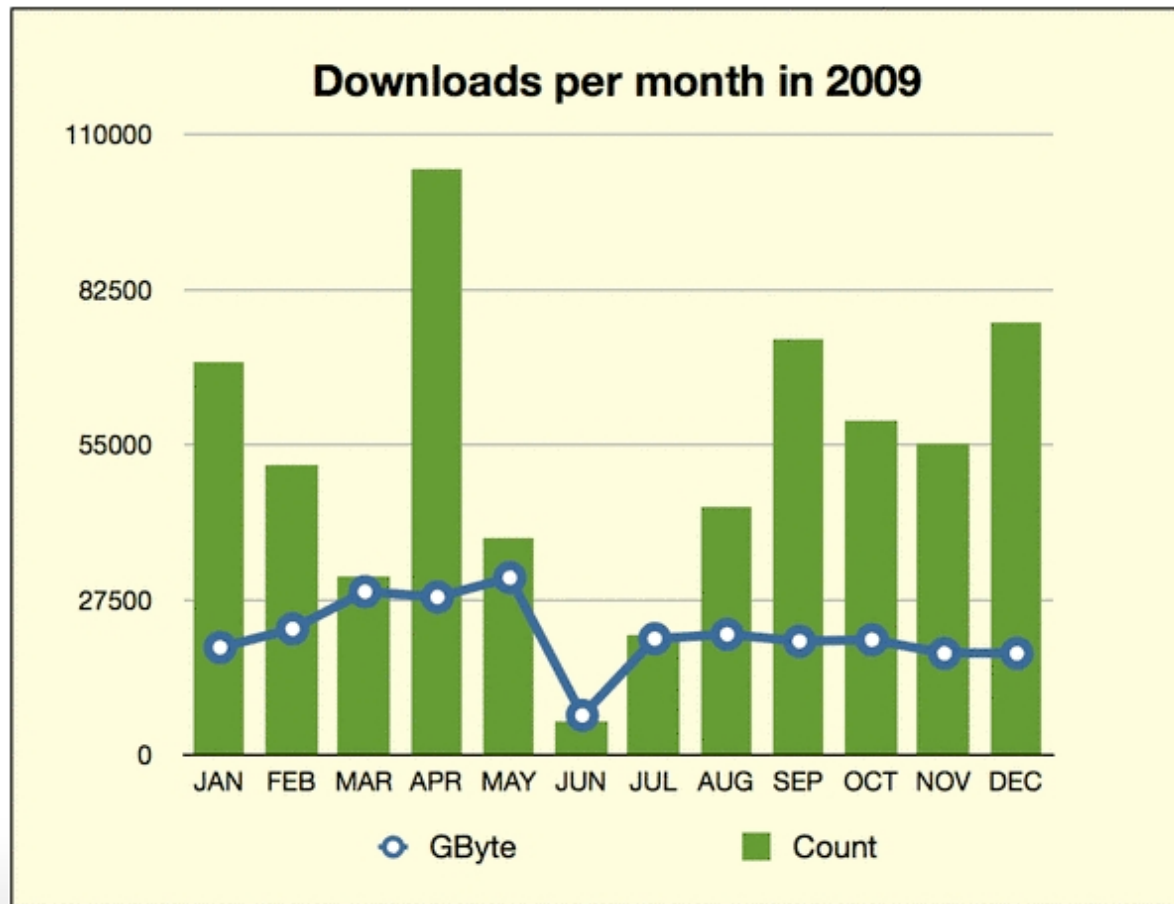
World Data Centre on Climate

- ▶ Approved in 2003
- ▶ Focus : long term archiving of Earth system model data and related observations
- ▶ Hosts several projects and Data Centres
- ▶ WDCC operates as a long-term data archive (10years +)
- ▶ WDCC is implemented within the CERA data and information system.
- ▶ Data are stored in conjunction with metadata.
- ▶ WDCC offers the publication service for primary data. (DOI)
- ▶ Approximately 5 person staff and 400 TB of data.
- ▶ Increase of a 1 PB/year starting Summer 2010

