

Big Data Research at DKRZ

Michael Lautenschlager and
Colleagues from DKRZ and
Scientific Computing Research Group

Symposium Big Data in Science
Karlsruhe October 7th, 2014

Big Data in Climate Research

- **Big data** is an all-encompassing term for any collection of **data sets so large and complex that it becomes difficult to process** using traditional data processing applications. (Wikipedia)
- **Big data** usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and **process data within a tolerable elapsed time**. Big data "size" is a constantly moving target, as of 2012 ranging from a few dozen terabytes to many petabytes of data.
(Snijders, C., Matzat, U., & Reips, U.-D. (2012). 'Big Data': Big gaps of knowledge in the field of Internet. *International Journal of Internet Science*, 7, 1-5.)
- Gartner's **"3Vs" model** for describing big data: increasing **volume (amount of data), velocity (speed of data in and out), and variety (range of data types and sources)**. (Laney, Douglas. "3D Data Management: Controlling Data Volume, Velocity and Variety". Gartner. Retrieved 6 February 2001.)
- Big Data in climate research:
 - Satellite data from Earth observation (2 V)
 - Climate model data (1.5 V)
 - Data from observational networks (e.g. Tsunami and Earth quarks) (3V)

Outline

- The German Climate Computing Center (DKRZ)
- Infrastructure at DKRZ: present and future
- Climate model data characteristic
- DKRZ's HPC infrastructure developments
- Climate model data production workflow
- Future Developments

The German Climate Computing Center (DKRZ)

Founded in 1987 as a national institution

Operated as a non-profit limited company with four shareholders

- Max Planck Society for Research (55%)
- The City of Hamburg represented by the University of Hamburg (27%)
- Alfred Wegener Research Institute in Bremerhaven (9%)
- Helmholtz Center for Research in Geesthacht (9%)



Mission

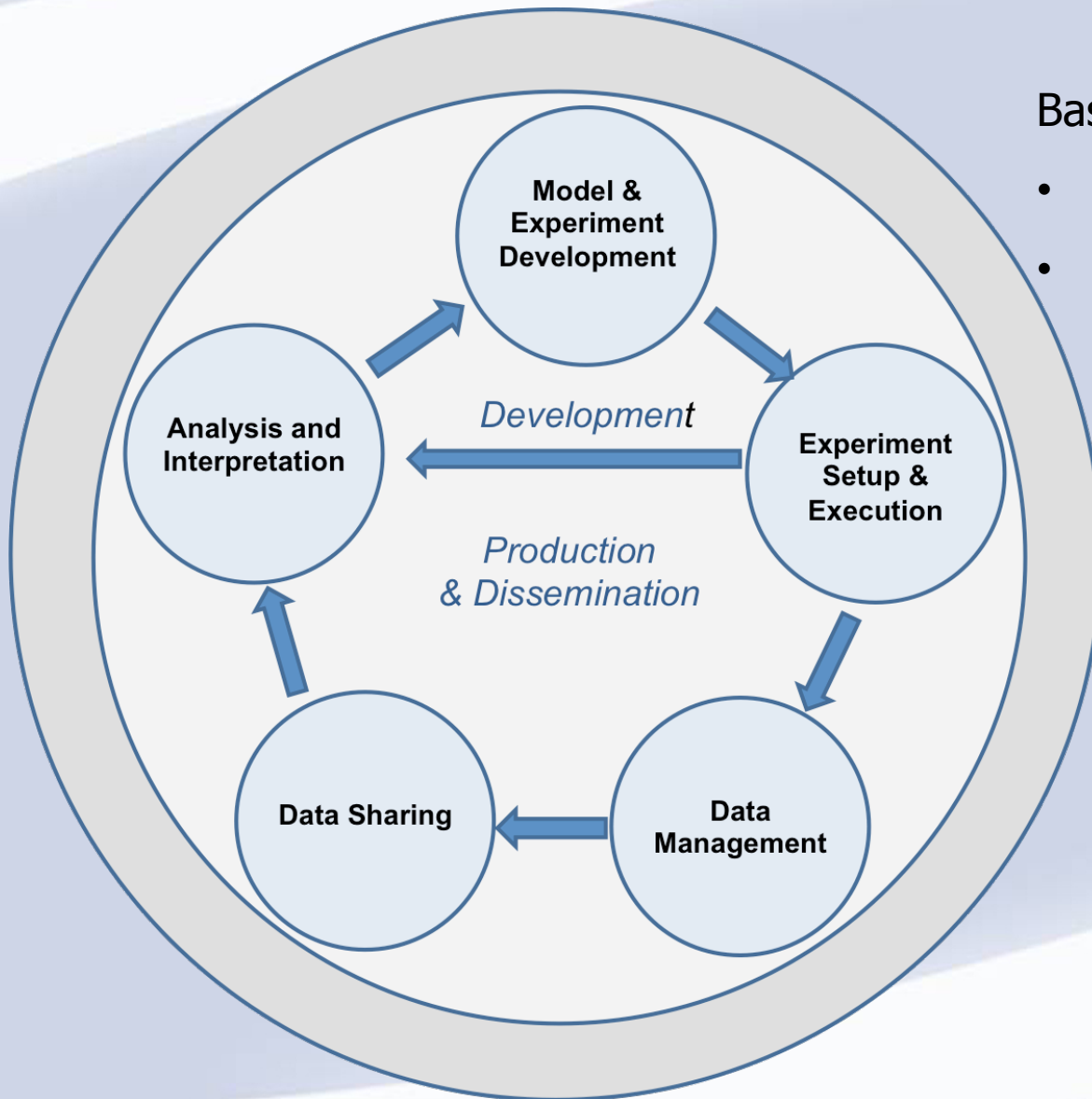
DKRZ – Partner for Climate Research

Maximum Compute Performance.

Sophisticated Data Management.

Competent Service.

