

# 41. Sitzung des Wissenschaftlichen Lenkungsausschusses der Deutsche Klimarechenzentrum GmbH

---

*Beginn der Sitzung: 30. November 2020 um 9:34*

## **Teilnehmer**

Dr. Andreas Baumgärtner, DLR PT (BMBF)  
Dr. Hendryk Bockelmann, DKRZ  
Prof. Dr. Claus Böning, GEOMAR Kiel  
Dr. Frauke Feser, Helmholtz-Zentrum Geesthacht  
Dr. Bernadette Fritzscht, AWI Bremerhaven (Vorsitzende des DKRZ-Usergroup-Komitees)  
Prof. Dr. Andreas Hense, Institut für Geowissenschaften, Abt. Meteorologie Universität Bonn  
Prof. Dr. Thomas Ludwig, DKRZ  
Dr. Mathis Rosenhauer, DKRZ (Protokoll)  
Prof. Dr. Robert Sausen, DLR Oberpfaffenhofen (Vorsitz)  
Hannes Thiemann, DKRZ  
Dr. Martin Werner, AWI Bremerhaven  
Dr. Sönke Zaehle, Max-Planck-Institut für Biogeochemie

## **1. Annahme der Tagesordnung**

Die Tagesordnung wird um die Punkte 5a und 5b erweitert. Die erweiterte Tagesordnung wird angenommen.

## **2. Organisatorisches**

### **a) Annahme des Protokolls der 40. Sitzung**

Das Protokoll wird angenommen.

### **b) Ort und Termin der nächsten Sitzung**

Die nächste Sitzung wird am 31.05.2021 um 10:Uhr beginnen.

## **3. Bericht DKRZ**

### **a) Nutzung HLRE-3 (H. Bockelmann)**

Die auf Mistral abgerufene Rechenzeit lag in den vergangenen Monaten bei ca. 90%. Der Community-Anteil ist in letzter Zeit noch deutlich angestiegen. Trotz der hohen Auslastung sind die Wartezeiten recht kurz, was insbesondere an Wochenenden zu Leerphasen führen kann.