WDCC



World Data Centre on Climate

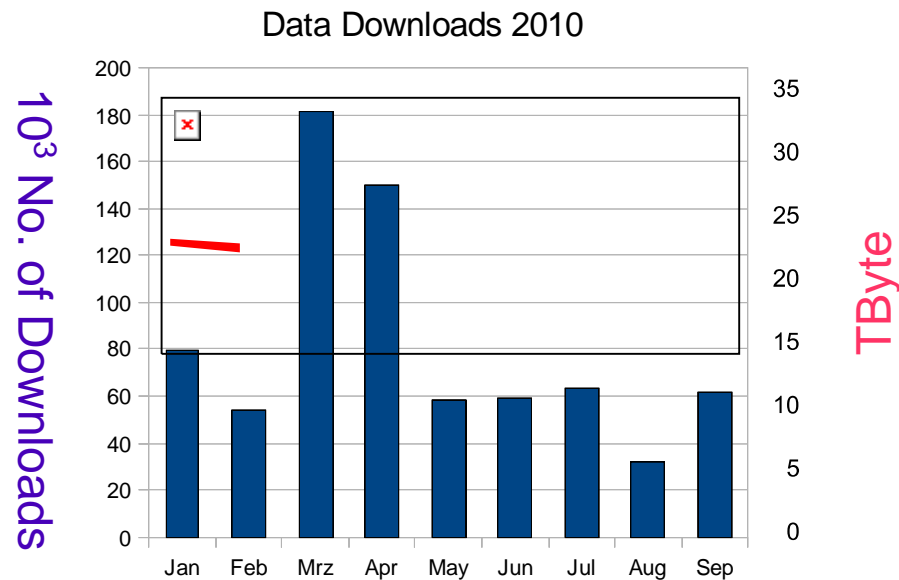


WDCC



World Data Centre on Climate

Downloads per month in 2010



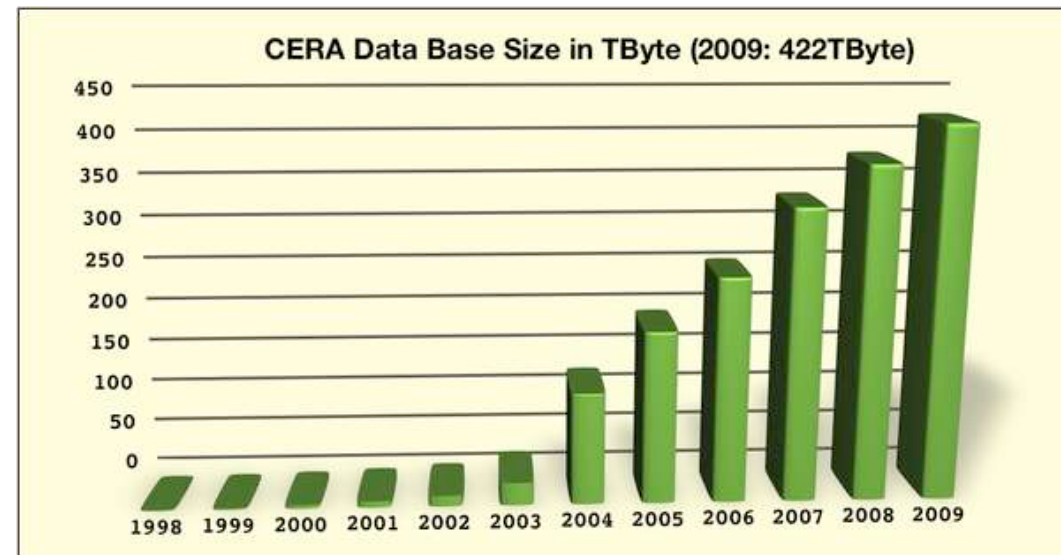
WDCC



World Data Centre on Climate

Calendar year 2009:

- ▶ 800 active users
- ▶ Data from
 - 80 projects
 - 1400 experiments
 - 170000 datasets
 - 8.5 Billion records
- ▶ more than 600000 downloads
- ▶ more than 425 TByte in total