Climate Modelling Support at DKRZ



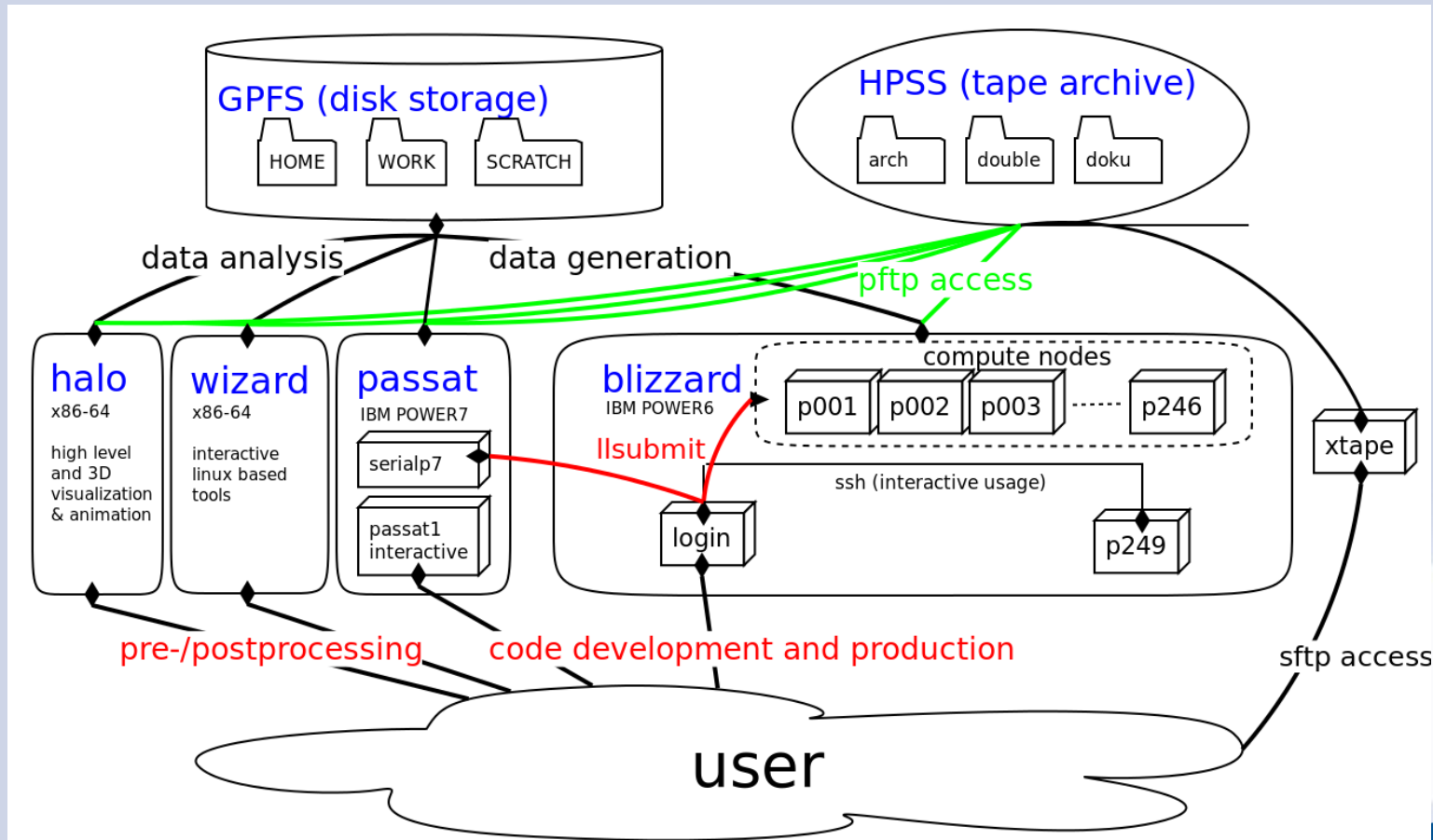
Basic Workflows:

- Climate Model Development
- Climate Model Data Production

HLRE-2: Architecture

In 2015
replacement
by HLRE-3

Current machine: IBM Power6
8.500 cores, 158 TFLOPS, 20 TB main memory
6 PB disk space, 100 PB tape archive space



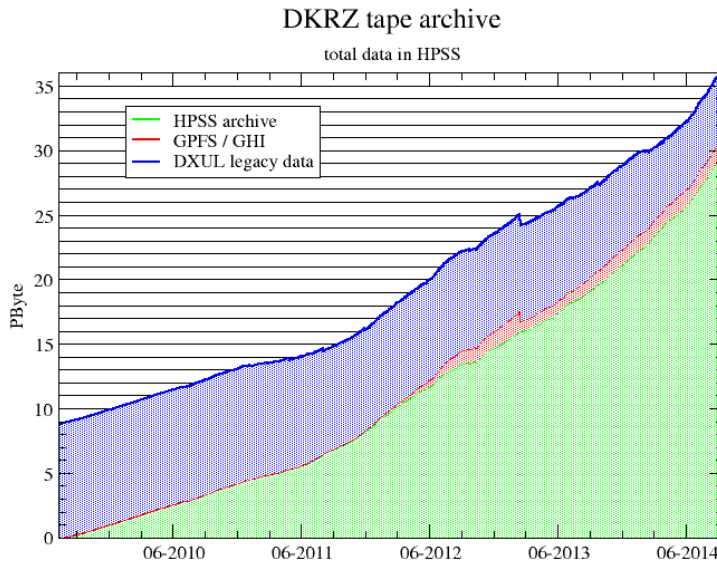
HLRE-2 to HLRE-3

Increases every ten years:
Processor speed: 500x
Disk capacity: 100x
Disk speed: 11x (17x with SSDs)

	2009	2015	Factor
Performance	150 TF/s	3 PF/s	20x
Nodes	264	2,500	9.5x
Node performance	0.6 TF/s	1.2 TF/s	2x
System memory	20 TB	170 TB	8.5x
Storage capacity	5.6 PB	45 PB	8x
Storage throughput	30 GB/s	400 GB/s	13.3x
Disk drives	7,200	8,500	1.2x
Archive capacity	53 PB	335 PB	6.3x
Archive throughput	9.6 GB/s	21 GB/s	2.2x
Power consumption	1.6 MW	1.4 MW	0.9x
Investment	30 M€	30 M€	1x

HLRE-2: Data Archive

Focus: climate model data
and related observations



Wed Sep 10 14:00:00 2014

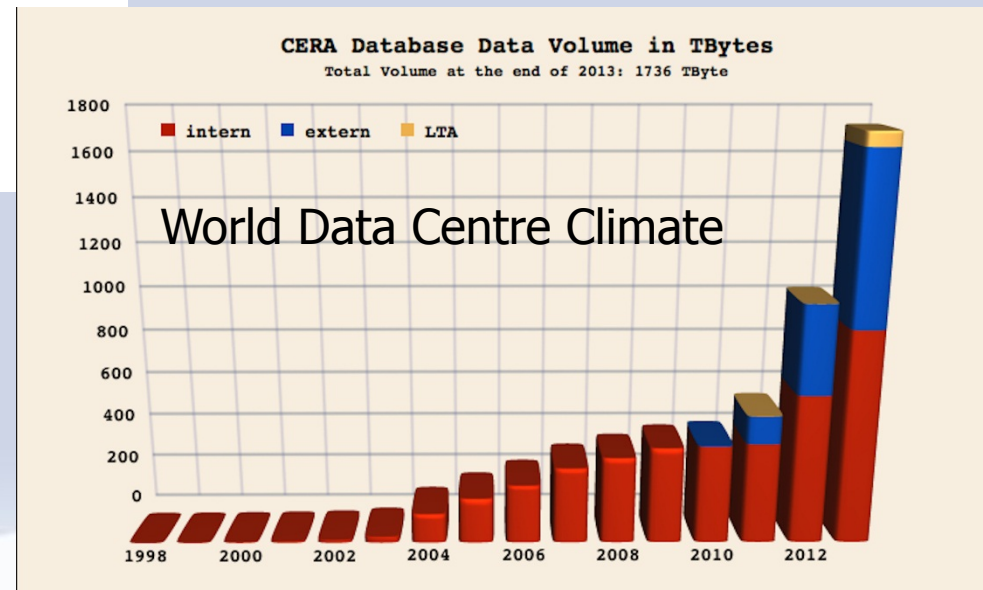
HPSS: 8 PB/year
(HLRE-3: 75 PB/y)

WDCC: 0.5 – 1.5 PB/year
(HLRE-3: 8 PB/y)

