

DKRZ Datenhierarchie

Von der Antragstellung bis zur Archivierung

07-May-2014

Hannes Thiemann
Deutsches Klimarechenzentrum (DKRZ)
Datenmanagement

- Antrag
- Ressourcen
- Archivierung + Zugriff
- Support
- LiveDemo

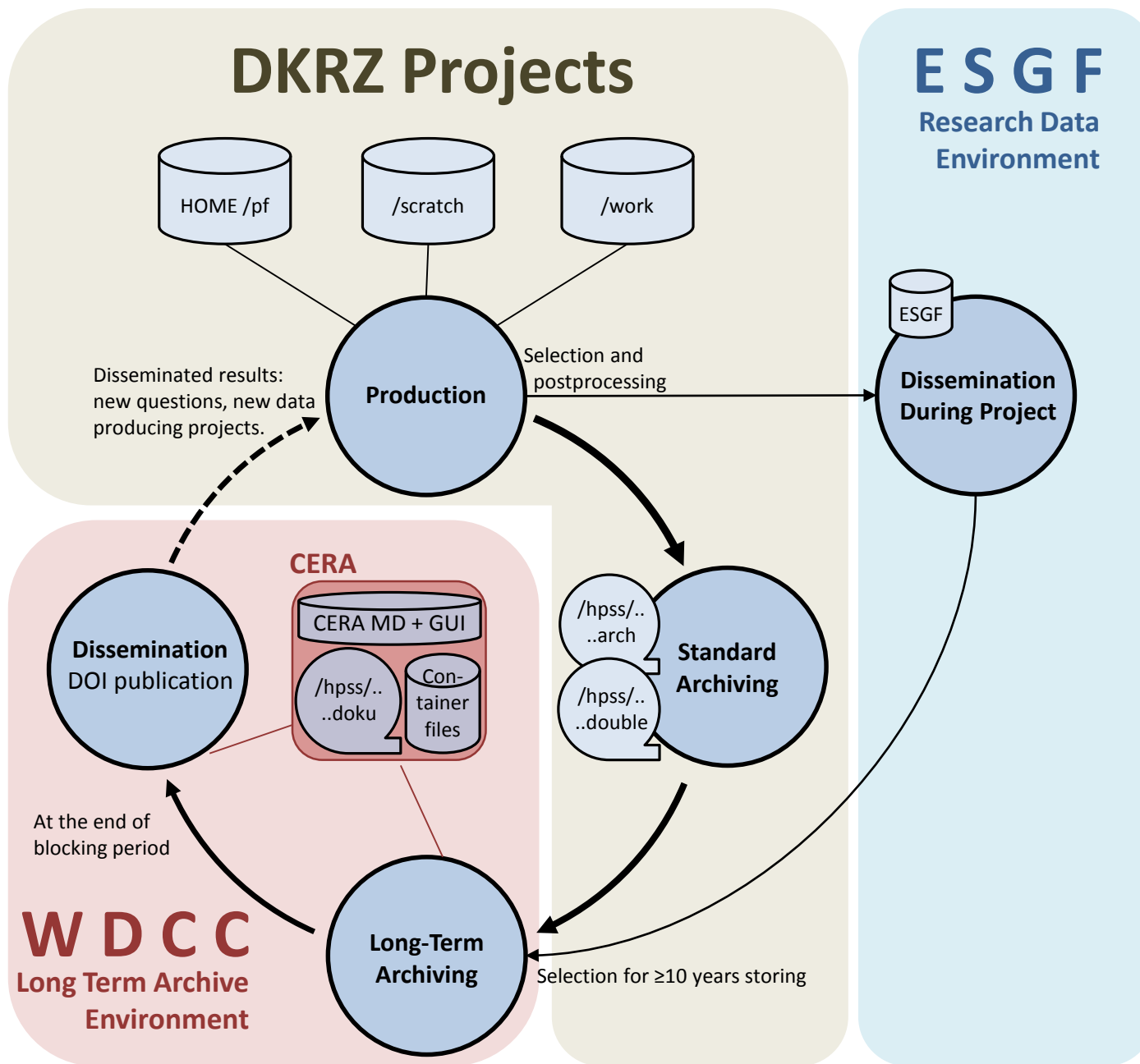
Antragsphase (I)