WDCC



World Data Centre on Climate

- ▶ Most active German Projects
 - COPS
 - REMO-UBA
 - CLM Consortial Runs
 - MILLENNIUM_COSMOS
- ▶ Anticipated projects
 - CMIP5
 - IPCC AR5
 - Global and Regional
 - STORM
 - EUCLIPSE
 - And many more
- ▶ Most active International Projects
 - CEOP
 - ENSEMBLES
 - DPHASE
 - Metafor
 - IS-ENES
 - IPCC

WDCC



World Data Centre on Climate

▶ Most active German Projects

- COPS
- REMO-UBA
- CLM Consortial Runs
- MILLENNIUM_COSMOS

▶ Anticipated projects

- CMIP5
- IPCC AR5
 - Global and Regional
- STORM
- EUCLIPSE
- And many more

WDCC

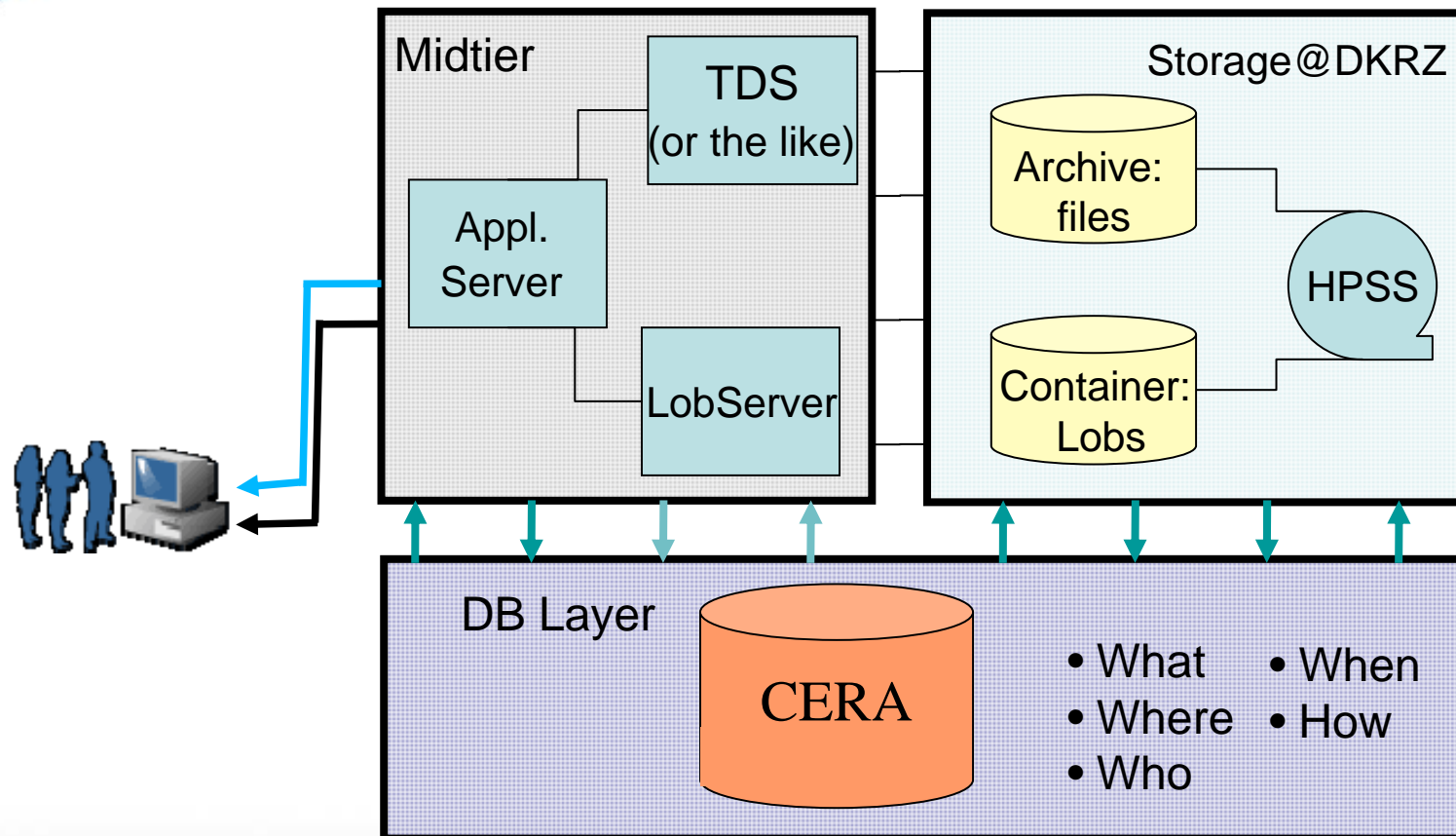


World Data Centre on Climate

▶ Most active International Projects

- CEOP
- ENSEMBLES
- DPHASE
- Metafor
- IS-ENES
- IPCC

WDCC: Data access



Tasks with participation of DKRZ

- ▶ Task 1.3.2: EUCLIPSE model output will be **inserted** in the WDC database system. Project data curation includes **dissemination of CMIP-5 data products** to project partners and **augmentation** of the CMIP-5 database by **GCM and cloud resolving simulations from EUCLIPSE**. Data and metadata ingestion into the WDC database system will use the CMIP5 data interfaces and should therefore be of no problem for each modeling partner. WDC data **quality insurance** includes **data collection, check of metadata and data structures** against WDC ingest interfaces and exceptions handling during ingest process (metadata and data). EUCLIPSE relevant data will be **documented and disseminated via a web-based data portal**. Partners: KNMI, MPG, METO, CNRS, MF-CNRM, DKRZ

Tasks with participation of DKRZ

- ▶ Task 2.1.5: **Integration of data analysis workflows** with respect to cloud processes into the WDCC infrastructure. In addition to existing CMIP-5 data metrics **data diagnostics** which will be developed within EUCLIPSE **will be inferred** with respect to integration **into the standard data processing workflows of WDCC**. Partners: DKRZ

Data management

EUCLIPSE data as part of CMIP5:

- Interface : CMOR-2 for NetCDF3/4 and METAFOR/CIM questionnaire for metadata
- data available via Earth System Grid Federation (PCMDI, BADC, WDCC/DKRZ)
- data access via IS-ENES data portal

EUCLIPSE data in addition to CMIP5:

- identical interface as for CMIP5 should be used
- amount of additional data should be restricted to 50 TB (project disk space)
- identical data access as for CMIP5 should be used (IS-ENES)

Data management

ECLIPSE access and quality control:

Access and quality Control could be handled similar to the CMIP5/AR5 QC (currently under discussion, see Table 1) :

- restricted access during quality control process, control by protect access or PCMDI
- free academic access after finalization of quality control, access control by WDCC according to rules of ICSU WDC system
- CMIP5/IPCC-AR5 QC L3
freezing of data, assigning of persistent identifier, fixed citation reference