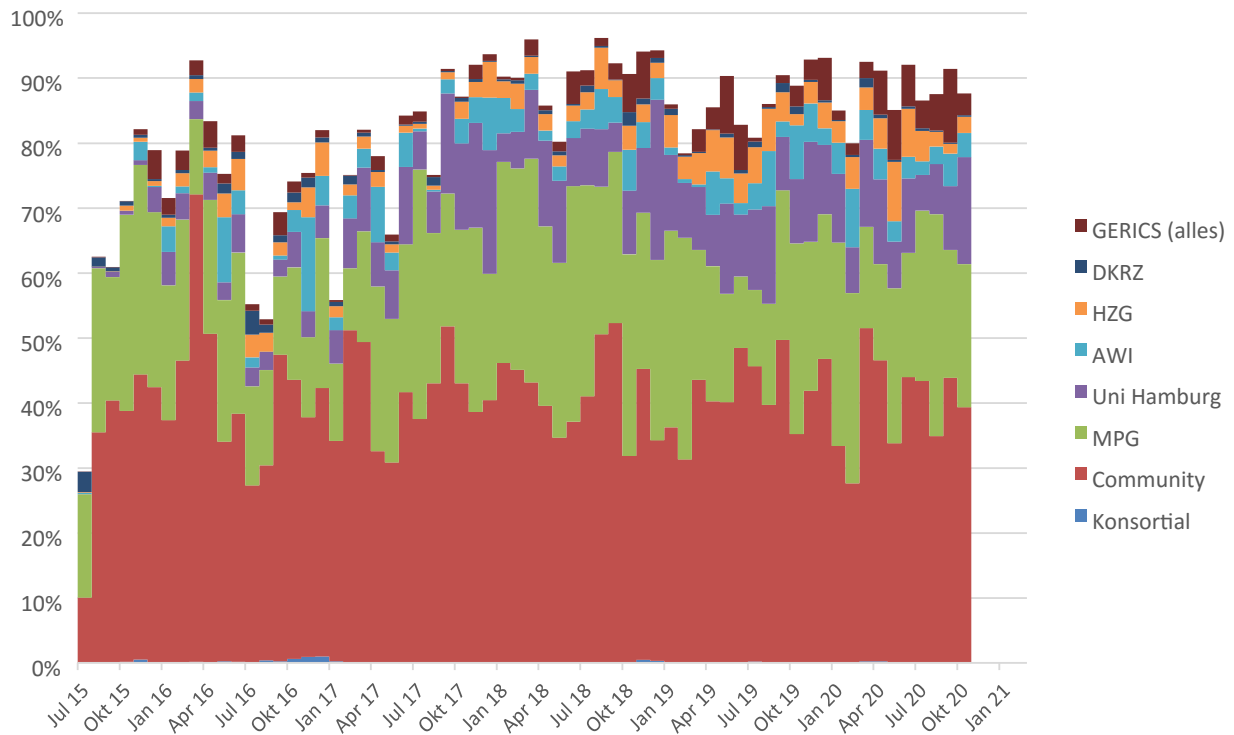


Abbildung 1: Durchschnittlicher Tagesverbrauch auf Mistral

Community-Projekte und solche der Gesellschafter haben ihren jeweiligen Anteil an Rechenzeit mit einer Abweichung von wenigen Prozent abgerufen.

Nach einer Standzeit von mehr als fünf Jahren ist die Beschaffung von Ersatzteilen für Mistral schwierig bis unmöglich. Daher wird die verfügbare Rechenzeit im Verlauf von 2021 weiter absinken. Zur nächsten Sitzung wird der dauerhafte Ausfall von Rechenknoten genauer quantifiziert werden können. Im Verlauf von 2020 waren durchschnittlich 5% der Knoten aufgrund defekter Hardware nicht verfügbar.

Auf den Plattenspeichersystemen war die Nutzung des Scratch-Bereichs stark angestiegen, sodass es hier notwendig war gegenzusteuern. Inzwischen konnte dieser Bereich reduziert werden und der Betrieb des Gesamtsystems ist bei einem Füllstand von 80% sehr stabil.

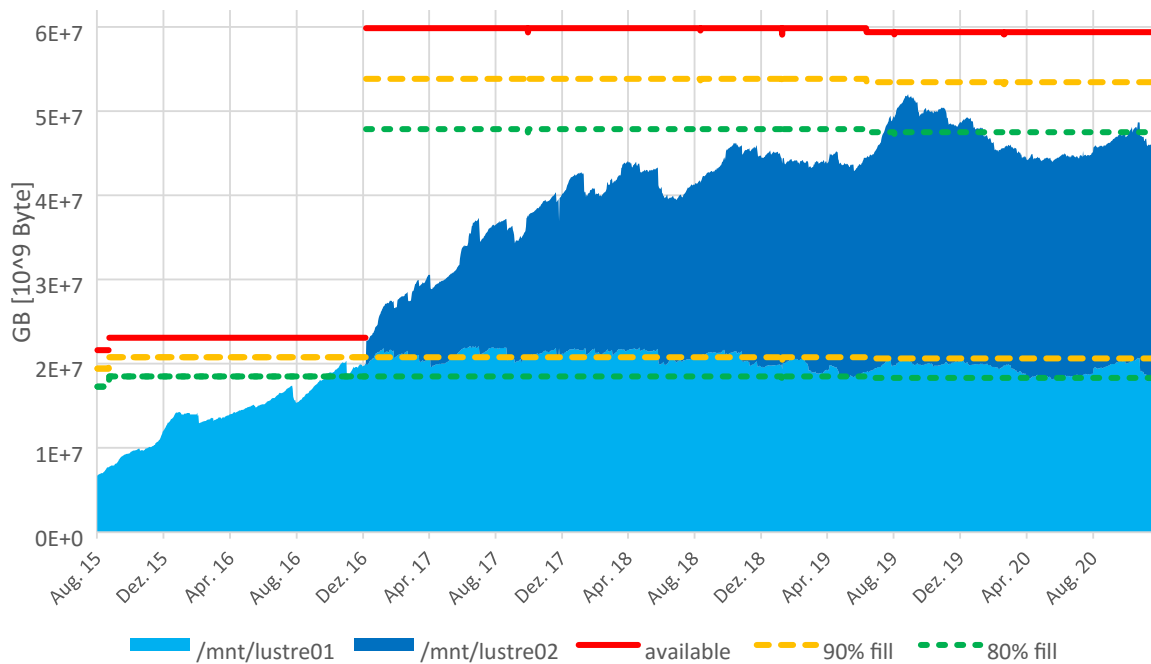


Abbildung 2: Belegung des Lustre Dateisystems [GB]

In 2020 sind bis zum jetzigen Zeitpunkt nur etwa 16 PB neuer Daten archiviert worden. Zum Übergang auf HLRE-4 wird das Bandarchiv voraussichtlich stärker genutzt werden, um einen unnötigen Transfer der Daten auf das Festplattensystem von HLRE-4 zu vermeiden. Insgesamt befinden sich nun etwa 130 PB im Archiv.