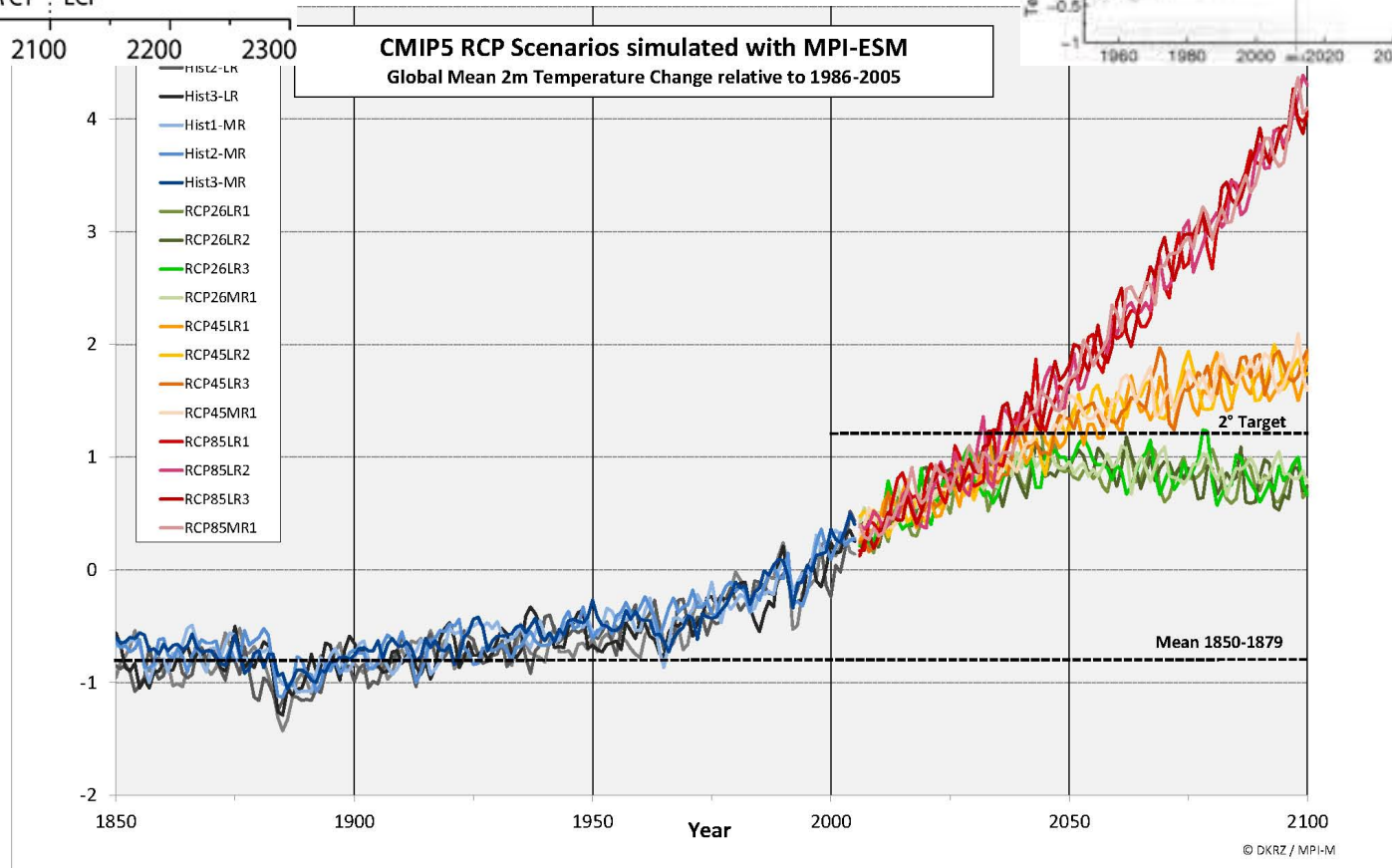
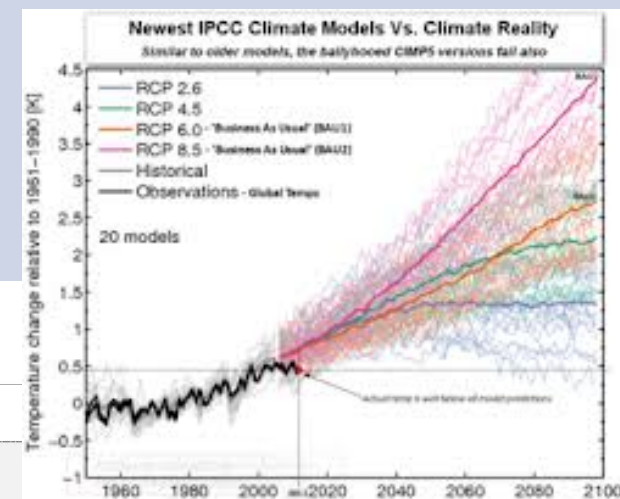
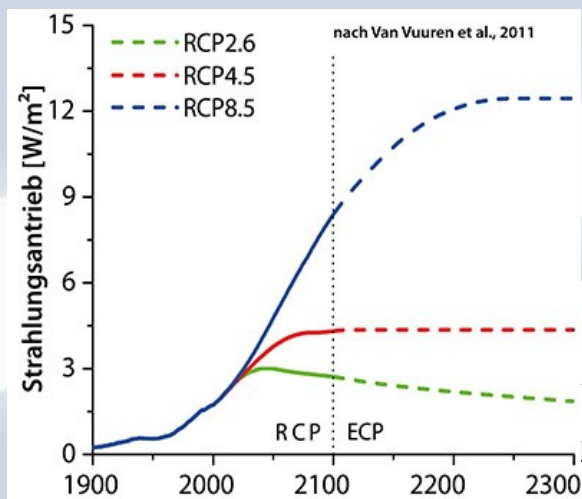
Content August 2014

HPSS: 36 PB

WDCC: 4 PB (fully documented)