- Das DKRZ stellt unentgeltlich Ressourcen (Rechenzeit, Archivkapazität, Fachberatung) für Wissenschaftler/innen zur Verfügung, die
 - auf dem Gebiet der Erdsystemforschung tätig sind, und
 - für ihre Arbeiten auf Höchstleistungs-DV-Ressourcen angewiesen sind.
- Die Ressourcen werden zu gleichen Teilen auf zwei Kontingente verteilt
 - Freies Kontingent, verwaltet vom WLA
 - Gesellschafter Kontingent

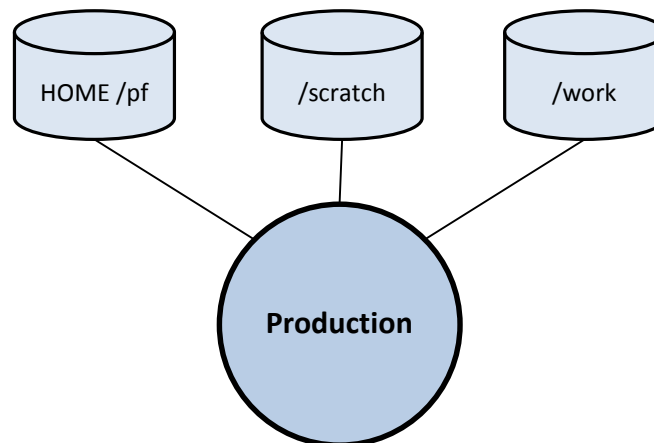
Antragsphase (II)

- Anträge auf Rechenzeit müssen bis zum 31. Oktober des vorherigen Jahres eingehen.
- Wenn Sie Ressourcen aus dem Gesellschafter-Kontingent nutzen wollen, melden Sie in Absprache mit Ihrer Institutsleitung bzw. dem/der EDV-Verantwortlichen Ihres Institutes ein HLRE-Projekt an.