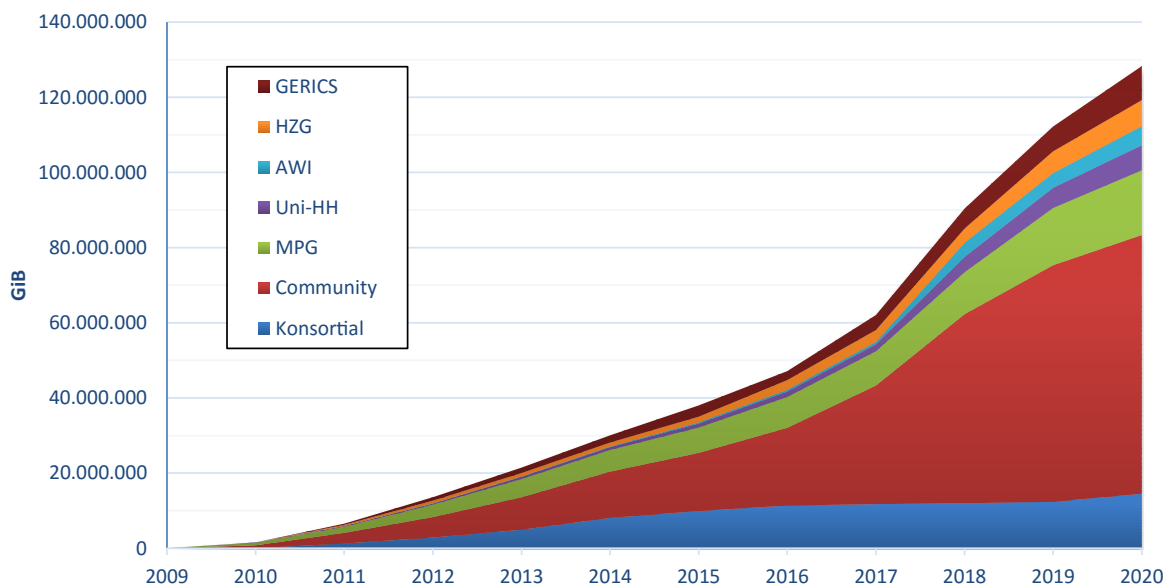


Abbildung 3: Belegung des Bandarchivs [GB]

### b) Perspektiven, u.a. Planung HLRE-4 (T. Ludwig)

Die Vertragsunterschrift mit Bull über HLRE-4 wurde vor einigen Tagen vollzogen. Hierzu gab es auch eine Pressemitteilung.

Der Zuschlag zum neuen HSM-System wurde erteilt. Der Aufbau beginnt in Kürze. Mit der Inbetriebnahme ist Anfang des kommenden Jahres zu rechnen.

Mit dem Umbau des Rechnerraumes für HLRE-4 wird Q1/2021 begonnen. Nach Abschluss der Bauarbeiten ist der Aufbau von HLRE-4 für Q2/2021 geplant. Im August 2021 könnte dann die Inbetriebnahme des neuen Rechners erfolgen. Eine Einweihungsfeier wird voraussichtlich erst Anfang 2022 stattfinden können.

Ursprünglich war die Inbetriebnahme der zweiten Phase zwölf Monate nach der ersten geplant. Aufgrund der Verzögerung von Phase eins wird bereits Q1/2022 mit dem Aufbau der Phase zwei begonnen. Die Entscheidung über die Ausstattung der zweiten Phase mit GPUs oder reinen CPU-Knoten ist noch nicht gefallen. Im März 2021 wird zu diesem Thema ein Bericht erstellt werden. Auf dieser Grundlage wird dann im Rahmen eines Treffens mit Gesellschaftern, dem WLA und der Benutzervertretung im April 2021 eine Entscheidung fallen.