Climate Modelling Example



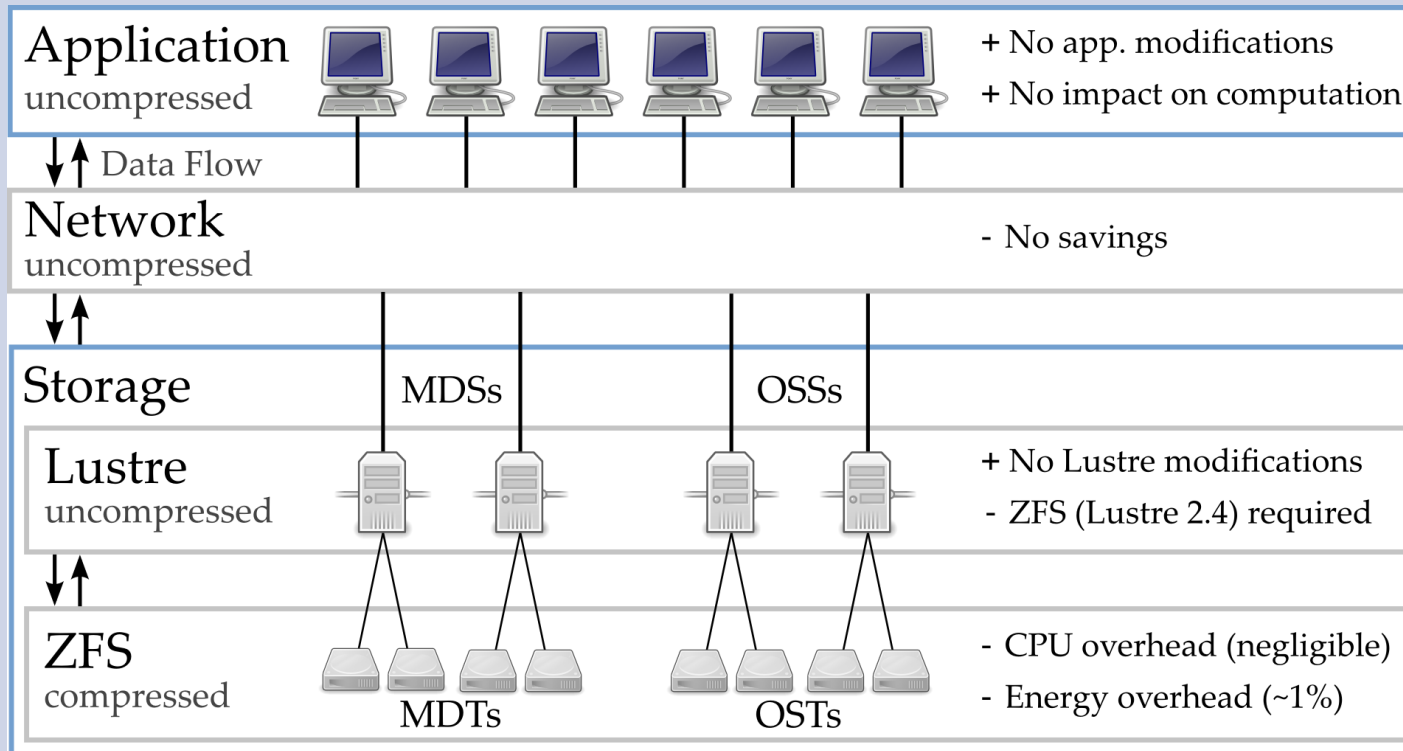
Data Amounts CMIP3/CMIP5

- CMIP3 / IPCC-AR4 (Report 2007)
 - Participation: 17 modelling centres with 25 models
 - In total 36 TB model data central at PCMDI and ca. ½ TB in IPCC DDC at WDCC/DKRZ as reference data
- CMIP5 / IPCC-AR5 (Report 2013/2014)
 - Participation: 29 modelling groups with 61 models
 - Produced data volume: ca. 10 PB with 640 TB from MPI-ESM
 - CMIP5 requested data volume: ca. 2 PB (in CMIP5 data federation)
 - Data volume for IPCC DDC: 1.6 PB (complete quality assurance process) with 60 TB from MPI-ESM
- Status CMIP5 data federated archive (August 2014):
 - 2.3 PB for 69000 data sets stored in 4.3 Mio Files in 23 data nodes
 - CMIP5 data is **more than 50 times** CMIP3
- Extrapolation for CMIP6 data federation:
 - **Volume: 150 PB**
 - **Number of files: 280 Mio Files**

Part of
WDCC/DKRZ

Research: More efficient disk usage

- Files system: Lustre-ZFS
 - lz4 compression increases energy consumption by max. 1%
 - But provides compression of 30% in average



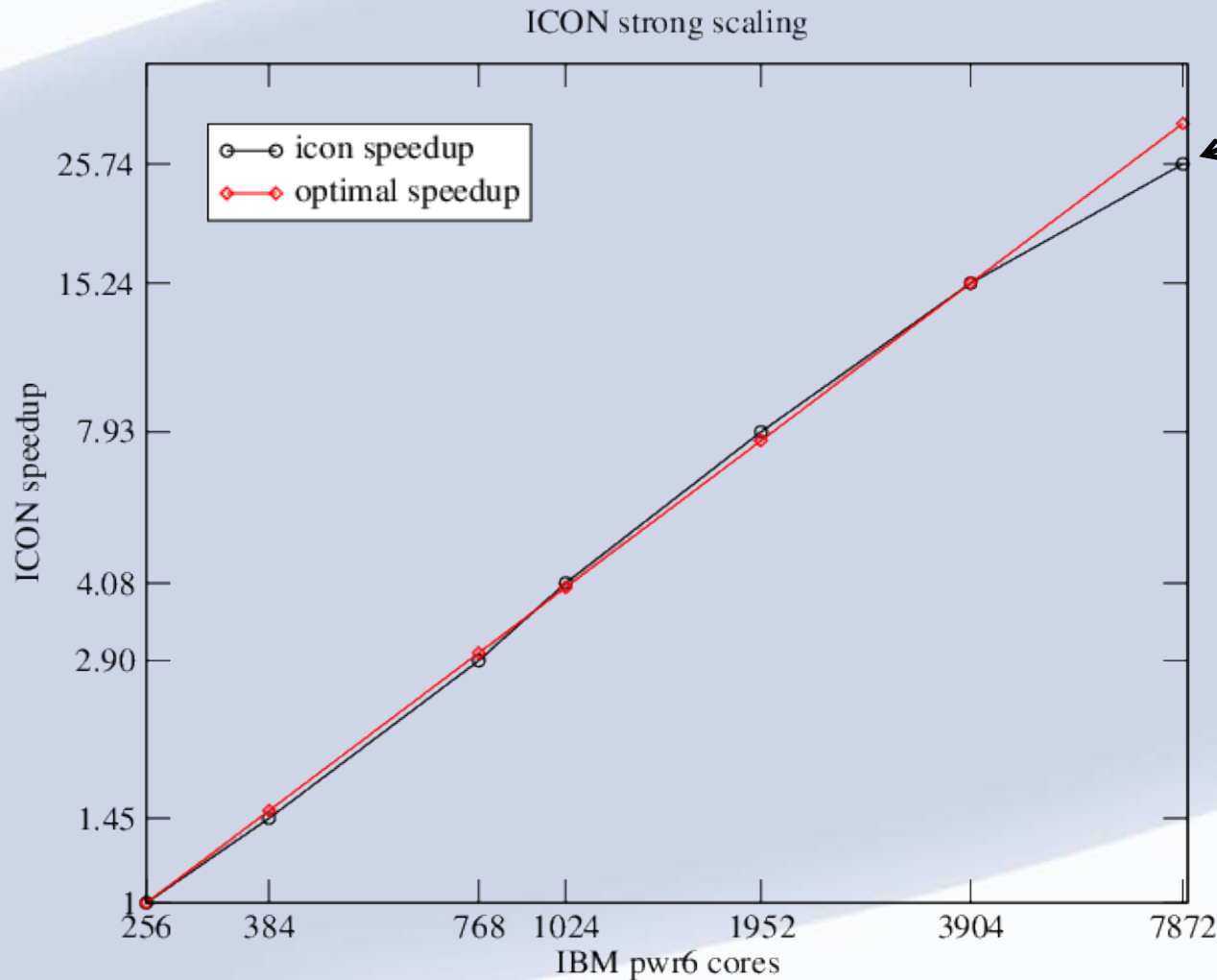
Potentially
applicable
to HLRE-3

MDS: meta
data server,
OSS: object
storage server,
MDT/OST:
meta data /
object storage
target

Parallelization on HPC: Scalability Objective of ICON

- able to conduct production simulations on target grids as large as 10^{10} grid elements ($10^4 \times 10^4 \times 400$)
- grid spacing finer than 400 m (100 m being the target)
- capable of efficiently using a diversity of advanced high performance computing resources
- scaling of the model to many tens, possibly hundreds, of thousands of cores

Parallelization on HPC: R2B07 GLOBAL 20km no I/O



Efficiency 83%

**Parallel I/O under
development**

CLIMATE MODELLING - A DATA INTENSIVE SCIENCE

Climate models produce extremely large amounts of data. In order to manage these data production rates, DKRZ operates data lifecycle services. With the ICSU World Data Center Climate (WDCC), DKRZ runs a fully documented long-term data archive with a size of currently 3 Petabytes. The entire tape archive is currently equipped to handle data production rates of up to 10 Petabytes per year. With the upcoming HLRE-3 system it will even manage annual rates up to 75 Petabytes. Beside data storage resources, a seamless end-to-end workflow from data production over data processing, data dissemination to data storage is applied to make optimal use of the huge but nevertheless limited HPC resources at DKRZ.