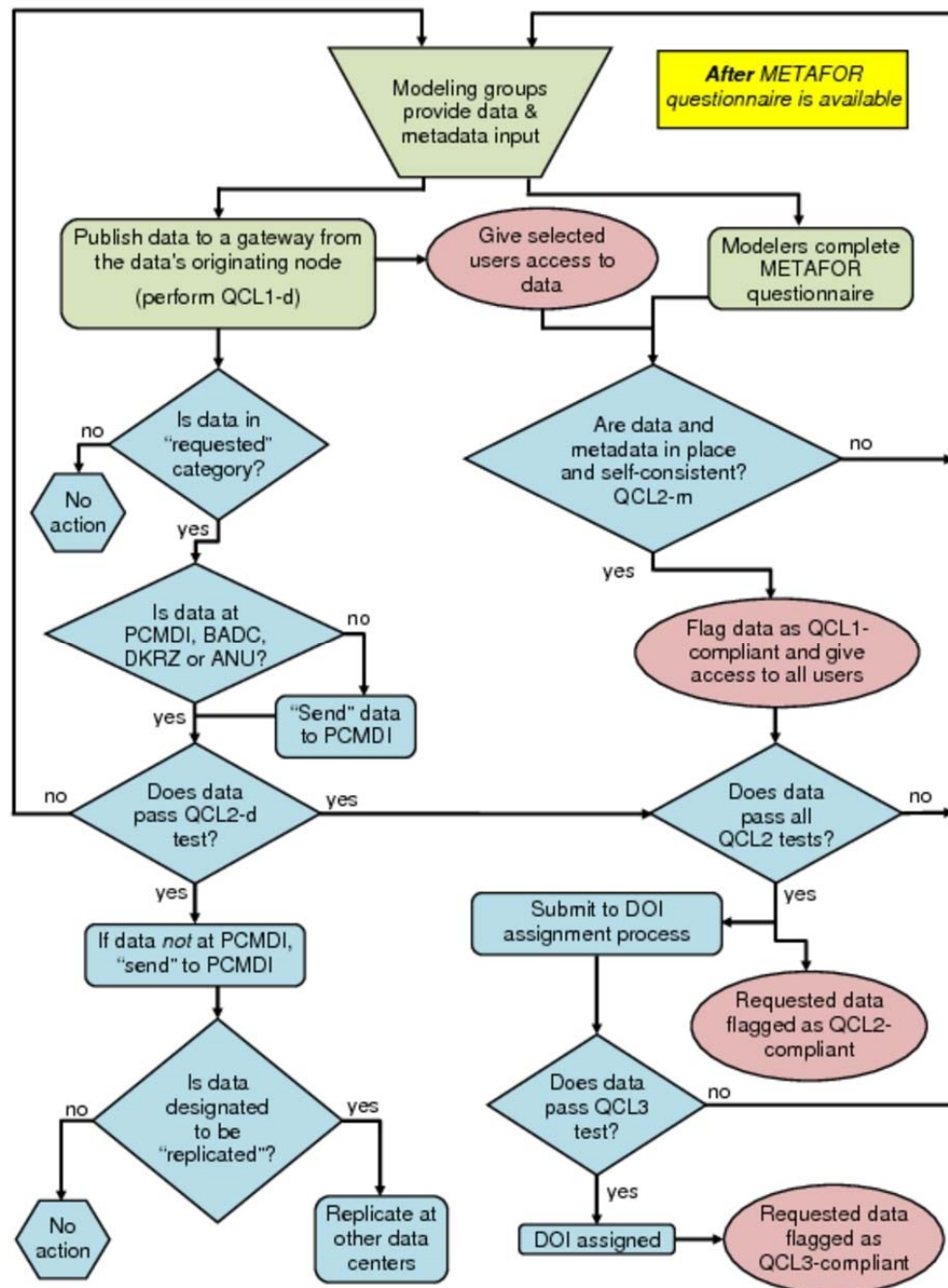
Data management

Quality control (under discussion):

	QC Level 1: CMOR2, ESG, CIM Conformance	QC Level 2: WDCC Conformance and subjective controls	QC Level3: STD-DOI Publication
Data	preliminary; no user notification about changes; performed for all data	preliminary; no user notification about changes; performed for core data	published and persistent data with version and unique STD-DOI citation; performed for core data
Access	constrained to data author (modeling center)	constrained to CMIP5 members and IPCC WG I community	open for IPCC process (WG I – III) and research community
Access Control	PCMDI on the behalf of WMO/WGCM	PCMDI on the behalf of WMO/WGCM	IPCC-DDC on the behalf of TGICA
Citation	no citation reference	preliminary citation reference	final citation reference
Quality Flag	“automated conformance checks passed”	“subjective quality control passed”	“approved by author” (in case of newer DOI available: “approved by author, but suspended”)

Table 1: CMIP5/AR5 Quality Control Levels.