#### **4. Bericht aus der DKRZ-User-Group (B. Fritsch)**

Für die zukünftige Nutzung von /pool/data wurde ein neues Konzept erarbeitet, welches im Folgenden vorgestellt wurde. Die Vortragsfolien hierzu befinden sich im Anhang.

In der anschließenden Diskussion kam die Frage auf, inwiefern eine adäquate Beschreibung der Daten sichergestellt werden kann. Die Abteilung Datenmanagement (DM) des DKRZ würde zu diesem Zweck die Datenverantwortlichen bei der guten Dokumentation der Daten beraten und unterstützen. Zur nächsten Sitzung wird ein Bericht von DM zum Stand der Dokumentation der Daten gegeben werden. Eine flexible Laufzeit der Projekte sollte ermöglicht werden. So hätten die Nutzenden eine gewisse Sicherheit bezüglich der dauerhaften Verfügbarkeit der Daten. Trotzdem sollten alle zwei Jahre Fortsetzungsanträge gestellt werden müssen, um eine fortwährende Betreuung und Nachfrage der Daten zu dokumentieren.

Für die Genehmigung eines neuen /pool/data-Projekts muss projektübergreifendes Interesse an den Daten nachgewiesen werden. Ein zitierfähiges Paper zu den Daten sollte bei den Datenverantwortlichen angeregt werden.

Der WLA unterstützt den Antrag. Das DKRZ soll die Verwaltung der Projekte mit den vorgeschlagenen Eckdaten übernehmen. Nach einer einjährigen Probephase wird der WLA erneut über das Verfahren beraten und ggf. Anpassungen vornehmen.

Auf der letzten Sitzung der Usergroup am 11.11.20 wurde die neu etablierte Vortragsreihe „Tech Talks“ des DKRZ lobend erwähnt. Auch die längerfristige Verfügbarkeit der Aufzeichnungen auf YouTube wurde als nützlich erachtet. Die spezifische Aufbereitung der Themen für die Gegebenheiten am DKRZ wurde von vielen Teilnehmenden begrüßt.

Der nächste Nutzerworkshop, welcher im Zusammenhang mit der Einweihungsfeier für HLRE-4 stattfinden soll, ist weiter in Planung.

## 5. Sonstiges

### a) Zitierbarkeit von Rechenzeit bzw. Ressourcen am DKRZ

Das DKRZ schlägt vor über HLRE-4 eine Veröffentlichung zu erstellen, die von Nutzern des Systems dann in wissenschaftlichen Veröffentlichungen, welche am DKRZ durchgeführte Rechnungen beinhalten, zitiert werden soll. Zusätzlich soll die Nutzung der DKRZ-Ressourcen mit einem Satz in den Acknowledgements von Artikeln gewürdigt werden. Community-Projekte sollten bei der Vergabe der Rechenzeit auf die neuen Anforderungen hingewiesen werden. Das RZ Jülich setzt ein vergleichbares Verfahren bereits ein.

<https://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/ComputingTimeAcknowledgements.html>

Die Erwähnung des DKRZs in den Acknowledgements wird vom WLA unterstützt. Ein verpflichtendes Zitat der DKRZ-Publikation in der Literatur-Referenzliste von wissenschaftlichen Artikeln wird hingegen abgelehnt.

### b) Nachfolge Stammer/Böning

Die Nachfolgeregelung für ausscheidende Mitglieder des WLA wird auf der kommenden Gesellschafterversammlung des DKRZ diskutiert werden.

## 6. Rechenzeitanträge

Im nichtöffentlichen Teil der Sitzung wurde unter anderem über die Rechenzeitanträge für Community- und Konsortial-Projekte beraten.