1. Data management plan

The data time line as well as volumes, structures, access patterns and storage locations have to be defined as accurate as possible for each DKRZ HPC project in order to realize a seamless workflow and efficient use of DKRZ resources.

2. DKRZ storage

Each DKRZ HPC project has to specify and to apply for compute and storage resources on an annual basis. Storage resources contain disc and tape storage. All resources are monitored on the basis of DKRZ HPC projects.

3. ESGF standardization



Climate data integration into ESGF (Earth System Grid Federation) requires standardization in order to make data interoperable within the federation. This data preparation process includes project specifications as well as adaptation of data and metadata to the ESGF data publication interface.

4. ESGF services

DKRZ offers a number of services to integrate („publish“), manage, discover and access climate data in the International ESGF. The data allocation includes definition of project specific publication and access policies, adaptation of data check routines and the data publication on the ESGF data node at DKRZ. User support for data publication and data access is included as well.

5. LTA DOKU



LTA DOKU stands for in-house longterm archiving in the DOKU(mentation) section of the tape archive at DKRZ. This service offers longterm archiving for data of DKRZ HPC projects as internal reference data only. A minimum set of metadata has to be supplied by data providers in order to characterize and identify them in the longterm archive of DKRZ. No additional information on data interpretation is provided. Focus here is set on internal data access from data providers.

6. LTA WDCC

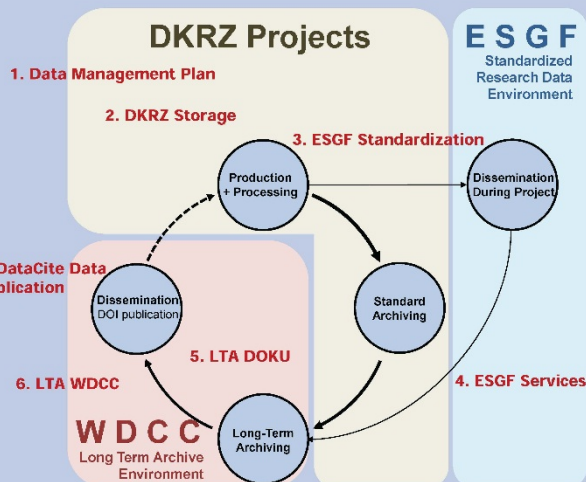
LTA WDCC stands for longterm archiving in the World Data Center for Climate (WDCC). This service is open for data from DKRZ HPC projects, data from ESGF but also for data from outside DKRZ. These data are fully integrated in the database system of the WDCC. The full set of metadata is provided in order to allow data interpretation even after ten years or more without contacting the data author. Connected to this archiving service is a fine granular data storage which allows field based data access (CERA container files) in distinction to the file based data access in the LTA DOKU service. Focus here is set on interdisciplinary data access.



7. DataCite data publication

DataCite is an international organization which aims to establish easier access to research data, increase the acceptance of research data as legitimate contribution in the scholarly record, and support data archiving to permit results to be re-used. Scientists are enabled to give and to get credit for the preparation of data products by formal data citations.

All data from the LTA WDCC service, that have passed a final quality assurance procedure are suitable for a DataCite data publication, i.e. citation metadata are published and a DOI (Digital Object Identifier) is minted. After receiving a DOI the data and key metadata remain unchanged, and the data is persistently accessible via its DOI.



DKRZ offers Services for End-to-End Workflows in Climate Modelling

Basic Workflows:

Climate Model Development

Climate Model Data Production

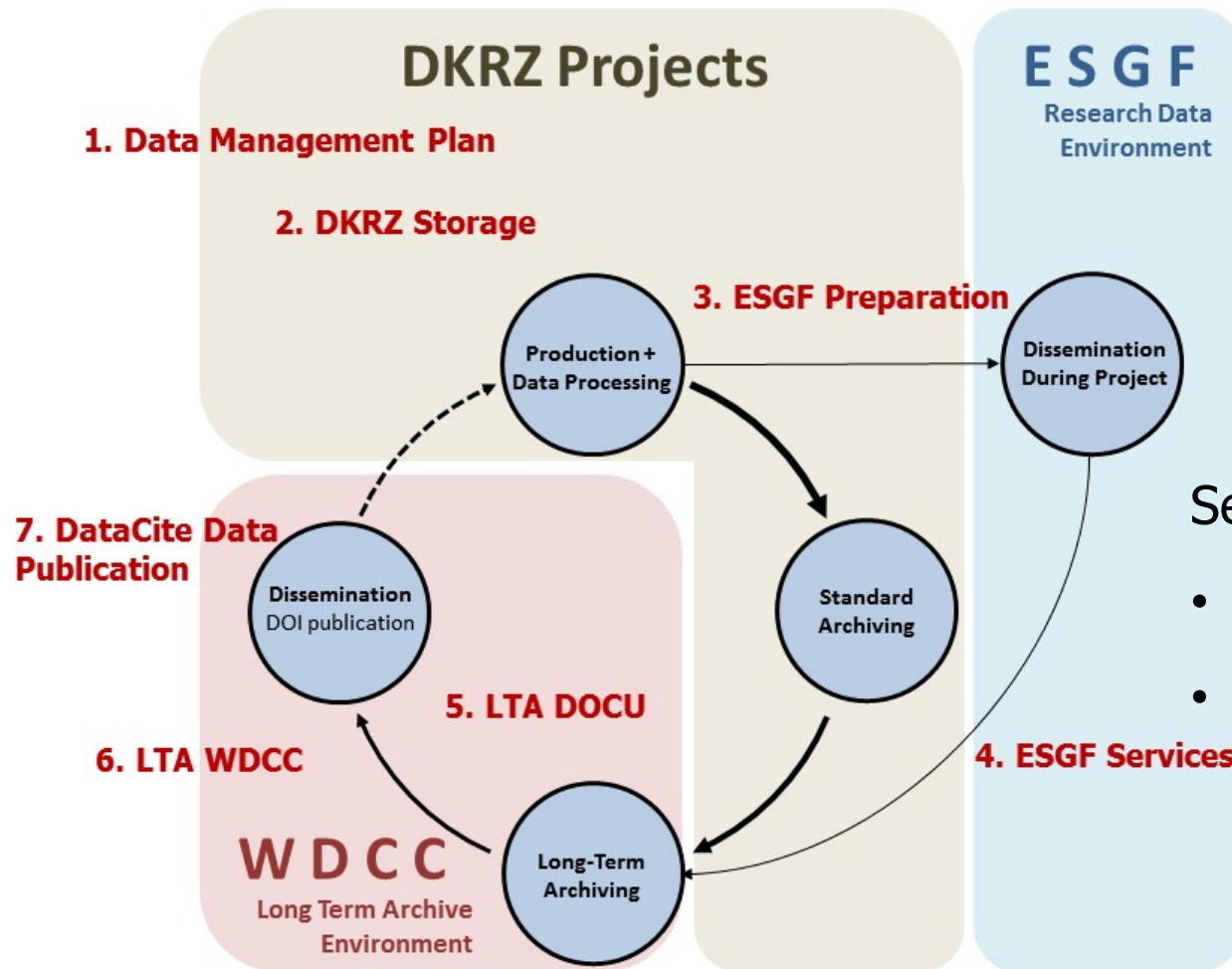
Poster:

ISC'14, Leipzig, Juni 2014

URL: https://www.dkrz.de/pdfs/poster/DataServices_ENG.pdf

www.dkrz.de

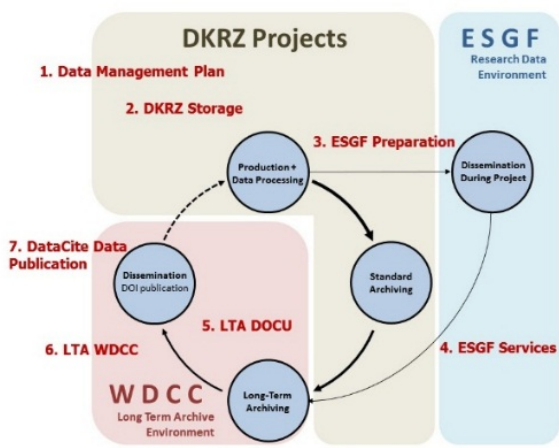
Climate Model Data Production Workflow



Separation:

- Scientific data analysis
- Long-term Archiving

Climate Model Data Production Workflow



1) Data Management Plan

The data **time line as well as volumes**, structures, access patterns and storage locations have to be defined as accurate as possible.

2) DKRZ Storage

Each DKRZ HPC project has to specify and to **apply for compute and storage** resources on an annual basis.

3) ESGF Standardization

Climate data integration into ESGF (Earth System Grid Federation) requires standardization in order **to make data intercomparable in the federation**.

4) ESGF Services

DKRZ offers a number of services to **manage, discover and access climate data** in the international Earth System Grid Federation (ESGF).

5) LTA DOKU

LTA DOKU stands for **in house long-term archiving** in the DOKU(mentation) section of DKRZ's tape archive. Focus here is on **internal data access from data providers**.

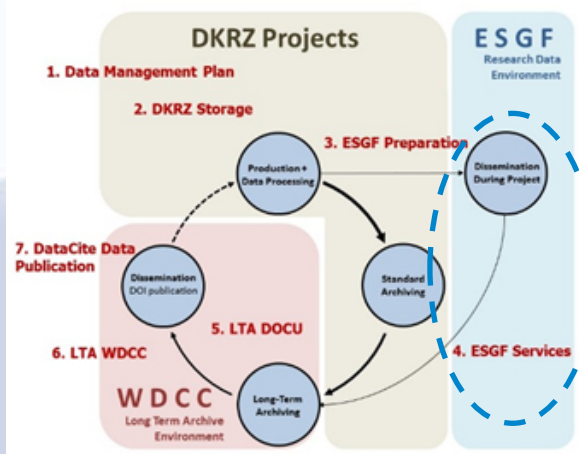
6) LTA WDCC

LTA WDCC stands for **long-term archiving in the World Data Center Climate (WDCC)**. This service is open for data from DKRZ HPC projects, data from ESGF but also for data from outside DKRZ. These data are fully integrated in the database system of the WDCC. The full set of **metadata** is provided in order to **allow for data interpretation even after ten years** or more without contacting the data author. Focus here is on **inter-disciplinary data access**.

7) DataCite Data Publication

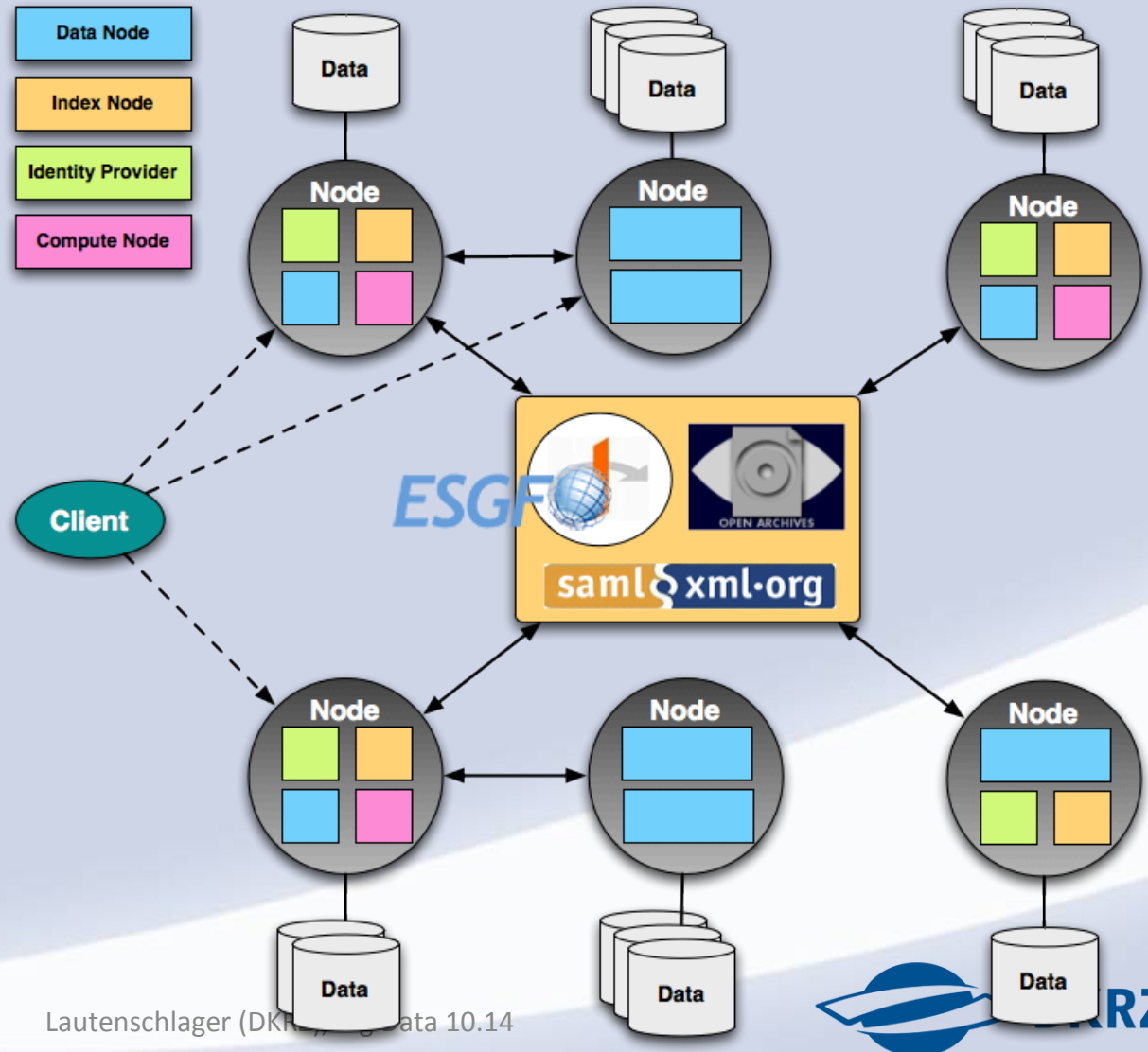
DataCite is an international organization which focuses on **publication of scientific data** in context with and to be used in the scientific literature. After passing the DataCite Data Publication services these data entities are irrevocable, has been assigned a DOI (Digital Object Identifier) and key metadata including the citation reference are registered in the DataCite repository. **The DOI allows for transparent data access and the citation reference allows for integration into the reference list of a scientific publication.**