DKRZ Projects



DKRZ Projects

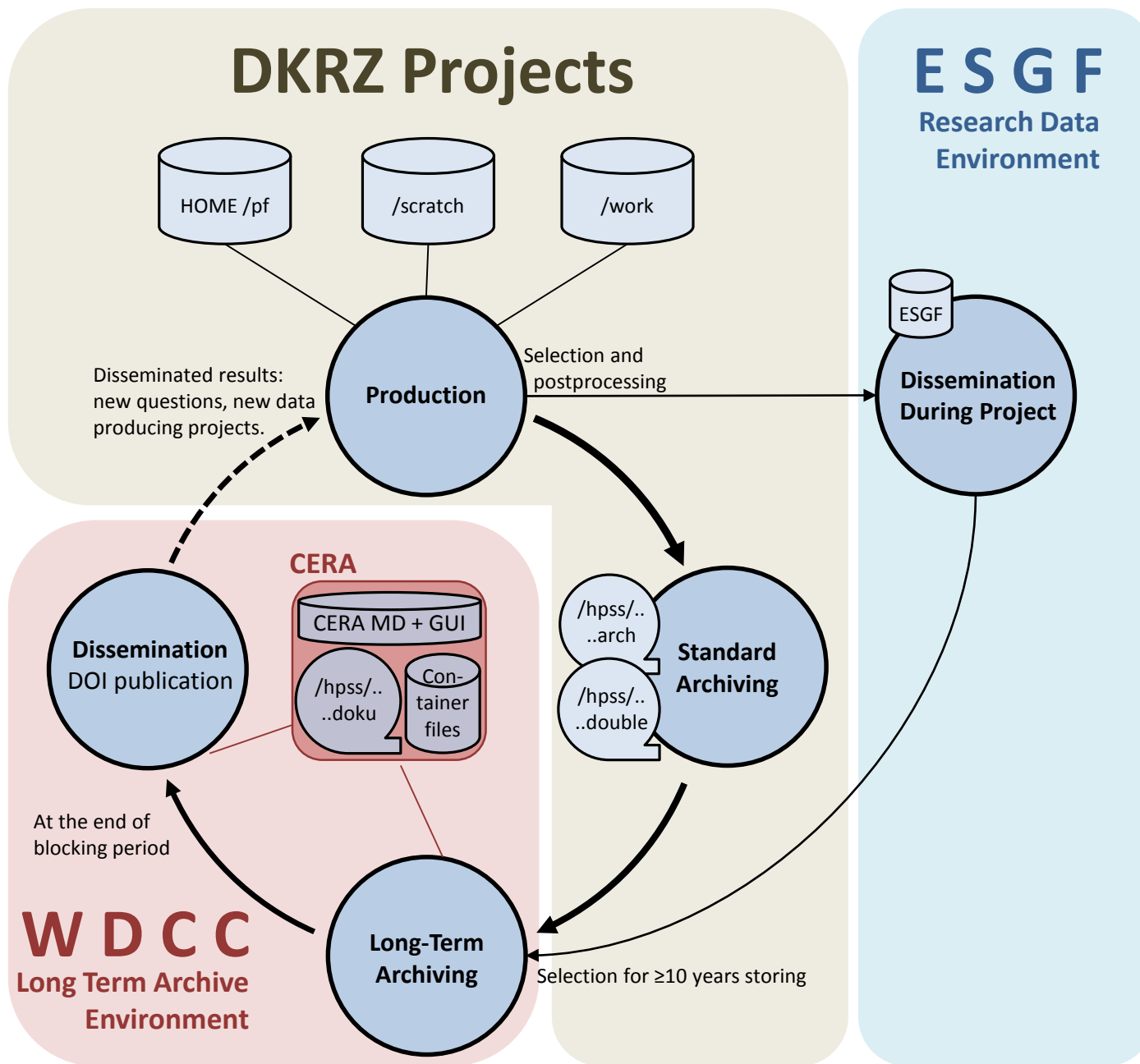


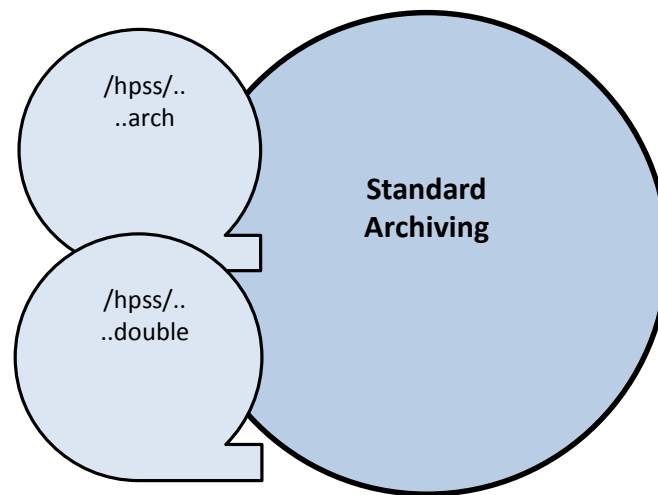
Data resources (Disk)

Diskspace (GPFS)

- HOME
 - /pf
 - < 8 GB (per User)
 - for source codes, ...
 - Backed up regularly
- SCRATCH
 - /scratch
 - <15 TB (per User)
 - no backup
 - old data automatically removed, granted period is 14 days
- WORK
 - /work
 - project based, quota applies
 - no backup
 - no data removed by DKRZ