Quality Control



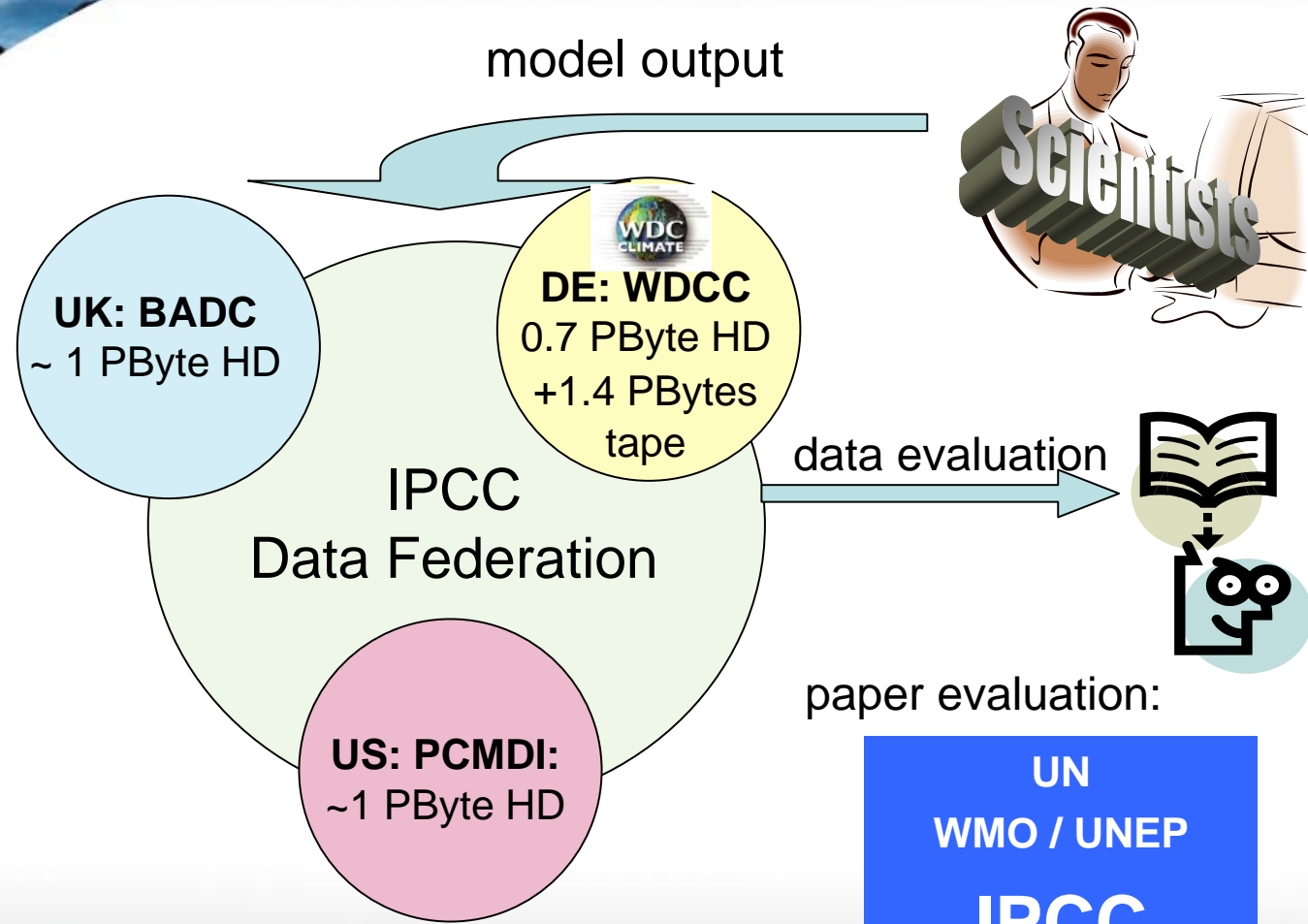


Data management

Working hypothesis for EUCLIPSE data management:

EUCLIPSE relevant data should be available on hard disks locally at WDCC/DKRZ using the CMIP5/IPCC-AR5 data interfaces for ingestion and access