Climate Model Data Federation (ESGF)



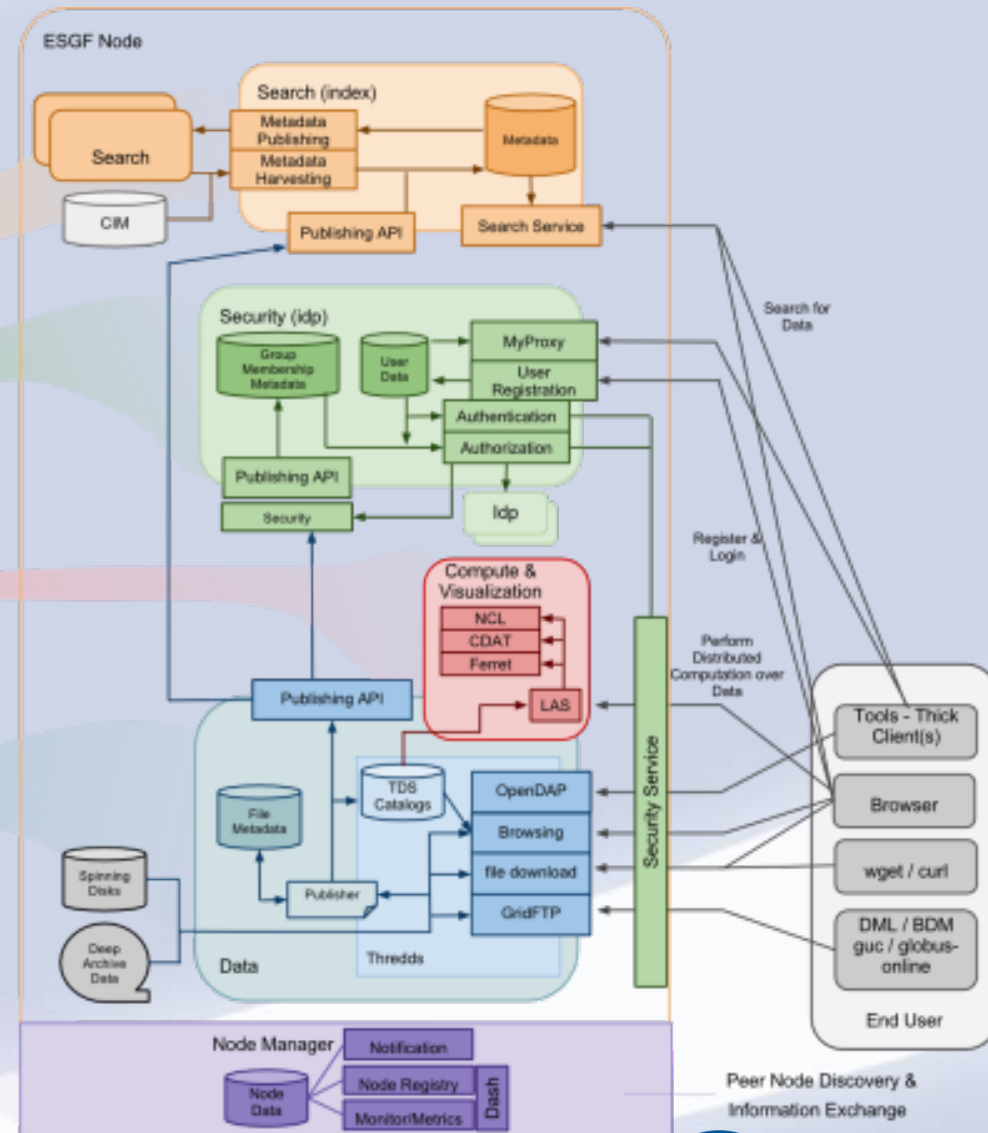
Service 4

ESGF P2P Architecture:
Data Node, Index Node,
Compute Node, IdP



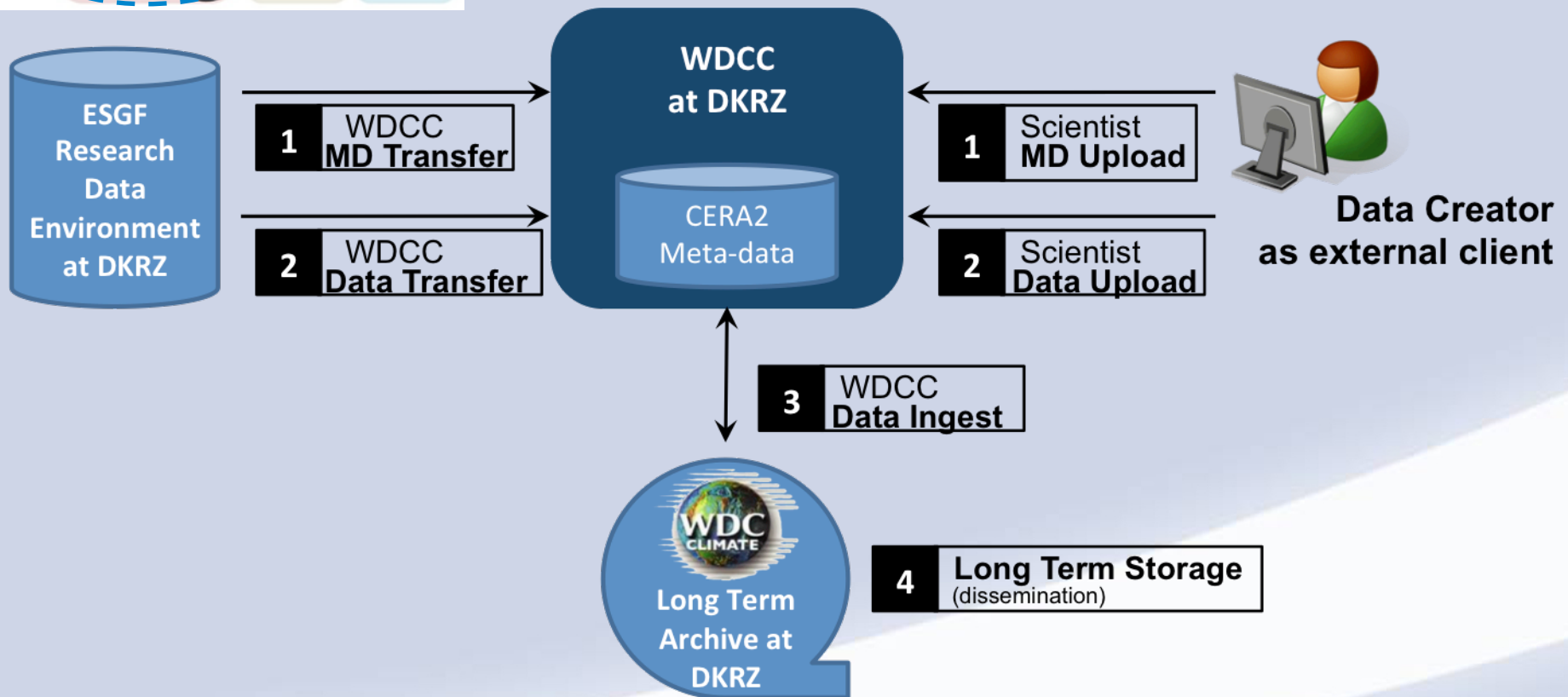
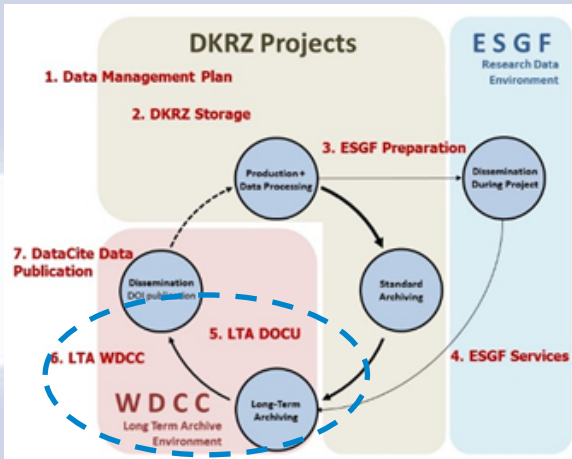
Service 4