DKRZ Projects



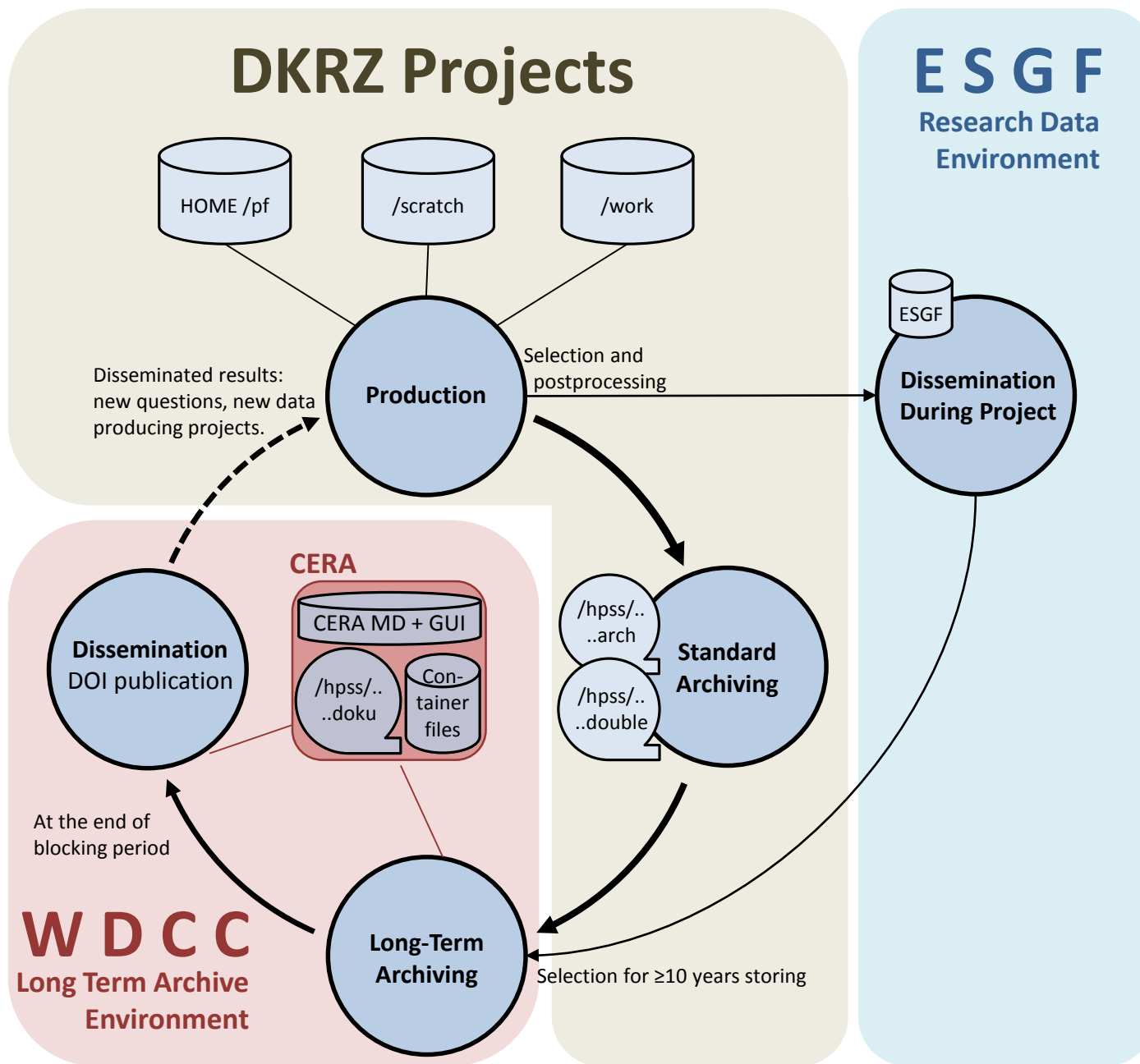


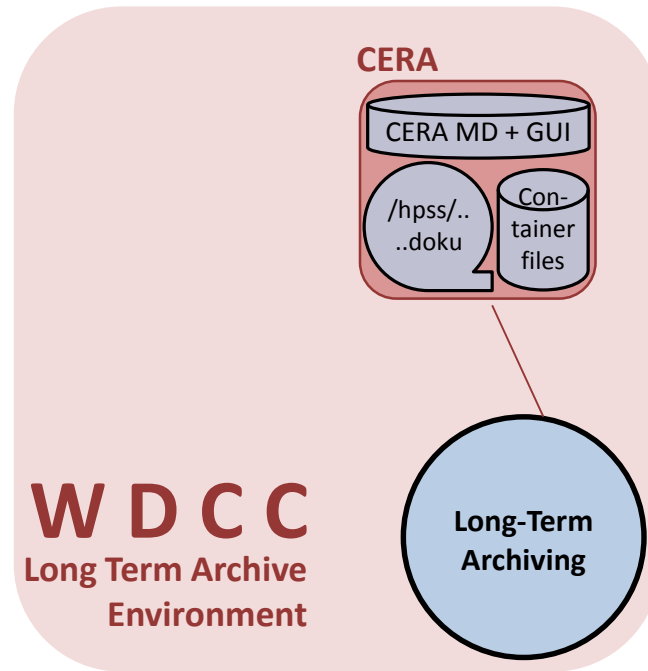
Data resources (Tape / HSM system)

ARCH	/hpss/arch	Single copy on tape	Size 20-50 GB recommended	Data removed one year after project has expired
DOUBLE	/hpss/double	2 copies on tapes in different fire compartments		

One quota for ARCH and DOUBLE

DKRZ Projects





Long-Term Archiving / Doku

- Service for DKRZ Projects
- Data moved from /arch to /doku
- By default to be removed 10 years after project expiration
- Two copies in different fire compartments
- Must not be changed
- Accessible by all registered users of DKRZ and WDCC
- Persistent CERA-Link
- Metadata needed (basic set)
 - Project description
 - Contact
 - Filelist