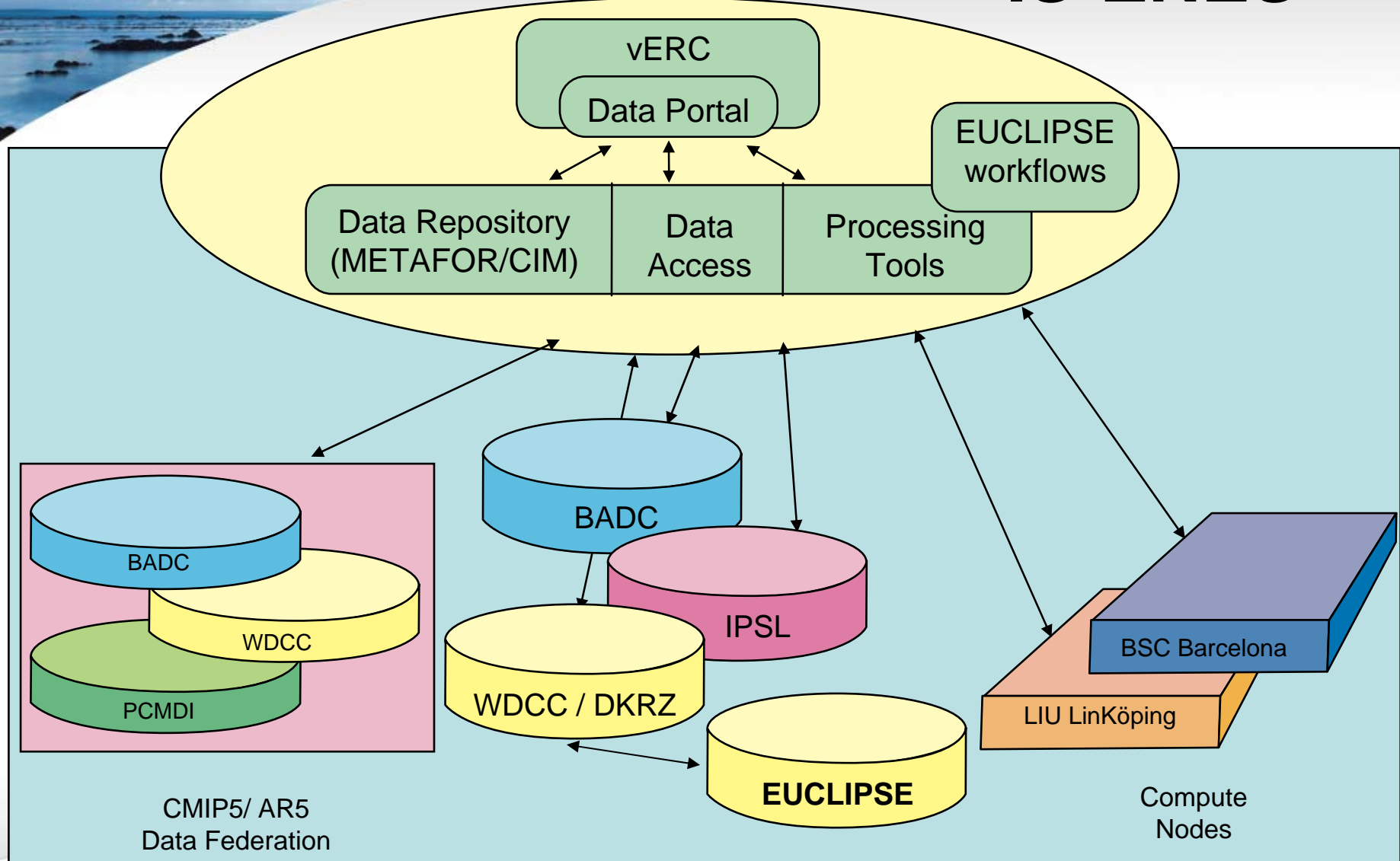
WDCC as IPCC / CMIP5 Data Node



- Interfaces:**
- CMOR-2/ NetCDF3
 - METAFOR CIM questionnaire for meta data

ESM Data Network

IS-ENES



M.Lautenschlager, H.Luthardt

EUCLIPSE Hardware

- This year
 - 2 data node servers (x86-64)
 - 50 TByte of disk storage
- Next year
 - Some tapes for backup and long term archiving

(*) procurement pending, waiting for release of funds. Delay might cause less storage