Search
Security
Compute
Data
Manager



LTA: Data ingestion

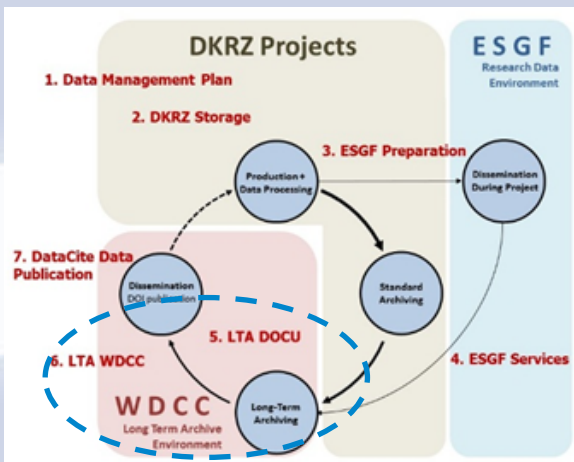
Services 5 + 6



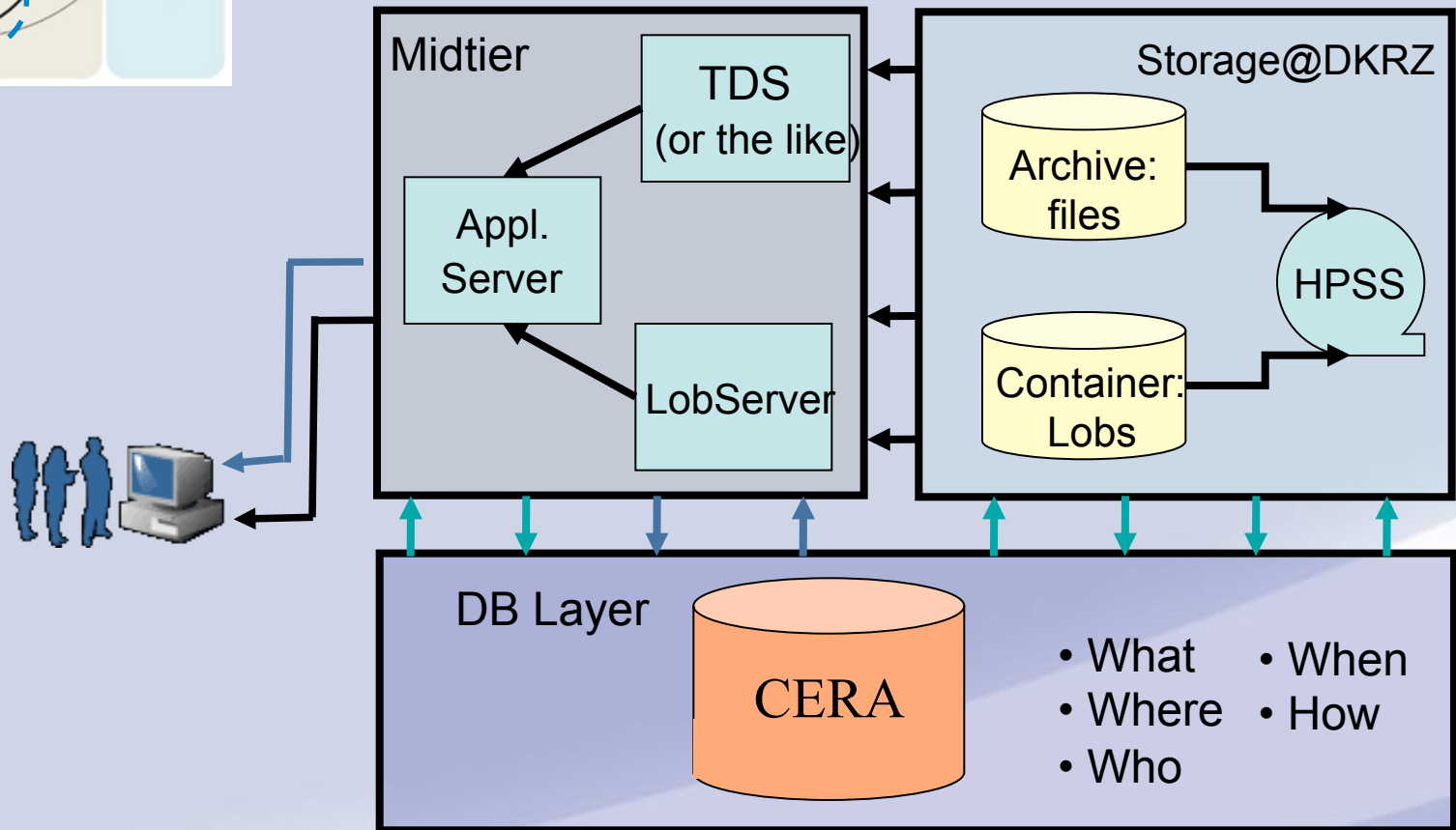
Long-Term Archive

Lautenschlager (DKRZ), Big Data 10.14

LTA: Data access



Services 5 + 6



Future Big Data Developments

- Seamless end-to-end workflow
 - Model data production – analysis – long-term archiving
- Model Code optimization
 - Many cores systems, GPUs
- File systems
 - Compression, parallel I/O, co-design of storage I/O APIs, object store
- Object Storage
 - OpenStack Swift
- Data access
 - Unique Identifier (DOI, Handle System) may replace directory structure
- Network
 - Bandwidth, SLAs, managed networks