Long-Term Archiving / Doku / Filelist

- Filelist available at
- `/hpss/arch/<prj>/_PROJECT.<prj>.filelist.GIGA`

- Example

```
-rw-r----- 426:1076 Mar 11 08:32 188011141120 189 giga  
/hpss/arch/bm0146/k204026/CEOP2/CEOP2_NCEP_GRID_2013.tar  
-rw-r----- 426:1076 Jan 24 08:26 16954480640 17 giga  
/hpss/arch/bm0146/k204026/CEOP2/CEOP2_NCEP_MOLTS_RAW_2008.tar
```

DOKU: addressing data

Aus
 /hpss/arch/bm0146/k204026/CEOP2/CEOP2_NCEP_GRID_2013.tar
 wird
 /hpss/doku/bm0146/DKRZ_LTA_146_ds00001
 /k204026/CEOP2/CEOP2_NCEP_GRID_2013.tar

User				
<tape>	<arch>	<dir>	<file>	
Tape	/hpss/arch/bm0146	k204026/CEOP2	CEOP2_NCEP_GRID_2013.tar	
DOKU				
<tape>	<doku>	<ds>	<dir>	<file>
Tape	/hpss/doku/bm0146	DKRZ_LTA_146_ds00001	k204026/CEOP2	CEOP2_NCEP_GRID_2013.tar

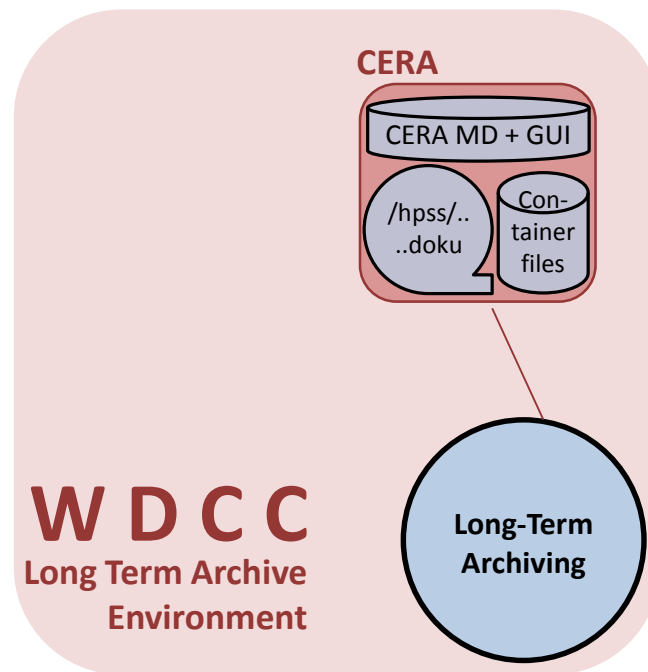
DOKU: addressing data

Metadata

- http://cera-www.dkrz.de/WDCC/ui/Entry.jsp?acronym=DKRZ_LTA_800_ds00001

CERA download

- http://cera-www.dkrz.de/WDCC/ui/externalFilesLTA?acronym=DKRZ_LTA_800_ds00001&pointerid=2462701



Long-Term Archiving / CERA

- Two tape copies in different fire compartments
- Stored within “container files”
 - Greater safety with regards to tape damage
 - Better utilization of HSM resources
- Must not be changed, new versions possible
- Metadata needed – complete set
- Access
 - Embargo possible
 - Access using CERA account (GUI, jblob*, wget)
 - ESGF data available through ESGF + CERA portal (if already archived)
- Prerequisite for data DataCite publication

Support

- Main entry for all support enquiries:

data@dkrz.de

Important Links

- Main Entry to CERA
 - <http://cera-www.dkrz.de/>
- List of projects in DOKU
 - <http://cera-www.dkrz.de/WDCC/DOKU>
- Interface to generate metadata
 - http://cera-www.dkrz.de/LTA_metadata
- Status of CERA services
 - <http://wdcc-status.dkrz.de>

Poster

- Lobster
- CERA Download
- B2FIND / EUDAT

Live Demo - > Hermann

- Eintragen von Metadaten:
- http://cera-www.dkrz.de/LTA_metadata
- Archivierung der Daten unter
- 1.) /hpss/doku (zur fachinternen Nutzung)
i.d.R. Speicherung von .tar Files
- 2.) CERA Container (für interdisziplinäre
Nachnutzung) i.d.R. Zeitreihen
- Zus. Absprache/Abstimmungen für 2.)