Es wurden Ressourcen für Neu- und Folgeprojekte über den Zeitraum vom 1.1.2021 bis 31.12.2021, sowie zusätzliche Ressourcen über den Zeitraum vom 1.1.2021 bis 30.6.2021 bewilligt. Im einzelnen sind dies:

	Beantragt	Bewilligt
Mistral [Node hours]	18.838.420	14.964.878
Lustre work [GiB]	26.162.886	19.625.656
HPSS arch [GB]	26.874.456	24.164.502
HPSS docu [GB]	5.026.058	3.432.073

*Ende der Sitzung: 16:17*

# Konzept für die Nutzung des Bereichs /pool/data am DKRZ

---

DKRZ: Karsten Peters-von Gehlen, Mathis Rosenhauer,  
Frank Toussaint, Hannes Thiemann

WLA: Sönke Zaehle

Usergroup: Bernadette Fritzsich





## /pool/data als Datenbereich für die projektübergreifende Nutzung von Daten

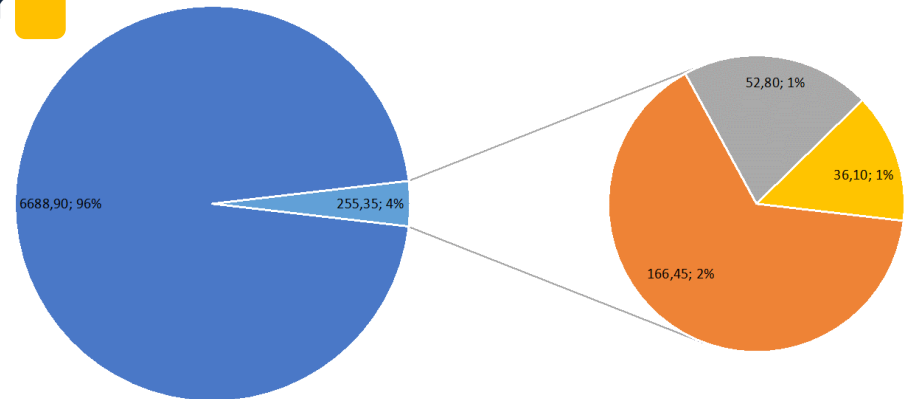
- 😊 von Nutzern sehr geschätzt
- 😊 30 Directories/Projekte, 6944 TB
- 😞 Historisch gewachsener Bestand
- 😞 Beschreibung, Verantwortlichkeiten
- 😞 Verfahren zur Vergabe bisher nicht einheitlich geregelt
- 😞 Plattenplatz als wertvolle Ressource

### Ziel:

- Transparentes Verfahren für Vergabe gewünscht
- Möglichst nahtloser Übergang für bestehende Daten

- Datenkategorien:

- Breite Community 
- Modellspezifisch 
- Beobachtungsdaten 
- Einzelne Institutionen/Nutzer 



- Datenablage:

- direkt in /pool/data (237,8 TB)
- verlinkt aus Projektverzeichnissen (6706,4 TB)



- Projektübergreifende Nutzung über **symbolische Links** in /pool/data
- Daten liegen in **Projektverzeichnissen** /work
- Daten haben eine bestimmte **Lebenszeit**
- Bestehende Gesellschafter- und Community-Projekte können für ihre Daten aus dem vorhandenen Kontingent solche Links bekommen
  - Begründung: Übergreifendes Interesse dokumentieren, Datenbeschreibung und Verantwortliche benennen
- **Pool-Projekte über Antrag**
  - Speicherplatz auf parallelem Filesystem
  - 2 Jahre Laufzeit
  - Begründung (s.o.)

Für Antragsverfahren werden Obergrenzen eingeführt:

- Summe < 1% des /work-Verzeichnisses (derzeit: 471 TB)
- Einzelprojekt < 82 TB
- Orientieren sich an bisherigen Daten

Entscheidung über Anträge

- DKRZ, laufend: bei **Unterschreitung beider Obergrenzen**
- WLA: sonst, zu den üblichen Calls

## 1. Beschreibung der Daten

- Bei Beantragung notwendig
  - *Titel/Bezeichnung des Datensatzes*
  - *Datenurheber/Kontaktperson (langfristig!)*
  - *Inhalt*
  - *Datenherkunft/Erstellungsmethode*
  - *Lizenz*
  - *Zitationsvorschrift*
- Bereitstellung eines Templates durch DKRZ

## 2. Information verbessern

- Übersicht auf DKRZ Webseiten  Nachnutzung
- Zugriffsmöglichkeiten (mistral, JupyterHub, FreVa, ...)

- Überführung der bisherigen Directories/Projekte relativ einfach möglich
- Nutzungskonzept in Usergroup Meeting 11.11.2020 abgestimmt
  - Probephase ?